UNIVERSIDADE FEDERAL FLUMINENSE
ESCOLA DE ENGENHARIA
PROGRAMA DE PÓS-GRADUAÇÃO EM
ENGENHARIA ELÉTRICA E DE TELECOMUNICAÇÕES

GUILHERME NUNES NASSEH BARBOSA

Desafios de segurança e predição de tráfego em centro de dados e redes de próxima geração

NITERÓI

2022

UNIVERSIDADE FEDERAL FLUMINENSE
ESCOLA DE ENGENHARIA
PROGRAMA DE PÓS-GRADUAÇÃO EM
ENGENHARIA ELÉTRICA E DE TELECOMUNICAÇÕES

GUILHERME NUNES NASSEH BARBOSA

Desafios de segurança e predição de tráfego em centro de dados e redes de próxima geração

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Engenharia Elétrica e de Telecomunicações da Universidade Federal Fluminense como requisito parcial para a obtenção do título de Mestre em Engenharia Elétrica e de Telecomunicações. Área de concentração: Comunicação de Dados Multimídia.

Orientador:
Diogo Menezes Ferrazani Mattos

NITERÓI

2022

Ficha catalográfica automática - SDC/BEE
Gerada com informações fornecidas pelo autor

B238d Barbosa, Guilherme Nunes Nasseh
Desafios de segurança e predição de tráfego em centro de dados e redes de próxima geração / Guilherme Nunes Nasseh Barbosa ; Diogo Menezes Ferrazani Mattos, orientador. Niterói, 2022.
67 p. : il.

Dissertação (mestrado)-Universidade Federal Fluminense, Niterói, 2022.

DOI: http://dx.doi.org/10.22409/PPGEET.2022.m.11059329743

1. 5G. 2. Segurança de redes. 3. Machine Learning. 4. Produção intelectual. I. Mattos, Diogo Menezes Ferrazani, orientador. II. Universidade Federal Fluminense. Escola de Engenharia. III. Título.

CDD -

Bibliotecário responsável: Debora do Nascimento - CRB7/6368

GUILHERME NUNES NASSEH BARBOSA

Desafios de segurança e predição de tráfego em centro de dados e redes de próxima geração

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Engenharia Elétrica e de Telecomunicações da Universidade Federal Fluminense como requisito parcial para a obtenção do título de Mestre em Engenharia Elétrica e de Telecomunicações. Área de concentração: Comunicação de Dados Multimídia.

Aprovada em 12 de Abril de 2022.

BANCA EXAMINADORA

Prof. Diogo M. F. Mattos, D.Sc. – Orientador, UFF

Profª. Dianne Scherly Varela de Medeiros, D.Sc. – UFF

Prof. Edelberto Franco Silva, D.Sc. – UFJF

Niterói

2022

Dedico este trabalho aos meus familiares.

Agradecimentos

Agradeço ao meu orientador Diogo Mattos pelos conselhos ao longo desta jornada tanto do ponto de vista pessoal quanto do ponto de vista acadêmico, mostrando-se um amigo dos alunos e nos fazendo compreender a necessidade da busca pela excelência. Agradeço ainda à professora Dianne Medeiros pela ajuda nesta trajetória, sempre disposta a ajudar os alunos, nos incentivando a buscar o melhor de nós.

Agradeço à minha mãe Angela e minha irmã Mariangela pelo incentivo em todos os momentos.

À minha namorada por todo apoio durante esse período e compreensão.

Aos amigos que pude fazer no Labgen - (Laboratório de Ensino e Pesquisa em Redes de Nova Geração) e Midiacom através dos projetos de pesquisa.

Aos colegas da Superintedência de Tecnologia da Informação STI - da UFF, em especial ao Superintendente Hélcio Rocha, e aos coordenadores Junior Barroso e Felipe Pimenta, por todo apoio no período que pude trabalhar junto com eles.

À empresa Equinix pelo incentivo e apoio de gestores e colegas.

Agradeço por fim as agências de fomento CNPq, CAPES, FAPERJ, FAPESP, bem como a Prefeitura da Cidade de Niterói e a Rede Nacional de Pesquisa pelas iniciativas de desenvolvimento.

Resumo

O aprimoramento das próximas gerações de redes móveis, em especial da quinta geração (5G), propiciou uma melhoria na qualidade do serviço, bem como possibilitou a expansão no desenvolvimento de novas aplicações. Mesmo com os desafios impostos pelo meio de propagação, as redes 5G oferecem taxas de transmissão e latências capazes de competir com meios confinados ofertados para o público residencial, facilitando, em alguns aspectos, a infraestrutura necessária para o acesso. Com capacidade incrementada, novos desafios são apresentados, pois a infraestrutura para a próxima geração de redes, requer a implementação de técnicas como redes definidas por *software* (SDN) e funções de redes virtualizadas (NFV) para operação descentralizada. Estas técnicas requerem análises constantes da infraestrutura, para que novas instâncias possam ser implementadas a fim de atender demandas específicas. Com isso, a predição de tráfego da utilização de recursos torna-se um desafio, uma vez que nesses ambientes há compartilhamento de recursos, sendo preciso identificar novas técnicas com baixo custo computacional para realizar predições. O alto volume de tráfego gerado nestas redes aumenta o grau de complexidade, pois além da necessidade de análise temporal, anomalias e ameaças de segurança podem permanecer ofuscadas e gerar danos irreparáveis. Assim, propõe-se uma comparação entre técnicas clássicas de predição utilizando modelos estatísticos e modelos de aprendizado de máquinas, para análise do tráfego em uma série temporal. Aplica-se a entropia de Shannon em características categóricas do fluxo para realizar a predição. A entropia é uma importante ferramenta, pois apresenta o grau de dispersão ou aleatoriedade de um sistema. O pré-processamento do tráfego, utilizando funções *wavelet*, capazes de decompor um sinal, apresenta-se com uma ferramenta promissora, uma vez que pode ser aplicada no reconhecimento de padrões e ataques em redes de computadores. Por fim, é realizado um comparativo entre duas técnicas clássicas, o modelo *Autoregressive Integrated Moving Average* (ARIMA) e o modelo *Long short-term memory* (LSTM). A técnica LSTM apresenta-se até 8 vezes mais eficiente do que a técnica ARIMA apresentando tempo total de execução, 42% mais eficiente.

Palavras-chave: Segurança de redes, 5G, 6G, Internet das Coisas, Machine Learning, Predição de tráfego.

Abstract

The improvement of the next generations of mobile networks, especially the fifth generation (5G), provided an improvement in the quality of service, as well as allowed the expansion in the development of new applications. Even with the challenges posed by the means of propagation, 5G networks offer transmission rates and latencies capable of competing with confined means offered to the residential public, facilitating in some respects the necessary infrastructure. With this increased capacity, new challenges are presented because the infrastructure for the next generation of networks requires the implementation of techniques such as networks defined by *software* (SDN) and virtualized network functions (NFV). These techniques require constant analysis of the infrastructure, so that new instances can be implemented in order to meet specific demands. Thus, the prediction of traffic of resource utilization becomes a challenge, since in these environments there is resource sharing, and it is necessary to identify new techniques with low computational cost to make predictions. The high volume of traffic generated in these networks increases the degree of complexity, because in addition to the need for temporal analysis, anomalies and security threats can remain overshadowed and generate irreparable damage. Thus, it is proposed a comparison between classical prediction techniques using statistical models and machine learning models, for traffic analysis in a time series. Shannon entropy is applied in categorical flow characteristics to perform prediction. Entropy is an important tool because it is able to present the degree of dispersion or randomness of a system. Traffic preprocessing, using *wavelet* functions, capable of decompose a signal, presents itself with a promising tool, since it can be applied in the recognition of patterns and attacks on computer networks. Finally, a comparison is made between two classical techniques, the *Autoregressive Integrated Moving Average* (ARIMA) model and the *Long short-term memory* (LSTM) model. The LSTM technique is up to 8 times more efficient than the ARIMA technique with a total execution time, 42% more efficient.

Keywords: Network Security, 5G, 6G, Internet of Things, Machine Learning, Network Predction.

Lista de Figuras

Lista de Tabelas

Lista de Abreviaturas e Siglas

Sumário

Capítulo 1

Introdução

A quinta geração (5G) e a sexta geração (6G) de sistemas de comunicações móveis representam uma nova era, na qual a conectividade apresentará fluidez e flexibilidade de interconexão entre pessoas, dispositivos inteligentes e máquinas. Estima-se que até 2025 a rede 5G abrangerá aproximadamente um terço da população mundial. Uma das principais características das redes de próxima geração é a capacidade de adaptação às necessidades dos usuários, permitindo que diferentes serviços se adequem dinamicamente para utilizar as vantagens da comunicação, como carros autônomos, telemedicina, através de cirurgias remotas, comunicação *Machine-to-Machine* (M2M), entre outros. Uma das principais características das próximas gerações de redes móveis, de acordo com a norma ITU-R M.2083-0 da *International Telecommunication Union* (ITU), é a melhoria das taxas de transmissão, na qual a rede 5G, por exemplo, operará com taxa de transmissão de até 20 Gb/s da Estação Rádio Base (ERB) para aplicações *enhanced Mobile Broad Band* (eMBB) e com taxas de transmissão típicas de usuários de até 1Gb/s. A latência de até 1 ms é também fator determinante para que aplicações *Ultra Reliable Low Latency Communications* (URLLC) sejam possíveis, como a operação de carros autônomos.

Desafios de segurança em redes de computadores, atualmente, se encontram principalmente na camada de aplicação, devido à diversidade e complexidade das aplicações existentes e à versatilidade hoje existentes para alavancar o desenvolvimento dessas aplicações. Com largura de banda consideravelmente superior e latências inferiores a 1 ms, a rede 5G e a 6G, por exemplo, fomentam o aprimoramento e desenvolvimento de aplicações principalmente com foco em sensoriamento remoto. Aplicações clássicas, como *Voice over New Radio* (VoNR), são susceptíveis a problemas de segurança elementares, como confidencialidade e privacidade de dados. Outros problemas completamente novos, como roubo de identidade virtual ou a extrapolação do consentimento do usuário atra-

vés do aprendizado dos seus dados baseado em hábitos também são objeto de pesquisas relacionadas à segurança de redes.

A sociedade está mais dependente do setor de telecomunicações e a quinta geração de redes móveis já está amplamente difundida em diversas redes comerciais, sendo objeto de investimento em diversas operadoras de diferentes países, tendo seu desempenho e flexibilidade associados ao conceito de redes sem fio definidas por *software* – *Wireless Software Defined Network* (WSDN). Através deste paradigma é possível realizar o fatiamento da rede (*network slicing*), compartilhando a infraestrutura entre usuários e garantido recursos necessários orientados à aplicação. Como exemplo, os serviços de *streaming* receberão recursos baseados em largura de banda, enquanto cirurgias remotas precisarão garantir latências extremamente baixas. Este mecanismo é obtido através da técnica de virtualização das funções de rede – *Network Functions Virtualization* (NFV), sendo o seu principal objetivo realizar a migração da utilização de *hardware* dedicado para máquinas virtuais, introduzindo agilidade, escalabilidade e interoperabilidade.

A predição o uso de recursos computacionais ou baseados em tráfego de dados é um fator essencial para garantir a correta distribuição dos recursos implementados a partir das funções de redes virtualizadas. Do ponto de vista operacional, há uma maior agilidade para configurar uma infraestrutura utilizando máquinas virtuais a partir de técnicas de automação. A capacidade de estimar os recursos necessários baseados em históricos de uso, são atividades desafiadoras. Há uma tendência natural do tráfego de rede tornar-se cíclico em função da repetição de atividades diárias das pessoas em suas movimentações, facilitando as técnicas de predição de carga. No entanto, uma anomalia na rede pode ser identificada através de uma simples mudança no comportamento da rede, como um pequeno aumento no número de pacotes transmitidos sem mudanças em aplicações que justifiquem tal alteração.

1.1 Motivação

A tecnologia 5G incita melhorias para a garantia de privacidade e segurança dos dados trafegados. Um maior número de dispositivos interconectados tem a capacidade de aumentar o tráfego e, consequentemente, a superfície de ataques. Privacidade e segurança demandam a implementação em diversas camadas, incluindo dispositivos, equipamentos de interface aérea, infraestrutura de rede de acesso de rádio na nuvem – *Cloud - Radio Access Network* (C-RAN), entre outros. As próximas gerações de redes são projetadas

para solucionar as funcionalidades ineficazes de suas antecessoras, como limitações de autenticação a partir de dispositivos, exposições relativas à privacidade do usuário e vulnerabilidades da interface de rádio [1]. O consórcio 3GPP estabelece a especificação 33.501 para definição dos requisitos de arquiteturais e procedimentais de segurança dos sistemas 5G, focando em autenticação, autorização de acesso e equipamentos, integridade de dados entre outros, garantindo a homogeneidade de elementos de rede.

As próximas gerações de redes terão um conjunto de desafios de segurança devido principalmente aos seguintes fatores: maior número de usuários, heterogeneidade de dispositivos interconectados, novas aplicações, incluindo de missão crítica, questões de privacidade do usuário e suporte a dispositivos baseados em *Internet of Things* (IoT). A utilização de paradigmas como *Software Defined Network* (SDN), NFV, *Mobile Edge Computing* (MEC), *Software Defined Wide Area Network* (SD-WAN) apresentarão desafios adicionais à infraestrutura [2] além dos desafios referentes a ataques aos dispositivos do núcleo até a borda da rede.

Ataques de *malware* são caracterizados como atemporais, pois também permanecem como ameaças às novas redes, porém, com uma maior capilaridade de disseminação, podendo impactar serviços públicos críticos. *Ransomware*, *Keylogger* e botnets são atualmente os principais ameaças desta categoria. As botnets são amplamente empregadas e utilizadas para a disseminação de *malware*. Outro ataque eficaz é o de negação de serviço – *Denied of Services* (DoS), considerado ainda uma das piores ameaças. Neste ataque, o atacante realiza tarefas como exploração de vulnerabilidades de protocolos com intuito de degradar o serviço, impedindo que usuários legítimos utilizem os recursos da rede ou sistema. Esses ataques cibernéticos possuem tamanha relevância, que a depender do contexto, podem inclusive preceder guerras, como no caso do conflito entre Rússia e Ucrânia, iniciado em fevereiro de 2022. As Forças Armadas da Ucrânia, bem como os bancos estatais foram alvos de ataques de negação de serviços distribuídos – *Distributed Denied of Services* (DDoS) antes da investida militar de fato [1].Esses ataques levaram à paralisação temporária de sites além da inviabilização de serviços financeiros. Desta forma, os ataques não estão contextualizados apenas no âmbito empresarial, mas também como influenciador em fatores geopolíticos regionais e globais.

[1]Disponível em: https://blogs.microsoft.com/on-the-issues/2022/02/28/ukraine-russia-digital-war-cyberattacks/

1.2 Contribuição

Esta dissertação propõe um comparativo entre técnicas de aprendizado de máquinas e técnicas clássicas de estatística, para realizar a predição em redes de computadores. Como principal contribuição, esta disertação avalia a capacidade de predizer o tráfego de rede de componentes lineares e não lineares de uma série temporal, através da criação de uma nova métrica baseada na entropia de Shannon das características categóricas do fluxo de rede. A entropia é uma ferramenta capaz de compreender o grau de dispersão ou aleatoriedade de um sistema, e os fluxos de redes, possuem um grau de aleatoriedade considerável referente a algumas características. Neste trabalho, a entropia foi utilizada com o objetivo de gerar novas características a partir das variáveis qualitativas existentes no fluxo, tais como IP de origem e destino. Por possuírem esta característica qualitativa, não são comumente utilizadas para realizar predição em redes. Entretanto, são informações elementares em qualquer fluxo de redes, e isto pode ocasionar um melhor aproveitamento da coleta de métricas para realizar predições e detectar anomalias. Com o objetivo de aproveitar toda a informação exisitente em uma série temporal, o pré-processamento do tráfego através de funções *wavelet* foi utilizado para decomposição do sinal. Do ponto de vista de implementação, esta técnica possuí poucos parâmetros de configuração quando utilizada a biblioteca disponível em Python [2] facilitando o reconhecimento de padrões e ataques a partir das componentes de alta e baixa frequência extraídas.

1.3 Organização da Dissertação

Esta dissertação está organizada da seguinte forma. O Capítulo 2 apresenta a fundamentação teórica das ameaças e desafios das redes de próxima geração com relação à predição de tráfego e detecção de anomalias. O Capítulo 3 descreve os trabalhos referentes à detecção de anomalias e predição de tráfego e sua aplicabilidade. O Capítulo 4 refere-se a uma avaliação de um modelo proposto para predição de tráfego e detecção de anomalias utilizando características não categóricas como fonte de dados. Por fim, o Capítulo 5 conclui a dissertação evidenciando as contribuições e apresentando futuros trabalhos.

[2] https://pypi.org/project/PyWavelets

Capítulo 2

Fundamentação Teórica

As gerações de redes móveis sempre são projetadas com intuito de aprimorar a qualidade do serviço e corrigir vulnerabilidades das gerações anteriores. A primeira geração (1G) com tecnologia analógica e padrão baseado no sistema *Advanced Mobile Phone System* (AMPS) era orientada somente à voz. Por sua modulação ser baseada em frequência – *Frequency Modulation* (FM) e o sistema não possuir mecanismos de criptografia, era possível, de maneira trivial, interceptar ligações através de receptores que operassem na mesma faixa de frequência nas proximidades de um aparelho. A segunda geração (2G) estabelece o inicio das comunicações móveis digitais na década de 1990. Um dos principais serviços ofertados é o de envio de curtas mensagens de texto. Nessa geração, existe um considerável avanço em processos de autenticação através do *Subscriber Identity Module* (SIM) do usuário e melhoria na confidencialidade. A terceira geração (3G) é marcada por uma ascensão da comunicação de dados, na qual o acesso à Internet era o principal serviço ofertado pelas operadoras. A partir das redes 3G, foram aprimoradas as questões relativas à autenticação, utilizando a autenticação mútua para evitar ataques com estações rádio base falsas. A quarta geração de redes móveis (4G) possui taxa de transmissão muito superior às suas antecessoras e o seu desenvolvimento foi baseado completamente no protocolo *Internet Protocol* (IP). As redes 4G empregam protocolos criptográficos com foco na autenticação do usuário oferecendo proteção contra ataques na camada física. A Tabela 2.1 apresenta um comparativo entre os mecanismos adotados por cada geração e os desafios a serem superados.

Tabela 2.1: Comparativo entre ameaças entre as gerações de redes móveis antecessoras ao 5G

Geração	Mecanismo de Segurança	Desafios de Segurança
1G	Sem medidas explícitas de segurança e privacidade.	Bisbilhotamento, interceptação de chamadas e nenhum mecanismo de privacidade.
2G	Proteção baseada em autenticação, anonimato e criptografia.	Estação base falsa, segurança de link de rádio, autenticação unilateral e spamming.
3G	Adotou a segurança 2G, acesso seguro à rede, Autenticação e Acordo de Chave (AKA) e autenticação mútua.	Vulnerabilidades de segurança de tráfego IP, segurança de chaves de criptografia, segurança de roaming.
4G	Nova criptografia (EPS-AKA) e mecanismos de confiança, segurança de chaves de criptografia, segurança de acesso do 3GPP e proteção de integridade	Maior segurança induzida por tráfego de IP, integridade de dados, segurança de Base Transceiver Stations (BTS) e interceptação de chaves de longo prazo.

2.1 Desafios de segurança na rede 5G e de próximas gerações

A segurança das redes 5G possuem características intrínsecas das comunicações móveis acrescida ao fato de ter uma conexão à Internet orientada a objetos. As próximas gerações de redes móveis, inclusive na quinta geração, segurança e privacidade são os principais focos. Diversos tipos de dados fazem parte da arquitetura de redes de próxima geração, entre dados dos usuários, dados sobre o usuário, arquivos de configuração, arquivos de registro (*logs*) entre outros. Tais informações, de acordo com Dutta e Hammad [3] serão utilizadas para habilitar funcionalidades do núcleo da rede e possibilitar a automação das decisões voltadas para aplicações e o gerenciamento de sistemas. Pode-se citar ainda como desafios adicionais, a classificação e proteção adequada dos dados inativos (*at-rest*), considerados menos vulneráveis, e dados ativos (*in-transit*). A privacidade deve ser considerada no projeto das redes e a configuração do ambiente deve garantir que apenas os dados necessários sejam coletados e armazenados. Um ataque tradicional DDoS, por mais simples que seja, pode interromper serviços e resultar em alto impacto para aplicações sensíveis à latência, como carros autônomos e tele cirurgia, por exemplo. Como as redes de próxima geração possuem relações diretas com redes definidas por *software* e funções de rede virtualizadas, é importante entender que as redes possuem diversas superfícies de ataque. A literatura expõe diversas técnicas amplamente conhecidas, porém com danos

e perdas potencialmente amplificados em função das novas aplicações orientas à Internet das Coisas. Do ponto de vista das comunicações móveis, incluindo a rede 5G, podem-se destacar as seguintes ameaças [4]:

- **Segurança das interfaces de rádio**: As chaves criptográficas da interface rádio são transmitidas por canais inseguros;
- **Integridade do plano do usuário**: A terceira e quarta geração de redes móveis possuem proteção, mas não para dados no plano do usuário. A integridade da camada de transporte e da camada de aplicação com criptografia são usadas se a integridade dos dados for necessária. Neste caso, um ataque *man-in-the-middle* e o sequestro de sessão são possíveis;
- **Tempestades de sinalização**: Quando um número de dispositivos M2M acessam a rede simultaneamente, a demanda de sinais de controle aumenta consideravelmente e forma a tempestade de sinalização;
- **Fatiamento de rede**: O fatiamento de rede permite o compartilhamento de recursos do núcleo da infraestrutura com maior eficiência, facilitando a alocação mais dinâmica para atender com agilidade diferentes aplicações, potencializando os riscos. Controles de segurança devem ser implementados para garantir o isolamento adequado das redes. Esses controles incluem primordialmente a categorização de cada fatia alocada, para proteção do fluxo de dados entre as fatias, mitigando assim, ataques laterais entre infraestruturas ou ataques que comprometam os recursos compartilhados;
- **Redes definidas por software - SDN**: Por permitir maior flexibilidade no gerenciamento das redes através de programação, e consequentemente centralizando de maneira lógica os planos de controle da rede, as SDN potencializam os desafios de segurança. O controlador SDN permite atualizar ou modificar regras de fluxo nos nós da rede, permitindo manipular o encaminhamento de dados. A centralização do controle da rede pode tornar o controlador um gargalo para a infraestrutura no caso de ataques. Através de programação, diversas funções de rede podem ser implementadas como aplicações, e caso haja o vazamento de credenciais e acesso à Interface de programação de aplicações *Application Programming Interface* (API), atacantes facilmente podem ter controle parcial ou total da rede;
- **Virtualização das Funções de rede - NFV**: Devido a sua característica dinâmica, desafios relacionandos a gestão de configuração devem ser abordados com

maior foco. Erros nas configurações podem ocasionar falhas de segurança propiciando ataques clássicos como *spoofing* e *sniffing*, por exemplo. As NFV ainda estão sujeitas à vulnerabilidades específicas relativas à virtualização como ataques *Side-channel*, *flooding* e sequestro de hipervisor. Devido ao acesso compartilhado da infraestrutura de vários clientes, ambientes comprometidos ou usuários mal intecionados com privilégios elevados podem interferir na operação da infraestrutura manipulando o tráfego de rede.

As redes definidas por *software* são ponto de atenção para as redes de próxima geração. Esse paradigma simplifica o gerenciamento das redes através de programação aplicada às funções do plano de controle. O plano de controle é logicamente centralizado para criação de políticas de encaminhamento e o plano de dados é distribuído para operar o tráfego baseado em políticas de encaminhamento. A centralização lógica das SDN apresenta numerosas vulnerabilidades [5]. As interfaces entre os planos são chamadas de *northbound*, entre o plano de controle e o plano de aplicação, e *southbound*, entre os elementos comutadores do plano de dados e do plano de controle corfome ilustrado na Figura 2.1. O plano de controle em particular é sujeito a ataques de negação de serviço devido à sua característica central, tornando-o ponto único de falhas. Ademais, há outras possíveis ameaças salientadas na literatura [6], entre elas: **acesso não autorizado**, referente ao controlador da redes ou às aplicações; **vazamento de informação**, através da descoberta de regras de fluxos ou políticas de encaminhamento, chaves criptográficas ou certificados para cada rede lógica; **modificação de dados**, no qual há alteração de regras de fluxos para adulterar pacotes, **aplicações maliciosas**, onde as aplicações permitem a inserção de regras fraudulentas; **negação de serviço**, seja através da inundação da comunicação do controlador-comutadores ou inundação da tabela de fluxo em cada comutador; **segurança do sistema SDN**, em que os comutadores OpenFlow podem operar no modo de falha autônoma, quando o comutador é desconectado do controlador, tornando-o o mesmo ser alvo de ataques. O protocolo OpenFlow permite a programação das redes para que os controladores SDN gerenciem, controlem e monitorem o tráfego através do plano de encaminhamento de ativos de rede como *switches* [7].

A próxima geração das redes de telecomunicações deverá suportar ainda uma quantidade significativa de terminais inteligentes, tais como celulares e sensores, disponibilizando aplicações de tempo real e provendo inteligência e confiança embarcadas na infraestrutura do núcleo e bordas da rede. Para atender a esses requisitos, a sexta geração das redes móveis, 6G, compreende o uso de novas tecnologias de inteligência artificial, cadeia de blocos (*blockchain)* e de fornecimento de serviços para Internet das Coisas. A rede 6G está sendo

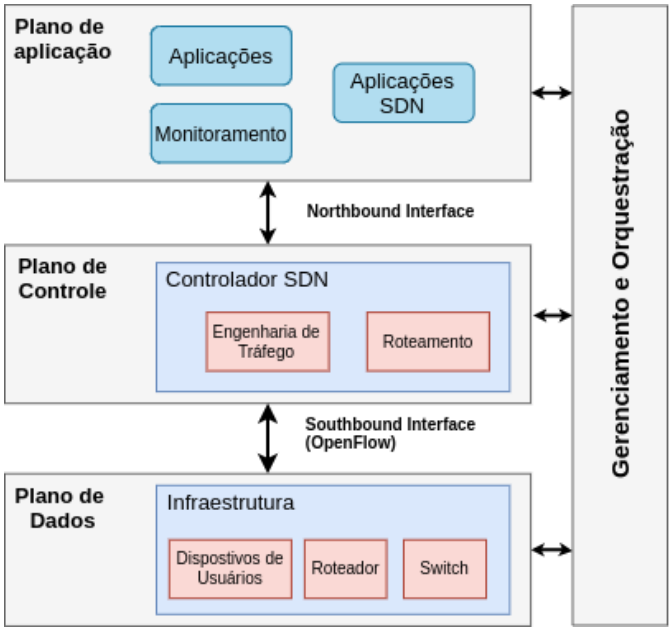


Figura 2.1: Separação entre planos das redes das redes definidas por *software*

desenvolvida para contornar as limitações existentes atualmente e possibilitar a utilização de novos paradigmas, como novas interações homem-homem e homem-máquina, utilização de frequências na faixa de terahertz, redes tridimensionais, comunicações quânticas, superfícies refletoras inteligentes, entre outras [8]. Estima-se que as taxas de transmissão serão da ordem de 1 Tb/s superando em cinquenta vezes a capacidade da rede 5G [9]. As aplicações 6G futuras apresentarão requisitos rigorosos e exigirão recursos de rede estendidos em às redes 5G desenvolvidas atualmente [10]. Na rede 6G, todos os dispositivos de ponta são concebidos para se conectarem à Internet e as aplicações de inteligência artificial serão amplamente usadas por esses dispositivos. A maioria das aplicações de inteligência artificial serão orientadas a dados, aumentando a preocupação com a segurança e privacidade das informações coletadas [11]. A privacidade dos clientes pode ser comprometida, caso haja o vazamento de dados ou o comprometimento dos modelos de aprendizado [12].

A utilização da rede 6G tem como objetivo aprimorar a fidelidade das comunicações, tendo como desafios estabelecer comunicações ultra confiáveis e de baixa latência URLLC. Esses conceitos permitem contribuir de maneira significativa em diversas áreas de missão crítica, permitindo que a comunicação tátil seja implementada para possibilitar que as interações físicas em tempo real sejam executadas, como a tele cirurgia. Outro conceito que ganhará notoriedade e demandará infraestruturas robustas é a holografia. Esta técnica emprega artifícios óticos para projetar luz e fornecer imagem em três dimensões, sendo objeto de estudos principalmente para a telemedicina, provendo atendimento médico em

áreas remotas ou realização de procedimentos cirúrgicos [13].

2.2 Principais desafios de predição de tráfego e detecção de anomalias

O aprimoramento das tecnologias permitiu que os meios de comunicação sem fio se tornassem populares, fazendo com que a tecnologia associada evoluísse continuamente e de maneira célere, para sustentar a comunicação de dados em tempo real com qualidade necessária para serviços de missão crítica. Todavia, a rede 5G possui inúmeros sensores e dispositivos como partes primordiais das comunicações, sendo imprescindível garantir robusta proteção do ponto de vista de infraestrutura, privacidade de usuários e, sobretudo, *software* destes dispositivos [14]. Lopez-Martin *et al.* propõem a utilização de codificadores automáticos variacionais condicionais (*Conditional Variational Autoencoders*) para comprender os rótulos de intrusão dentro das camadas de decodificação, possibilitando ser utilizado para predição de ataques e reconstrução de informações faltantes [15]. Por possuir uma única fase de treinamento, o modelo torna-se útil no que corresponde a otimização de recursos computacionais. Devido à sua complexidade, a rede 5G exige o desenvolvimento de arquiteturas e soluções com alta resiliência. Ahmad *et al.* categorizam os principais desafios como [4]:

- **Tráfego de rede repentino**: descreve um aumento relevante de aparelhos e dispositivos conectados à rede, podendo ser contornado mediante melhoria dos recursos existentes ou da adição de mais recursos conforme aumento da demanda, utilizando redes definidas por *software* ou funções de rede virtualizadas;
- **Integridade do plano do usuário**: por ter uma característica de dados de usuários, é fundamental a utilização de criptografia ponta a ponta. Aplicações específicas podem exigir a utilização de outras camadas de segurança, além da comunicação criptografada.

A comunicação móvel proporciona avanços científicos ao longo de suas gerações, flexibilizando e universalizando a troca de informações em tempo real, permitindo que usuários e dispositivos se tornem ubíquos, através de computação móvel, redes de sensores entre outros. Entretanto, essas características contribuem para que os ataques se intensifiquem, pois a superfície de ataque aumenta acompanhado a curva de crescimento de dispositivos conectados, possibilitando ainda que estes mesmos dispositivos sejam vetores de ataques

distribuídos, ocasionando roubo de informações e até mesmo guerra cibernéticas. O tráfego dentro de uma célula usualmente expõe flutuações recorrentes e possui rajadas a qualquer instante, assemelhando-se com o comportamento de pessoas, que possuem características aleatórias no deslocamento ao longo do dia. A análise de segurança das redes tem ganhado foco em diversos campos de pesquisa, sobretudo na detecção de anomalias. Entretanto, a detecção em tempo real torna-se desafiadora em função da quantidade de dados gerados pelos dispositivos, pois requer um monitoramento ininterrupto de eventos, processos e mensagens na infraestrutura [16]. As técnicas de detecção de anomalias em tempo real, em função dos desafios previamente listados, utilizam de forma majoritária ferramentas estatísticas por possuírem baixo custo computacional [17]. Estudos apresentam os métodos de Holt-Winters [18] e detecção do ponto de mudança (*change point detection*) [19] para detectar anomalias em redes de computadores. Ho Bang *et al.* propõem um modelo utilizando Cadeia Oculta Semi-Markoviana – *Hidden Semi-Markovian Model* (HSMM) para detecção de ataques de sinalização na rede *Long-Term Evolution* (LTE) [20]. Um processo é dito semi-markoviano quando a probabilidade de ocorrer uma mudança de um estado oculto para outro estado depende do tempo decorrido a partir do estado atual. Um dos benefícios do emprego desse modelo é a capacidade de capturar atributos estatísticos do tráfego de rede e, uma vez utilizando os parâmetros do modelo do HSMM, é possível identificar anomalias de acordo com sua probabilidade ou entropia [21].

O aprendizado de máquina é amplamente utilizado para empreender a detecção de anomalias e predição de tráfego. No entanto, para aplicações com necessidade de baixa latência, alguns algoritmos necessitam de uma sequência ininterrupta de dados, podendo ocasionar um custo computacional superior em função do grande fluxo de informações a serem processados. A arquitetura da rede 5G tende a ser complexa. Fu *et al.* abordam dois desafios para o gerenciamento do tráfego de rede [22]. O primeiro, é o fato da rede ser heterogênea e a simultaneidade de redes distintas com diversas características tornar a predição de tráfego mais complexa. O segundo ponto diz respeito da rede ser majoritariamente implementada utilizando SDN e NFV. Com o fatiamento da rede, os serviços são utilizados de maneira autônoma em infraestruturas compartilhadas e todo o tráfego gerado por cenários distintos é unificado na infraestrutura, tornando o núcleo da rede praticamente imponderável. O aprendizado por reforço profundo (*Deep Reinforcement Learning*) oferece ganhos expressivos particularmente em situações de alta latência e congestionamento da rede, por exemplo. Os métodos baseados neste modelo podem aprender informações de roteamento e padrão de tráfego e, assim, gerenciar de forma mais eficiente

os recursos da rede quando determinandas informações ocorrerem novamente, otimizando assim, o tempo de resposta do controlador [22].

As próximas gerações de redes são promissoras para promover a implementação de cidades inteligentes, agregando serviços públicos através de comunicação ubíqua de sensores, câmeras de segurança, entre outros. Além disso, a utilização de celulares inteligentes (*smartphones*) com o sistema operacional Android acrescenta um potencial risco. Sistemas baseados em Android representam 85% da utilização global e, consequentemente, se tornam alvos de ataques massivos [23]. É possível instalar aplicativos de terceiros, elevando significativamente a oportunidade de criação de *botnets* ou vazamento de informações pessoais.

A comunicação do sensores das cidades inteligentes ocorrerá primordialmente pela transmissão de dados através das redes móveis. Assim, da perspectiva de gerenciamento da interface de transmissão aérea, a estimativa das informações de estado do canal – *Channel State Information* (CSI), que representa as propriedades do canal de rádio, continua a ser um dos desafios elementares das redes 5G. O estado do canal, representado pelo CSI, tem um impacto significativo na alocação de recursos de rádio e gerenciamento de interferência, sendo utilizado para a determinação dos parâmetros da camada física do enlace. Devido à impossibilidade de envio frequente do valor do CSI pelo receptor, é essencial que o transceptor esteja apto para estimar precisamente o valor de CSI para permitir uma comunicação efetiva e otimizada no enlace. Métodos tradicionais para estimar o CSI possuem alta complexidade computacional e não são adequados para o uso em 5G devido ao emprego de tecnologias que aumentam o tráfego de dispositivos móveis, como – *Multiple Input Multiple Output* (MIMO) massivo e ondas milimétricas. Luo *et al.* sugerem um algoritmo de predição de CSI, denominado OCEAN, baseado em dados históricos de comunicação 5G [24]. Primeiramente, são identificadas diversas características importantes que afetam o CSI em um enlace de rádio, como faixa de frequência, localização, horário, temperatura, umidade do ar e tempo. Em seguida, considerando a relação espaço-temporal do CSI, os autores projetam um arcabouço de aprendizado que consiste em uma combinação de duas redes neurais convolucionais (CNN) e uma memória longa de curto prazo (LSTM). A arquitetura do sistema proposto compreende duas etapas de treinamento, uma *offline* e outra *online*. A primeira etapa é responsável por treinar a rede com dados históricos. A etapa *online* ocorre com o sistema em operação. Nessa etapa, em um intervalo de tempo constante, uma atualização do valor de CSI aferido é utilizada para treinar novamente a rede. Assim, utilizando a retroalimentação *online*, os valores previstos sempre são corrigidos e se adaptam às mudanças reais suportada pelo

canal, propiciando conclusões mais estáveis nas aplicações em sistemas de comunicação 5G.

2.3 Algoritimos para detecção de anomalias e predição de tráfego

A rede 5G requer uma maior automação relacionada à infraestrutura de telecomunicações. Técnicas de aprendizado de máquina são amplamente abordadas em predições de tráfego e detecção de anomalias, em especial pelo alto desempenho, compensando o custo computacional. Tais técnicas recebem dados com a finalidade de obter modelos capazes de descrever as observações realizadas, possibilitando a evidenciar comportamentos até então, desconhecidos. A partir dos modelos, decisões podem ser realizadas sobre tarefas específicas de forma precisa [25]. O núcleo da rede 5G deve ser escalável, e para isso, implementam-se novas instâncias sob demanda utilizando SDN e NFV. Para que isso seja realizado de forma automatizada e utilizando os recursos de rede de forma eficiente, é necessário prever a carga de uso da rede continuamente. Nesse sentido, a utilização de técnicas de aprendizado de máquina é fundamental [26]. Com o aumento de tráfego ocasionado por uma maior quantidade de dispositivos conectados, a segurança também exige esforços para detectar anomalias e ameaças antecipadamente na rede 5G, principalmente em ambientes com utilização de canal compartilhado e computação na borda, uma vez que brechas de segurança são decorrentes da comunicação com redes abertas, facilitadas pela virtualização de funções de rede [27].A detecção de ataques derivados de botnets, por exemplo, pode ser realizada através de aprendizado de máquina utilizando um sistema de detecção de intrusão baseado em redes neurais profundas – *Deep Neural Network* (DNN)[28], [29]. O aprendizado de máquina pode ser classificado em diversas categorias, sendo as principais listadas a seguir[30]:

- **Aprendizado supervisionado**, no qual as observações são fornecidas através de pares de entrada-saída e o objetivo do algoritmo é identificar uma função que relacione as entradas com as saídas. Assim, o algoritmo é capaz de predizer as próximas saídas baseando-se em amostras com padrões já rotulados. O treinamento é mantido até que se encontre um modelo mais otimizado possível de precisão e acurácia. Podem-se citar os modelos de máquina de vetor de suporte (*Support Vector Machine* - SVM), Redes Neurais, árvores de decisão, entre outros;
- **Aprendizado não supervisionado**, as observações que são fornecidas para os

algoritmos referem-se somente aos dados de entrada e sem rotulação, tendo o algoritmo como objetivo, agrupar as entradas em grupos aproximados denominados (textitclusters. K-Médias (*K-Means*), Floresta de Isolamento (*Isolation Forest*) e Rede de Crença Profunda (*Deep Belief Network*) são os principais algoritmos de aprendizado não supervisionado;

- **Aprendizado semi-supervisionado** caracteriza-se como uma interseção entre os modelos supervisionado e não-supervisionado. Os algoritmos são capazes de aprender a partir de um conjunto de dados parcialmente rotulado e generalizam o aprendizado para os demais dados não rotulados;

- **Aprendizado por reforço**, os algoritmos baseiam-se em um modelo de recompensas e punições, oferecidos a partir da interação do modelo com o ambiente. Não há mapeamento direto entre entradas e saídas e os resultados são obtidos a partir de retroalimentação (*feedback loop*) entre o sistema de aprendizado e o ambiente. A cada iteração, as ações disponíveis são anunciadas ao modelo no seu estado atual e, após a mudança de estado, recebe um sinal de reforço que tem o objetivo de fomentar um procedimento desejado, através de, ações que potencializam a recompensa a longo prazo [31]. Exemplos de algoritmos dessa categoria são *Q-Learning*, *Q-Learning* profundo (*Deep Q-Learning* - DQL) e Estado-Ação-Recompensa-Estado-Ação (*State–Action–Reward–State–Action* - SARSA).

As próximas gerações de redes, incluindo a 5G, possuem maior complexidade para análise de tráfego em tempo real em virtude da alta demanda de dispositivos conectados. Essa característica induz maiores esforços no processamento de fluxos e detecção de anomalias. Neste caso, o aprendizado profundo (*deep learning*) torna-se uma ferramenta valiosa. Esta técnica é uma área do aprendizado de máquina com o objetivo de identificar padrões complexos em estruturas de dados originários de representações inteligíveis.[32]. Trinh *et al.*, por exemplo, utilizam um modelo semi-supervisionado com uso de redes neurais recorrentes para detectar atividades legítimas [33]. A detecção é importante sobretudo em áreas metropolitanas, onde podem ocorrer anomalias causadas por aglomerações inesperadas e degradação no serviço de maneira involuntária.

As próximas seções abordarão os principais algoritimos baseados em redes neurais para identificação de anomalias e predição de tráfego em redes de computadores. Neste trabalho, entretanto, apenas os modelos ARIMA e LSTM são utilizados para avaliar a utilização da entropia das características categóricas do fluxo como métrica de predição.

2.3.1 Redes Neurais Recorrentes – RNN

As Redes Neurais *Feed-Forward*, conhecidas apenas como redes neurais, fazem parte dos primeiros algoritmos que fomentaram a inteligência artificial. Reproduzindo o comportamento do cérebro humano, foram utilizadas primordialmente na resolução de problemas de classificação e, com o avanço do poder computacional, tornaram-se capazes de trabalhar em diversos outros campos. As Redes Neurais *Feed-Forward* realizam o processamento de dados em sequência. A estrutura dessas redes é formada por três camadas, pelas quais o fluxo de informações transita de maneira unidirecional. A informação movimenta-se da camada de entrada que recebe as informações, representadas na figura por i_1 e i_2, para a camada oculta que aplica uma função matemática específica nos dados da camada anterior. Essa função é conhecida como função de transferência, representada na figura por h_k. Por fim, a informação termina na camada de saída, representando o resultado do treinamento da rede neural, ilustrado na figura pelas saídas o_1 e o_2. Por considerarem apenas a entrada atual, essas redes não possuem memória, não sendo viáveis para realizar predições de séries temporais. A estrutura desta rede neural pode ser exibida conforme ilustrado na Figura 2.2.

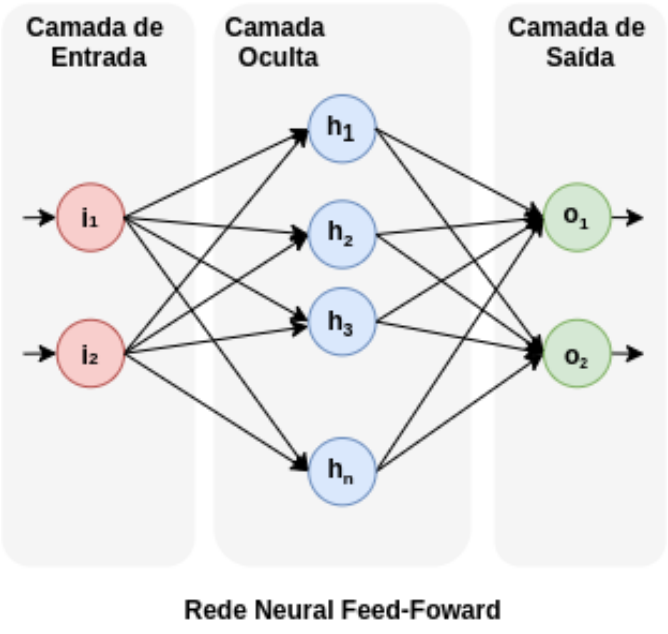


Figura 2.2: Camadas de uma Rede Neural *Feed-Forward* e interação entre os neurônios. A rede neural é formada por três camadas, entrada, oculta e saída. O fluxo de informação percorre a rede da entrada para a saída. Não há retroalimentação entre neurônios.

As redes neurais recorrentes – *Recurrent Neural Networks* (RNN) são uma variação de rede neural *feed-forward* que adicionam a capacidade de memorizar os estados ocorridos para processar os valores subsequentes de dados, dispondo de grande potencial para

realizar predições em séries temporais [34]. A Figura 2.3 apresenta um exemplo generalizado de RNN. Observa-se que, diferentemente das redes neurais *feed-forward*, há uma exoneração no sentido de fluxo da informação, podendo fluir através de conexões cíclicas, representadas na figura pelas setas tracejadas. Essas conexões permitem o acesso a estados anteriores, agregando a capacidade de memória à rede. Estudos utilizam as redes neurais recorrentes para predição de tráfego em virtude de capturar comportamentos mais complexos e não lineares, comparadas aos modelos estatísticos tradicionais, podendo exibir dependências de longo prazo [35]. Ramakrishnan e Soni comparam alguns modelos de redes neurais recorrentes, como o modelo de memória longa de curto prazo (*Long Short-Term Memory* - LSTM) e unidades recorrentes fechada (*Gated Recurrent Units* - GRU) com modelos estatísticos tradicionais, para mensurar o desempenho de cada um na predição do volume de tráfego, de pacotes por protocolo e distribuição de pacotes. A análise mostra que a LSTM possui o melhor desempenho dentre os modelos. As redes móveis 5G produzem dados sequenciais em larga escala [36], tais como fluxos de tráfego de dados e latência de aplicativos. Dessa forma, é interessante utilizar a RNN para aprimorar a análise de dados temporais nas redes móveis.

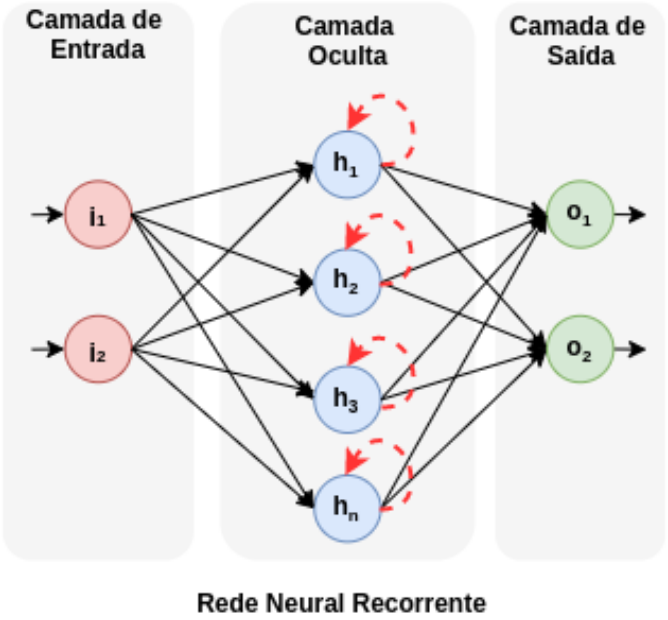


Figura 2.3: Interação entre os neurônios de uma Rede Neural Recorrente. A rede apresenta a capacidade memorizar estados passados e usá-los no processamento dos próximos dados. Há a retroalimentação de informação nos neurônios da camada oculta.

2.3.2 *Long Short-Term Memory* – LSTM

A rede neural LSTM é uma variante da RNN, entretanto os nós da rede possuem um estado interno de memória, podendo ser utilizado para armazenar e recuperar informações durante diversas iterações. Este modelo vem sendo extensivamente utilizado para modelagem de dados contínuos como processamento de linguagem e predição de séries temporais através de reconhecimento de padrões. Uma célula básica do modelo LSTM é apresentada na Figura 2.4. A célula é composta por três portas lógicas denominadas Porta de Esquecimento (*Forget Gate*), Porta de Entrada (*Input Gate*) e Porta de Saída (*Output Gate*). ***Forget Gate*** é responsável pela remoção dos valores que não são mais relevantes no estado da célula. Possui como entrada dois valores, sendo o primeiro h_{t-1} relativo ao valor da célula anterior e x_t, que representa uma entrada em um dado instante. Ambos são inseridos em uma função de ativação denominada **sigmoide**, fornecendo uma saída binária. Em seguida, o valor é multiplicado por uma matriz de peso (W_f), e adiciona-se o enviesamento (*bias*) b_f. O enviesamento é um fator de correção para ajustar o modelo. O modelo é representado pela Equação 2.1.

$$f^{(t)} = \sigma(W_f[h^{(t-1)}, x^{(t)}] + b_f). \tag{2.1}$$

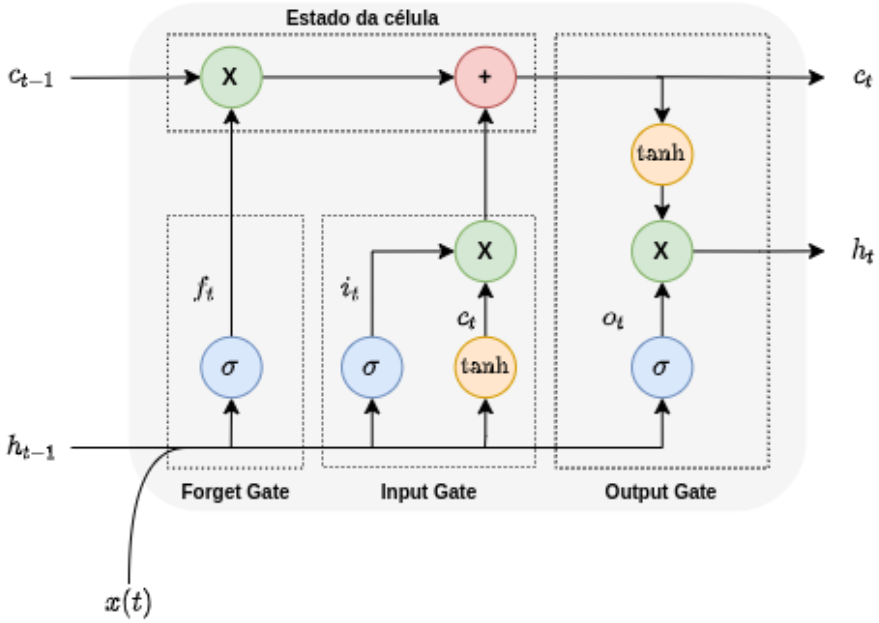


Figura 2.4: Célula básica de uma rede neural de memória longa de curto prazo (LSTM). As portas de entrada (*input gate*), esquecimento (*forget gate*) e saída (*output gate*) controlam o esquecimento ou o aproveitamento de estados anteriores. A memória interna e novos dados são ativados por funções *sigmoides* e de tangente hiperbólica.

2.3.3 Redes Neurais Convolucionais – CNN

Dentro do contexto de inteligência artificial e aprendizado de máquina, as redes neurais convolucionais – *Convolutional Neural Networks* (CNN) são um tipo de rede neural profunda (*Deep Neural Network* - DNN) utilizadas de maneira mais eficiente com dados de entrada com características multidimensionais, por exemplo, imagens. As CNNs podem ser utilizadas para classificar ou agrupar dados de saída, de acordo com um grau de similaridade atribuído pelo algoritmo. A influência para criação da rede neural convolucional é originária da estrutura do córtex visual do cérebro humano, com o objetivo de processamento das informações visuais. O termo visão computacional resume tais características biológicas através de processos e modelagens utilizando, sobretudo, algoritmos capazes de analisar imagens e classificá-las, semelhante ao cérebro humano. A estrutura clássica de uma CNN é composta por cinco camadas, conforme mostra a Figura 2.5. As principais camadas são as camadas convolucionais, camadas de agrupamento e camadas densas [37]. A **camada convolucional** contém filtros de tamanhos específicos, responsáveis por realizar a operação de convolução dos dados originados na camada de entrada, sendo imagens ou mapa de características, resultando em um novo mapa de características que irá alimentar a próxima camada. Matematicamente, uma imagem ou mapa de características é representado por uma matriz, tendo como componentes n_A, n_L e n_C representando altura, largura e número de canais respectivamente. Para o caso de uma imagem RGB, considera-se $n_C = 3$. Por convenção, considera-se que o filtro K é quadrado com dimensão ímpar, denominado por f, permitindo que cada pixel da imagem seja centralizado no filtro e, assim, considere todos os elementos em sua vizinhança. O produto convolucional entre a imagem e o filtro é uma matriz bidimensional, resultado de uma operação de multiplicação elementar entre o filtro e uma parte da imagem, conforme mostra a Figura 2.6 e expressada matematicamente como na Equação 2.2.

$$conv(I, K)_{x,y} = \sum_{i=1}^{n_A} \sum_{j=1}^{n_L} \sum_{k=1}^{n_C} K_{i,j,k} I_{x+i-1,y+j-1,k}, \tag{2.2}$$

em que K é a matriz representando o filtro aplicado à matriz de entrada I para a operação de convolução $conv$.

A **camada de agrupamento** é utilizada após a camada de convolução com o objetivo de reduzir as amostras das características extraídas da camada de entrada, sem impacto no número de canais, reduzindo a redundância de dados [37]. Por fim, a **camada densa** possui o objetivo de descrever de maneira mais detalhada as características extraídas da

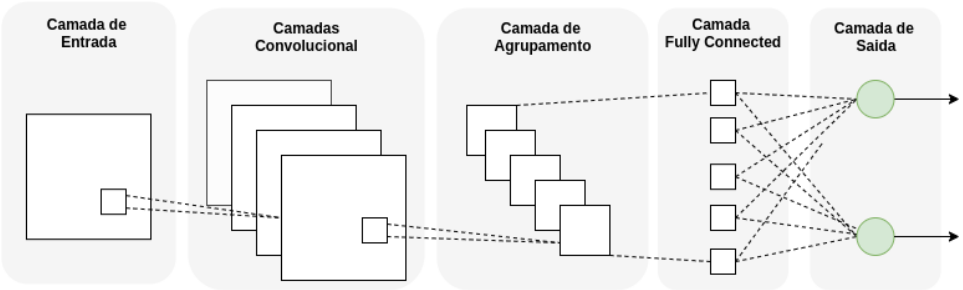


Figura 2.5: Estrutura simplificada de uma rede neural convolucional (*Convolutional Neural Network* - CNN). O aprendizado profundo com CNN é caracterizado pela repetição de camadas convolucional e de agrupamento. A cada par de camadas de convolução e agrupamento são extraídas características de mais alto nível. A camada densa (*Fully Connected*) realiza a classificação através das características extraídas. A consolidação do resultado ocorre camada de saída.

camada anterior. Uma função de ativação é utilizada para resultar na probabilidade de cada amostra na camada de saída.

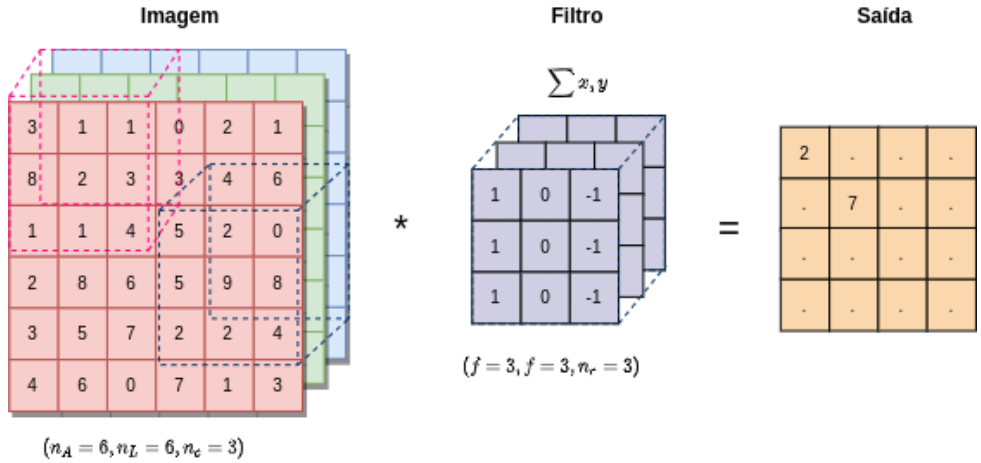


Figura 2.6: Operação convolucional entre imagem e filtro. O resultado da convolução é proveniente da multiplicação matricial entre segmentos da imagem de entrada e os filtros usados.

2.3.4 Codificadores Automáticos (*Autoencoders*)

Codificadores automáticos, *autoencoders*, são algoritmos de aprendizado de máquina não supervisionado utilizados para identificação e codificação dos dados de entrada. É amplamente utilizado como um pré-processamento de outras redes neurais com objetivo de reduzir a dimensionalidade dos dados e, por conseguinte, ignorar ruídos existentes no sinal, tendo a capacidade de aprimorar a entrada de dados de algoritmos supervisionados. A estrutura do codificador automático é composta por três camadas: uma camada de

entrada, uma camada oculta e uma cada de saída, sendo a camada oculta utilizada como **codificador** e a camada de saída como **decodificador** [38]. O codificador é representado por uma função $f(x)$ que transforma os dados de entrada em uma função h. O decodificador é uma função $g(x)$ que transforma a representação h para um valor reconstruído $\overline{x}$. Os codificadores automáticos são majoritariamente ferramentas para compressão de dados. Atualmente, duas aplicações práticas comuns dos codificadores automáticos são a eliminação de ruído, visto que transforma os dados em uma versão mais compacta, codificada com perdas, e a redução de dimensionalidade para visualização de dados. Com as restrições de dimensionalidade e esparsidade apropriadas, os codificadores automáticos podem aprender projeções de dados que são mais interessantes do que a análise de componentes principais (*Principal Component Analysis* – PCA) ou outras técnicas simples.A Figura 2.7 ilustra a estrutura do *autoencoder*.

Wu, Nekovee e Wang propõem um método de inferência da interferência dinâmica em um canal gaussiano multiusuário baseado em aprendizado profundo e codificadores automáticos [39]. A proposta é um mecanismo de codificador automático adaptativo. A intensidade da interferência é prevista por meio de um processo de aprendizado profundo, com a aprendizagem em linha (*online*) em tempo real do conhecimento do nível de interferência. Os resultados mostram que o codificador automático proposto funciona de forma mais robusta em um canal de interferência para todos os níveis. A melhoria é mais notável para os cenários de interferência forte e muito forte. A proposta estabelece uma base para permitir uma constelação adaptável para sistemas de comunicação 5G, nos quais condições de rede heterogêneas são consideradas.

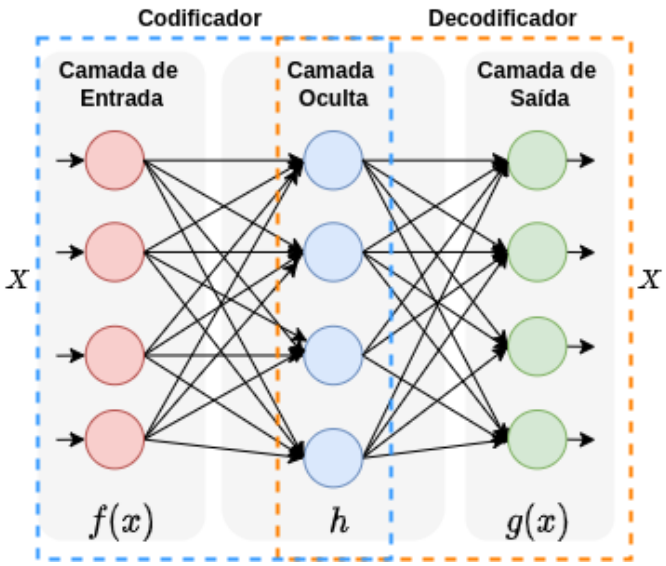


Figura 2.7: Estrutura geral dos Codificadores Automáticos

2.3.5 Máquina de Vetor de Suporte de Classe Única – OCSVM

Máquina de Vetor de Suporte – *Support Vector Machine* (SVM) é uma técnica de aprendizado supervisionada, para classificação binária ou de múltiplas classes. Nesse último caso, aplica-se uma SVM para cada classe. O modelo é baseado na teoria de aprendizagem estatística e seu objetivo é classificar um conjunto de dados através de um espaço multidimensional definindo um hiperplano de separação entre as classes de tal modo que os dados com as mesmas características estejam agrupados do mesmo lado do hiperplano. A proposta do algoritmo é encontrar os pontos mais próximos das linhas de cada uma das classes, conforme mostra a Figura 2.8, sendo esses pontos denominados **vetores de suporte**. Em seguida, calcula-se a distância do vetor de suporte até o hiperplano, denominado **margem** A finalidade do SVM é maximizar essa distância, pois assim será encontrado o hiperplano ideal do modelo, de forma a criar um limite de decisão entre as classes. É essencial que a margem tenha a maior amplitude possível para garantir a maximização da distância entre as classes. O SVM utiliza uma família de funções denominada ***kernel***, que possui diversos tipos, como linear, polinomial, *sigmoide*, dentre outras. O objetivo da função *kernel* é transformar o conjunto de dados, de modo que uma superfície de decisão não linear possa ser transformada em um plano de dimensão superior, fazendo com que a separação entre as classes seja tratada linearmente. Embora existam diversas funções de ativação disponíveis para implementação do LSTM, este trabalho teve como foco apenas a utilização do modelo e não sua otimização, embora seja de fundamental importância para utilização em equipamentos com limitação de recursos computacionais.

Uma variação do modelo tradicional é o modelo de classe única da máquina de vetor de suporte (*One-Class Support Vector Machine* - OCSVM), utilizado amplamente para detecção de anomalias. Para aplicação do modelo OCSVM utilizam-se no treinamento dados de uma única classe, ditos normais, ou dados contendo uma pequena fração de amostras anômalas. Com isso, o OCSVM é capaz de detectar amostras fora da classe alvo e amostras com novas características. As amostras que não pertencem à classe alvo estão distantes do hiperplano de decisão e são classificadas como pontos discrepantes (*outliers*), que nesse caso representam as anomalias. Diversas aplicações utilizam esse modelo. O OCSVM é utilizado para aprender o comportamento normal dos sensores de veículos conectados e automatizados [40]. Para detectar variações de anomalias, os autores realizam um processamento nos dados originários dos sensores para mitigar a influência de ruídos utilizando um filtro de Kalman estendido.

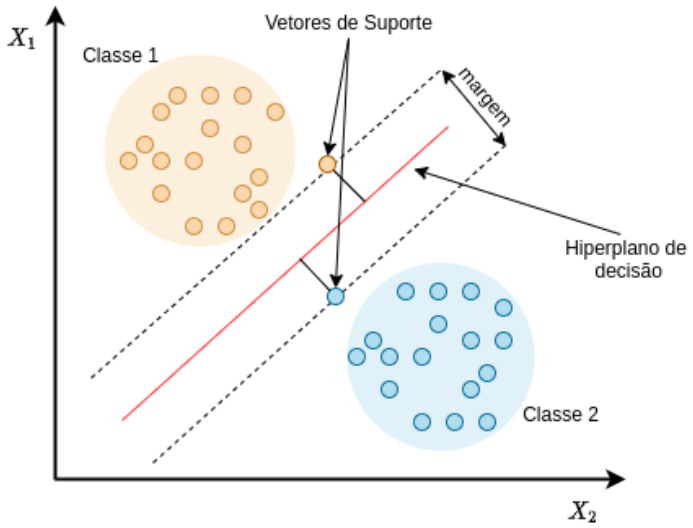


Figura 2.8: Estrutura geral do algoritmo máquina de vetor de suporte (*Support Vector Machine* - SVM). O algoritmo busca amostras de dados, vetores de suporte, que definem o hiperplano de separação entre classes. A seleção dos vetores de suporte visa maximizar a margem de separação entre classes.

2.3.6 Aprendizado Federado

O aprendizado federado (*Federated Learning* - FL) é um modelo de aprendizado com o objetivo de permitir que dispositivos móveis treinem de maneira colaborativa modelos preditivos compartilhados, mantendo os dados de treinamento localmente [12]. Isso garante, sobretudo, a segurança com relação aos dados, pois não é necessário o envio de informações sensíveis por parte dos usuários para um servidor central, sendo uma das principais técnicas que garantem a privacidade de uso. Um maior tráfego na borda da rede e o aumento considerável do poder computacional por parte dos dispositivos de usuários (*User Equipament - UE*), o aprendizado federado é uma técnica promissora para utilização nas redes de próxima geração. O objetivo desse é que cada dispositivo receba o modelo atual de um servidor central e, em seguida, utilize os próprios dados locais para treinamento local. Como cada cliente realiza um treinamento com dados distintos, são geradas pequenas atualizações locais que são enviadas para o servidor central. Por sua vez, o servidor central garante a agregação das atualizações originárias de todos os clientes, sendo calculada a média entre os modelos de todos participantes para melhorar o modelo compartilhado, através do algoritmo de média federada (*Federated Average* - FedAvg) [12]. Após a geração no modelo mais atual, novamente o servidor central envia para os clientes a última atualização. Assim, o aprendizado federado proporciona a geração de modelos mais eficientes com menor latência, pois é possível a utilização imediata do modelo no próprio dispositivo. O aprendizado federado é fortemente baseado em mecanis-

mos de aprendizado de máquina treinados e otimizados através do método do gradiente descendente estocástico (*Stochastic Gradient Descent* - SGD). Atualmente, as principais implementações do aprendizado federado dependem da ponderação da contribuição local de diferentes clientes para definição do fluxo de otimização do modelo global através do algoritmo FedAvg.

A Tabela 2.2 apresenta de forma sintetizada as principais características de cada modelo. Destaca-se que para cada um, é necessário realizar uma ponderação entre custo computacional e desempenho.

Tabela 2.2: Comparativo entre os principais modelos de aprendizado de máquinas usados para a detecção de anomalias e predição de tráfego em redes 5G.

Modelo	**Finalidade**	**Aplicação**	**Vantagens**	**Desvantagens**
Autoencoders	Aprendizado de representações e compactação	Detecção de anomalias	Análise de dados sequenciais	Alto custo computacional
CNN	Modelagem de dados espaciais	Detecção de anomalias	Reconhecimento de padrões	Dificuldade de parametrização
LSTM	Análise de dados sequenciais	Predição de tráfego	Modelar relações de longo prazo	Alto consumo de memória
OCSVM	Classificador	Detecção de anomalias	Eficaz em espaços multidimensionais	Tempo elevado de treinamento
RNN	Análise de dados sequenciais	Predição de tráfego	Captura dependência temporal	Alto fluxo de dados para treinamento

2.4 Modelos estatísticos para detecção de anomalias e predição de tráfego

Um dos maiores desafios para as redes de próxima geração será o gerenciamento de rede devido à alta complexidade de dispositivos interconectados e roteamento altamente dinâmico. O ITU trabalha na modernização das recomendações relativas à qualidade de serviço (QoS) e qualidade de experiência dos usuários (QoE). A recomendação ITU-T Y.3172[1] prevê a introdução de mecanismos de aprendizado de máquina para o gerenciamento e a orquestração de funcionalidades nas próximas gerações de redes. Os modelos estatísticos baseados em séries temporais são os métodos clássicos com desempenho satisfatórios para realizar predições [41], sobretudo em fluxos de rede, por possuírem capacidade analítica suficiente e implementação com baixo custo computacional. A previsão de séries é um campo essencial do aprendizado de máquina aplicado a redes 5G [42]. A modelagem de

[1]Disponível em https://www.itu.int/rec/T-REC-Y.3172-201906-I/en.

séries temporais é uma área de pesquisa com abrangência em diversos campos científicos e diversos modelos de predição de séries temporais evoluíram ao longo do tempo.

As séries temporais são representações matemáticas de fenômenos que ocorrem continuamente durante um intervalo de tempo. São divididas em três componentes: tendência, sazonalidade e irregularidade. A componente de **tendência** refere-se a uma perspectiva de longo prazo, a **sazonalidade** diz respeito a eventos sistêmicos associados à um período recorrente e, por fim, as **irregularidades** são oscilações não sistemáticas de curta duração [30]. Existem objetivos basilares para análise de uma série, sendo eles a cognição do mecanismo gerador da série e a predição de pontos futuros. No que diz respeito ao mecanismo gerador da série, é fundamental identificar o comportamento, isto é, descrever se existem ciclos, tendências ou sazonalidade e pontos de periodicidade relevantes. Com base nisso, a predição do comportamento é possível. Cabe ressaltar que a escolha do melhor método e seus respectivos parâmetros para uma dada série, tem como objetivo redução de erros de predição, pois estimar o futuro envolve incertezas.

As próximas seções descreverão os principais modelos estatísticos utilizados em diversos campos de estudo, principalmente na predição de séries temporais.

2.4.1 *Auto-Regressive Integrated Moving Average* – ARIMA

O modelo ARIMA é um dos mais populares para realizar predição utilizando séries temporais. Inicialmente proposto por George Box e Gwilym Jenkins, também conhecido como método Box-Jenkins é um modelo aplicado em casos de não estacionariedade. Existem ainda variações do modelo como o VARIMA(*Vector Auto-Regressive Integrated Moving Average*), utilizado para múltiplas séries temporais e o SARIMA (*Seasonal Auto-Regressive Integrated Moving Average*), utilizado em situações que existam possíveis sazonalidades nos pontos da série. Todos esses modelos possuem bom desempenho para análises de curto prazo, enquanto o modelo SARIMA possui a melhor capacidade de análise a longo prazo. A estrutura do ARIMA é composta por três coeficientes, sendo o primeiro denominado auto-regressivo p, seguido do coeficiente de diferenciação d e por último o coeficiente de médias móveis da série q. O modelo ARIMA [43] é formulado por:

$$y'_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i y'_{t-i} + \varepsilon_t + \sum_{i=1}^{q} \beta_i \varepsilon_{t-i}, \tag{2.3}$$

em que o coeficiente α_i refere-se ao termo auto-regressivo da série, β_i relaciona-se à média móvel e ε_t corresponde à parte residual do modelo. As principais etapas para utilização

do modelo ARIMA podem ser descritas em três estágios [43]:

1. **Pré-processamento na série**. O pré-processamento é realizado através do teste *Augmented Dickey–Fuller* (ADF) para identificar se a série é estacionária. Em caso negativo, diferenciações são realizadas na série, quantas vezes forem necessárias, até obter a estacionariedade. O número de diferenciações é caracterizado através do parâmetro d;

2. **Cálculo dos valores da função de autocorrelação amostral (ACF) e autocorrelação parcial (PACF)**. O cálculo dos valores das funções ACF e PACF é feito para a série estacionária obtida, determinando os parâmetros p e q respectivamente. Para fins de desempenho, esses parâmetros podem ser obtidos através da análise da métrica *Akaike Information Criterion* (AIC), no qual o objetivo é mensurar a qualidade relativa de um modelo estatístico;

3. **Teste do modelo e realização de predições**. Por fim, são realizados testes no modelo que apresenta o melhor desempenho e as predições da série são realizadas. A avaliação do modelo pode ser realizada através da relação entre os valores preditos e observados. Métricas calculadas pelo *Root Mean Square Error* (RMSE) são comumente utilizadas para mensurar a qualidade de um modelo preditivo.

2.4.2 *Seasonal Auto-Regressive Integrated Moving Average* – SARIMA

Uma das variações do ARIMA é o modelo SARIMA. Esse modelo tem o propósito de realizar análises mais apuradas em séries com características predominantes de sazonalidade e periodicidade, podendo ser útil em predições de tráfego de redes sem fio [44] e detecção de anomalias em redes [45]. Por ser uma variação do modelo ARIMA, o SARIMA pode ser representado por SARIMA$(p, d, q)(P, D, Q)_s$. A primeira parte do modelo, representada pelos parâmetros p, d e q é não sazonal, enquanto a segunda parte é sazonal e constitui o fator de sazonalidade. Os parâmetros P, D e Q representam respectivamente o número dos termos de sazonalidade da parte auto-regressiva, o número de diferenciações sazonais e a parte sazonal de médias móveis. O fator de sazonalidade contribui para analisar características como uso de banda, que tendem a ter comportamentos cíclicos. Hanbanchong e Piromsopa utilizam o modelo SARIMA para detectar anomalias predizendo o uso de banda através da sazonalidade existente [46]. Em diversos outros campos de estudo, no qual a série temporal é utilizada como fonte de análise para estabelecer pontos futuros,

o modelo SARIMA é amplamente utilizado. Condições climáticas, predição de carga energética e propagação de doenças infecciosas são temas de estudo que frequentemente utilizam este modelo.

2.4.3 Classificador Bayesiano

O classificador Naïve Bayes tem sua origem no **Teorema de Bayes** onde as probabilidades de evento estão condicionadas à probabilidade de hipótese com resultados previamente conhecidos. Seja um conjunto de dados $X = (x_1, y_1)$,...,(X_N, y_n) com x amostras e y classes correlatas para um problema de classificação, sendo $x \in \mathbb{R}$ e $y \in [1, K]$ [30], o Teorema de Bayes é descrito por:

$$P(y = i|x) = \frac{P(i) * P(x|i)}{P(x)}, \tag{2.4}$$

em que $p(i)$ é a probabilidade de uma hipótese ser verdadeira a partir da amostra de uma classe e $p(y|x)$ é a distribuição de probabilidades desconhecidas no espaço amostral x.

É regularmente utilizado para dados de alta dimensionalidade [47]. A probabilidade condicional é utilizada para predição de ataques e tráfegos regulares, podendo ser utilizada na detecção de ataques em redes definidas por software [48]. Em redes *Ad-Hoc*, o classificador Naïve Bayes é utilizado como parte de um modelo de detecção de ataques de Negação de Serviço Distribuído [49]. Nesse caso, os autores aplicam o classificador para decompor o tráfego de rede em dois padrões, normal e ataques DDoS. Assim, consideram cinco características do tráfego para determinar qual padrão um fluxo pertence, sendo eles o tamanho do pacote, a porta, o IP de origem, o IP de destino e a variação do atraso (*jitter*). O classificador também é utilizado para classificação de textos, sendo possível detectar anomalias inseridas em cargas útéis (*payload*) do protocolo HTTP conforme apresentando em [50]. Essa detecção é importante porque uma das principais formas de ataque HTTP ocorre através da modificação do *payload* do pacote.

2.4.4 Hidden Markov Model – HMM

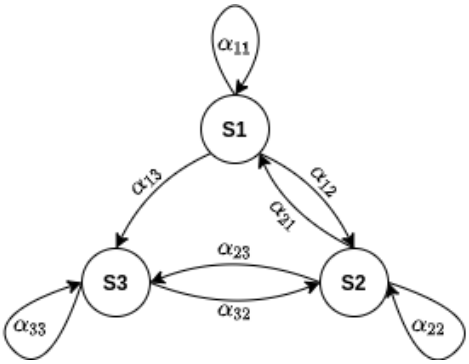
Processos estocásticos são definidos através de variáveis aleatórias, as quais representam características determinísticas em um intervalo de tempo t. Os processos estocásticos são utilizados para analisar o comportamento de sistemas em que o grau de incerteza é consideravelmente alto. Esses processos podem ser classificados em dois cenários, em relação

ao estado e ao tempo, com características discretas e contínuas para cada um dos cenários. Um processo estocástico é considerado *markoviano* se a probabilidade condicional de um estado futuro depende apenas do estado presente e não dos estados anteriores. Essa probabilidade pode ser descrita como probabilidade de transição e é expressa matematicamente por:

$$P(q_{t+i} = S_j | q_t = S_i). \tag{2.5}$$

A Equação 2.5 representa a probabilidade do estado q_{t+1} ser S_j no momento $t+1$ dado que o estado q_t é igual a S_i no instante t.

Um processo markoviano é classificado como **Cadeia de Markov** se as variáveis aleatórias são definidas em um espaço de estados discretos. A Cadeia de Markov representa sistemas que podem a qualquer instante de tempo t estar em um dado estado S. A mudança entre um estado e outro ocorre através de uma matriz de transição, que descreve as probabilidades de o sistema mudar do estado S_0 para o estado S_n. A Figura 2.9 mostra uma Cadeia de Markov com 3 estados e as probabilidades de transição entre esses estados, representadas por $\alpha_{i,j}$.



(a) Diagrama das probabilidades de transição

$$S = \begin{pmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ 0 & a_{3,2} & a_{3,3} \end{pmatrix}$$

(b) Matriz de probabilidade de transição

Figura 2.9: (a) Exemplo de representação de uma cadeia de Markov com probabilidades de transição entre três estados. Os estados são representados pelas variáveis s_i e as probabilidades de transição entre estados, pelas variáveis α_{ij}.(b) Matriz de transisão de probabilidades do estado S.

O conjunto de probabilidades da matriz de transição de uma Cadeia de Markov é caracterizado pela Equação 2.6 e deve obedecer às propriedades das Equações 2.7 e 2.8.

$$a_{ij} = P(s_{t+i} = S_j | s_t = S_i) \qquad (2.6)$$

$$a_{ij} \geq 0 \qquad (2.7) \qquad \sum_{i=1}^{N} a_{ij} = 1 \qquad (2.8)$$

O **Modelo Oculto de Markov** (*Hidden Markov Model* - HMM), em sua essência, é a variação de um processo estocástico Markoviano. O HMM é caracterizado por duas componentes, uma não observável e outra observável. A primeira componente representa o estado de um sistema previamente modelado, enquanto a segunda representa as observações já realizadas. Os processos não observáveis representam um conjunto de estados interligados através da matriz de probabilidades, enquanto os processos observáveis representam as saídas de cada estado. Como exemplo, alertas de um sistema de detecção de intrusão - (IDS) [51] podem ser caracterizados como um processo estocástico observável em um modelo oculto de Markov, no qual a sequência de observações representa os alertas e a sequência de estados ocultos representam o estado do evento de segurança [52]. O modelo é representado de forma reduzida através de uma tupla com três elementos, (A, B, π), em que A representa a matriz de transição, B a distribuição de probabilidades das observações e π o vetor de probabilidade inicial. Di Bernardino e Brogi mostram que o modelo também pode ser representado com parâmetros adicionais [53], da seguinte forma:

1. Sendo N o número de estados do sistema, o conjunto de estados descritos individualmente é dado por

$$S = \{S_1, S_2, ..., S_N\}; \qquad (2.9)$$

2. Existe um número M de observações realizadas, cujo conjunto é dado por

$$O = \{O_1, O_2, ..., O_M\}; \qquad (2.10)$$

3. A transição entre estados é dada pela matriz de transição de probabilidades A que possui dimensão $N \times N$ e é definida por $A = [a_{ij}]$, cujos elementos são dados por

$$a_{ij} = P(s_{t+i} = S_j | s_t = S_i), \qquad 1 \leq i, \quad j \leq N; \qquad (2.11)$$

4. A matriz de probabilidades de observações $B = [b_{ij}]$ possui dimensão $N \times M$ e os elementos são descritos através de

$$b_{ij} = P(o_t = O_j | s_t = Si), \qquad 1 \leq i \leq N, \quad 1 \leq j \leq M; \qquad (2.12)$$

5. O vetor de probabilidade inicial é definido por

$$\pi_i = P(s_1 = S_i), \qquad 1 \leq i \leq N. \tag{2.13}$$

No HMM existem dois tipos principais de estruturas, classificadas como ergódica, ou sem restrições, e esquerda-direita (*left-right*). No modelo com estrutura ergódica, cada estado pode transitar entre quaisquer outros, sendo esse modelo completamente conectado. O modelo com estrutura esquerda-direita não permite transições entre um estado e estados anteriores [51], sendo mais relevante para detecção de ataques, principalmente os que possuem diversas etapas antes de atingirem o objetivo.

A Tabela 2.3 apresenta os modelos estatísticos descritos nas seções anteriores, apresentando as principais características, finalidades, pontos positivos e negativos de cada um. É importante frisar, que ambos os modelos com a finalidade de predição de tráfego, possuem custo computacional relativamente altos, referente principalmente a grande quantidade de dados necessários para treinamento dos modelos.

Tabela 2.3: Comparativo entre os principais modelos de aprendizado de máquinas usados para a detecção de anomalias e predição de tráfego em redes 5G.

Modelo	**Finalidade**	**Aplicação**	**Vantagens**	**Desvantagens**
ARIMA	Análise de séries temporais	Predição de tráfego	Análise de curto prazo	Custo computacional
SARIMA	Análise de séries temporais com sazonalidade	Predição de tráfego	Captura dependência entre dados consecutivos	Custo computacional
HMM	Predição de estados	Detecção de anomalias	Modelar relações de longo prazo	Tempo de treinamento
Bayesiano	Classificador	Detecção de anomalias	Eficaz para múltiplas classes	Presume independência das características

[1]Informações deste capítulo foram extraídas do texto: Segurança em Redes 5G: Oportunidades e Desafios em Detecção de Anomalias e Predição de Tráfego baseadas em Aprendizado de Máquina

Capítulo 3

Trabalhos Relacionados

3.1 Predição de tráfego em redes de computadores

Sivanathan *et al.* focam em identificar dispositivos IoT em uma rede através das características de tráfego geradas por cada dispositivo e desenvolvem uma estrutura para classificação destes utilizando características de tráfego obtidas na camada de rede [54]. Na proposta, 28 dispositivos inteligentes são utilizados, entre câmeras, luzes, tomadas, sensores de movimento entre outros. O tráfego de todos os dispositivos é coletado por um período de seis meses. A análise baseia-se em inferências estatísticas, aplicando as seguintes métricas para a caracterização dos dispositivos: volume, duração e taxa média do fluxo, tempo de hibernação, número das portas, endereços de consultas DNS, intervalo das consultas NTP (*Network Time Protocol*) e conjuntos de cifras do *handshaking* TLS (*Transport Layer Security*). Uma medida de custo é realizada para obtenção de cada uma das variáveis, de acordo com necessidade de processamento dessas medidas. O custo é então classificado em baixo, médio ou alto. O trabalho propõe ainda uma métrica de relevância dos atributos, no qual o impacto de cada variável no resultado da classificação é considerado. Assim, é possível realizar uma escolha de quais variáveis utilizar para otimizar a implementação (*online*) sem comprometer o desempenho da classificação.

Sciancalepore *et al.* realiazam estudo acerca da predição de tráfego em redes móveis, para tratar do paradigma de divisão de recursos de infraestrutura de rede. Inicialmente como motivação, os autores abordam o a questão da tecnologia 5G, que irá impulsionar o que eles chamam de fatiamento da rede para multiplos inquilinos. Do ponto de vista comercial, isto pode ser interessante para operadoras de rádiofrequência, em função da otimização dos recursos, visto que poderão alocar mais clientes de diferentes modelos para compartilhar a mesma infraaestrutura. A ideia central é realizar uma política de novas

demandas de divisão da rede baseado em *Service Level Agreement* (SLA). Os autores comentam que a divisão de rede permite que operadoras compartilhem sua infraestrutura de rede física para a criação simultânea de diversas redes lógicas independentes, gerenciadas de maneiras diferentes. Esses segmentos de rede são alocados de forma temporária para os respectivos locatários. Esse conceito proporciona novas oportunidades de geração de receita da infraestrutura de rede, uma vez que uma maior utilização da capacidade da infraestrutura pode ser alcançada admitindo mais pedidos de alocação. Os autores descrevem que algoritmos de controle de admissão bem estruturados impulsionam a multiplexação de tráfego entre as divisões de redes realizadas e são fatores determinantes para otimização dos recursos. Os autores propõem um arcabouço composto por três módulos, sendo o primeiro módulo de predição, um segundo módulo para controle de admissão e um terceiro para provisionamento. O módulo de predição é desenvolvido para avaliar as demandas de cada locatário e, inicialmente, aborda que as requisições podem ter um paradigma global, ou seja, são distribuídas uniformemente por toda a rede. Outra abordagem é o fator de mobilidade, no qual as requisições que podem ser afetadas em ambientes com células adjacentes. Neste primeiro cenário, os autores utilizam o modelo de Holt-Winters para analisar a predizer as requisições.

Yue *et al.* propuseram uma estrutura baseada no algoritmo *Random Forest* para realizar a predição de largura de banda de redes LTE em tempo real [55]. Segundo os autores, os métodos convencionais de aprendizado de máquina ou estatística são baseados no histórico de sequências curtas, não sendo fácil identificar padrões temporais em estruturas de dados mais complexas. Segundo os autores, uma sequência de dados históricos mais longa não apenas pode ajudar na predição, mas pode diminuir o desempenho.

Feng *et al.* avaliam a predição de tráfego em redes móveis utilizando redes neurais como base do estudo. Inicialmente os autores citam modelos amplamente já abordados como o SARIMA e *Support Vector Regression* (SVR) [56]. A proposta do trabalho está no desenvolvimento de uma estrutura denominada DeepTP que consiste na extração de características da rede para modelar as dependências espaciais existentes no tráfego da rede móvel. Um outro módulo deste framework é responsável por selecionar características a partir de outras torres de celular próximas, seleciona características de tráfego de outras torres de celular pelo tráfego de uma determinada célula para representar sua influência. Primordialmente, esse módulo produz dois tipos de correlação dos recursos de tráfego: a correlação positiva, que representa a influência de torres de celular adjacentes com padrão de tráfego semelhante, e a negativa, que representa a influência do tráfego oposto. Por fim, baseado nestas características é utilizada uma RNN baseada em LSTM para realizar

a predição. Existem desafios para a predição de tráfego em redes móveis, em função das particularidades que este meio possuí, principalmente no que diz respeito à mobilidade e a correlação espacial.

O volume de tráfego em cada estação rádio-base, possuí características distintas umas das outras. Diversos fatores compõem essa heterogeneidade no tráfego, como a localização e os hábitos de onde a estação está localizada. Estações em áreas de entretenimento por exemplo, podem oscilar rapidamente em períodos específicos durante noite, enquanto estações localizadas nas próximas das de serviços de transporte podem ter picos sazonais durante o dia, principalmente nas horas de maior movimentação das pessoas. O modelo proposto pelos autores, é implementado através de dois módulos sendo eles o extrator e o sequencial. O primeiro é responsável pela geração das características do tráfego, no que diz respeito a parte temporal e espacial, enquanto o segundo trata da modelagem que irá definira a relação sequencial entre as características extraídas. A extração espacial tem como resultado duas características, sendo a positiva e negativa. A positiva reflete a influência de áreas com comportamento similar em relação à área avaliada, enquanto a negativa são as áreas que possuem padrões opostos. O extrator da característica espacial calcula inicialmente a correlação entre o tráfego da área avaliada e o tráfego de outras áreas, implementado através de rede neural *feed-forward*.

Frank *et al.* propõem um modelo de predição com foco na integração de serviços de cidades inteligentes, baseado na análise de dados de redes sem fio [57]. Os autores utilizam uma rede neural denominada *Multilayer Perceptron* para prever o número de usuários conectados em uma área geográfica e assim otimizar a alocação de largura de banda. Foi utilizado um método heurístico denominado PSO (*Particle Swarm Optimization*) para melhoria dos parâmetros de treinamento da rede neural MLP. Os autores focam na otimização do número de neurônios da rede para consequentemente aprimorar a taxa de aprendizado. Para validação da proposta, os autores verificaram o ganho em relação à economia de largura de banda total, baseado no número de usuários conectados em um dado instante de tempo.

Mei *et al.* abordam a predição em tempo real da largura de banda e do *handoff* entre as redes 4G e 5G [58]. É aplicado um modelo de Rede Neural Recorrente para explorar os padrões temporais e de evolução da largura de banda em cenários de mobilidade fixa, ou seja, que possuem um padrão de repetição. São propostos dois modelos de predição de *handoff* sendo a primeira uma predição binária, prevendo se o dispositivo fará a mudança da rede 4G para 5G e vice-versa e a predição contínua, no qual é realizada a predição da

probabilidade do acesso à rede 5G em uma futura janela de tempo curta. Os modelos de predição são baseados em classificação e regressão, atingindo mais de 80% de precisão na previsão de *handoffs* entre as redes 4G e 5G.

Benslimen *et al.* propõem um arcabouço utilizando o modelo *Autoregressive Integrated Moving Average* (ARIMA) para predizer ataques e o modelo LSTM para predizer anomalias e falhas[59]. Para realizar a predição, os autores utilizam métricas tradicionais dos recursos computacionais, tais como uso de memória, uso de processamento, e tráfego de entrada e saída de rede por exemplo.

Wang *et. al.* sugerem técnicas de aprendizado profundo para caracterizar a relação espaço-tempo na predição de redes móveis [60]. Para isso, empregam uma arquitetura baseada em codificadores automáticos (*autoenconder*) e LSTM para avaliar a correlação espaço-tempo conforme distribuição do tráfego na rede. Os autores utilizam o codificador automático para modelar e extrair características espaciais e a rede neural LSTM para modelagem temporal, facilitando assim a escalabilidade dos recursos do núcleo.

Alawe *et al.* examinam a escalabilidade dos recursos do núcleo da rede 5G, composto principalmente por SDN e NFV. O plano de controle no núcleo da rede 5G estabelece a Função de Acessibilidade e Mobilidade *Access and Mobility Function* (AMF) que atua diretamente nas requisições de conexão dos clientes. Assim, a AMF representa um gargalo no plano de controle de redes móveis [26]. São comparados dois modelos de redes neurais, DNN e LSTM, tendo o último um desempenho superior.

Nie *et al.* é predição de tráfego em redes de malha utilizando Rede de Crença Profunda *Deep Belief Network* (DBN) [61]. No primeiro momento, os autores utilizam a transformada discreta *wavelet* para extração das componentes de baixa frequência do tráfego de rede, que caracterizam as relações de longo prazo, adotando em seguida a DBN para efetuar a predição dessa componente. Já as componentes de altas frequências, representam flutuações dispares no tráfego e os autores utilizam um modelo gaussiano para caracterizá-las, estimando os parâmetros através da Máxima Verossimilhança.

3.2 Predição de ataques

Wang *et al.* apresentam o desenvolvimento de um framework para detecção e mitigação de ataques *Link-Flooding* em redes definidas por *software* [62]. Ataques desse tipo possuem uma característica de difícil detecção, pois o tráfego malicioso se assemelha muito com tráfego legítimo. Este ataque tem como objetivo saturar os links de comunicação, de

forma a degradar os possíveis caminhos que levam a um determinado serviço. A proposta consiste em três módulos, sendo um para detectar o *Link-Flooding*, o segundo para avaliar se o ataque é ou não *Link-Flooding* ou simplesmente um tráfego atípico e o terceiro um mecanismo para mitigar o impacto, realizando engenharia de tráfego para balancear o tráfego através dos nós.

Bartos *et al.* abordam o tema de alertas de priorização de estado e correlação de eventos como tarefas desafiadoras [63]. Os autores afirmam que a probabilidade de que o ataque recentemente descoberto ocorra novamente em um curto período é alta. Os autores propõem um aprendizado de máquina para estimar a probabilidade de que uma entidade, ou seja, um *host* ou uma rede, possam se tornar uma fonte de ataque. Para este fim, os autores criam uma pontuação denominada *Future Misbehavior Probability*-(FMP). O objetivo é introduzir alguns conhecimentos, desde diferentes fontes, sobre uma entidade, rede ou host específico. A pontuação representa o comportamento esperado de uma dessas entidades, com base no aprendizado de máquina e, em seguida, atribui um valor a ele, que prevê eventos futuros. Nesse cenário, os autores analisam dois modelos: Redes Neurais e *Gradient Boosted Decision Tree* (GBDT). Inicialmente é comparado os valores *Brier Score*, que é um fator para medir a precisão das previsões probabilísticas. Tanto as redes neurais quanto o GBDT realizaram bem, atingindo valores de *Brier Score*, próximos a zero. O GBDT, no entanto, teve um desempenho melhor.

Bilal *et al.* propõem um modelo para prever os recursos necessários no instante ideal de tempo, para manter a elasticidade em uma rede com funções virtualizadas [64]. A capacidade de diferentes funções desta rede aumenta ou diminui, com intuito de otimizar a utilização de CPU. Desta forma, é apresentado duas abordagens para prever a utilização da CPU no dia seguinte. A primeira é um modelo de agendamento *offline* que permite ajustar a elasticidade em redes virtualizadas, prevendo eventos em dias normais. O segundo, é baseado em uma abordagem de agendamento que prevê a utilização de CPU no dia seguinte durante momentos transientes, devido a algumas circunstâncias atípicas. É proposto então, um algoritmo que utiliza ambas as estratégias para lidar de forma eficiente com a elasticidade em redes virtualizadas. Baseado nos padrões de utilização de recursos do plano de controle e dados, em escala diária, foram utilizados modelos de séries temporais para prever a carga de uso um dia antes, e para avaliar a tendência de transição de cargas.

Dias *et al.* autores analisam o problema de anomalias para centro de dados e dissertam sobre detectá-las em tempo real, propondo um algoritmo que seja simples e ao

mesmo tempo robusto. A análise desse problema tem como principal motivação outros algoritmos de detecção de anomalias, que geralmente precisam de treinamento prévio com uma quantidade considerável de informação, para que os algoritmos possam descrever um possível comportamento anômalo. Esta proposta consiste em um algoritmo baseado em score, para que uma dada sequência de valores em uma série temporal, receba uma pontuação e baseado neste valor, seja possível realizar a detecção de uma anomalia. O algoritmo inicialmente executa a discretizarão dos valores de uma característica da série, por exemplo, percentual da utilização de CPU, transformando valores escalares em valores em uma faixa entre que zero e teta (variável definida seguindo parâmetros pré-definidos). A partir desta discretização, a métrica sequência é definida, onde o seu comprimento também é pré-definido. Neste caso os autores utilizam dois como parâmetro. A sequência é composta pelo atual valor discretizado e o valor anterior, formando assim um vetor com dimensão igual a dois. O algoritmo se mostrou mais eficientes do que os outros do *framework* em todas as etapas exceto no cálculo da pontuação [65].

Tartakovsky *et al.* o problema de detecção de anomalias para redes de computadores em tempo real, demonstrando a dificuldade de realizar esta tarefa com métodos tradicionais. Para isto, citam como exemplos o SPRT *(Sequential Probablity Ratio Test)*, CUSUM *(Cumulative Sum)* e EWMA *(Exponentilly Weighted Moving Average)*. Os últimos dois métodos, são originários da área de análise estística denomiada *sequential changepoint detection*, que possuem o objetivo de desenvolver e analisar de forma ágil a detecção de uma anomlia no momento do evento. O método escolhido para realizar o estudo é o Shiryaev–Roberts, e faz parte da técnica do tipo *changepoint*. Os autores o descrevem como sendo pouco conhecido da comunidade de segurança, possuí um custo computacional tão baixo quanto o CUSUM e o EWMA, porém apresenta um modelo ótimo para determinadas configurações multi-cíclicas. Os autores descrevem o estudo acerca do *Quickest changepoint detection*, que se trata de técnicas para detectar a mudança de estado, geralmente entre normal e anormal. A configuração sequencial assume que a série é acumulada uma por vez, e que comportamento dos dados está em um estado normal. No caso de haver uma mudança para o estado anormal, o objetivo é detectar esta mudança o mais rápido possível. Para validar a proposta, os autores realizaram o teste utilizando o algoritmo em um ataque real do tipo DDoS, conhecido como *SYN flood*. O foco deste ataque é congestionar o link da vítima com requisições do tipo *SYN* até estressar por completo os recursos do alvo, de modo que ele possa não mais responder requisições legítimas. Os autores apresentam uma imagem do dataset, no qual são exibidas o número de tentativas de conexões, e embora seja possível visualizar graficamente o ataque, é impossível

descrever o seu momento exato de início.

Os trabalhos apresentados neste capítulo utilizam métricas tradicionais para realizar a predição de tráfego e anomalias, tais como taxa de transmissão, uso de memória e processamento entre outros. A proposta deste trabalho tem como foco o aproveitamento de características categóricas do fluxo de rede para realizar a predição, aplicando a entropia de Shannon e gerando novas métricas. O aproveitamento destas características tais como IP de origem e destino, mostra-se promissor em função das mesmas serem elementaresm em qualquer fluxo de rede, utilizadas na maioria dos casos para classificação e criação de pontuação em *blacklists* por exemplo.

Capítulo 4

Predição de tráfego em redes sem fio de larga escala

O problema de predição de tráfego de redes está relacionado à modelagem do volume de tráfego entre seus nós. A previsão de tráfego visa antecipar a caracterização do fluxo de rede que acontecerá no futuro [35]. Além de prever o volume de tráfego, o problema de predição também aborda a questão de classificação de protocolos e predição de distribuição de protocolos. O problema de classificação consiste em uma série de tipos de protocolos, determinando quais devem aparecer na rede em etapas futuras. Uma extensão do problema envolve a previsão da distribuição de características dos pacotes. No entanto, a predição de dados trafegados pela rede depende majoritariamente da natureza estatística dos dados e da dependência temporal. A autossimilaridade e a característica primordialmente não linear dos dados, possuem propriedades estatísticas que dificultam particularmente a predição. Distribuições como Poisson ou Gaussianas modelam insuficientemente a natureza não linear dos dados. Além disso, do ponto de vista da interdependência, o tráfego de rede é caracterizado por autocorrelação de longo prazo, a qual a maioria dos modelos estatísticos não consegue capturar [35].

O padrão IEEE 802.11, popularmente conhecido como WiFi, é uma das principais tecnologias para funcionamento da IoT [66]. As tecnologias de comunicação sem fio são amplamente implantadas e abragem áreas extensas com grande número de usuários, como os *campi* universitários [67]. Assim, é fundamental estudar o crescimento dos dados móveis devido ao seu impacto no gerenciamento e segurança das redes, principalmente quando a densidade de uso aumenta. A análise de tráfego é uma ferramenta crucial para definir metas de gerenciamento e projetos para novas infraestruturas de redes de próxima geração.

Pesquisas anteriores utilizam o modelo matemático autorregressivo integrado de médias móveis (ARIMA) para predizer o crescimento do tráfego. O modelo normalmente é aplicado para análise *offline* e, portanto, a complexidade do algoritmo não é crítica. Entretanto, em um cenário de predição *online*, isolar o tráfego em componentes, como tendências, rajadas e ruídos, deve seguir uma abordagem estatística para viabilidade e, então, cada componente fará parte da predição separadamente. As redes neurais também são utilizadas para predição em tempo real do tráfego de rede, pois a rede neural artificial encontra padrões complexos nos dados recebidos. A transformada *wavelet* é adequada para predição multidimensional, pois transforma naturalmente um sinal em várias resoluções. A aplicação da transformada *wavelet* permite revelar tendências locais detalhadas descritos por Tant [68].

4.1 Proposta de mecanismo híbrido de predição em redes de larga escala

O mecanismo proposto aplica a Transformada *Wavelet* Discreta para extrair componentes lineares e não lineares da série e para cada componente são aplicadas as as técnicas ARIMA e LSTM. Cada modelo fornece a predição de cada componente do sinal, assim, uma das métricas para comparação da eficiência de cada modelo é o RMSE de cada característica. O RMSE é apresentado na Equação 4.1 e representa a medida do desvio médio entre os valores observados e previstos. A predição final de cada característica é a composição da melhor predição da componente linear e a melhor predição da componente não linear. A caracterização do fluxo foi baseada utilizando as 5 tuplas do NetFlow: IP de origem, IP de destino, porta de origem, porta de destino e tipo de protocolo de transporte. Estas características foram utilizadas, pois na mairoria dos modelos preditivos são utilizadas caractéristicas não categóricas, como *bytes*, número de pacotes, entre outros. São executadas duas atividades principais que compõem a predição do comportamento do tráfego. A primeira separa os dados coletados em fluxos fracionados que representam a coleta de um dia para treinamento dos modelos. A segunda calcula a entropia de cada característica da tupla para a janela de tempo de uma hora. Por fim, o processamento do fluxo é dividido em três etapas principais sendo detalhada na Figura 4.1. As principais etapas deste modelo são aplicação da entropia de Shannon, aplicação da transformada *wavelet* e aplicação dos modelos preditivos, sendo apresentadas a seguir.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(predction_i - actual_i)^2}. \tag{4.1}$$

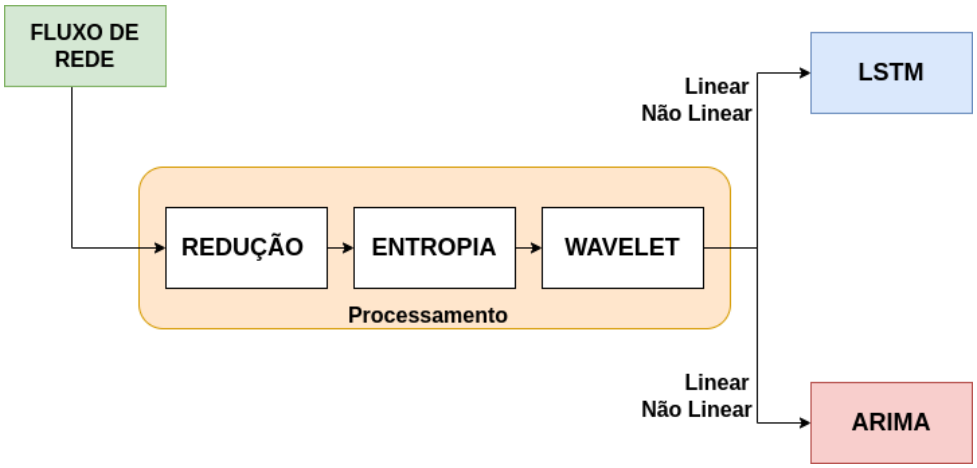


Figura 4.1: Proposta da arquitetura para predição de tráfego e detecção de anomalias

- **Cálculo da Entropia de Shannon**: A entropia da informação, também conhecida como Entropia de Shannon, é uma medida para analisar o grau de incerteza ou concentração da distribuição da informação. Originalmente, a entropia foi descrita como uma medida em sistemas termodinâmicos, mas Claude Shannon estendeu-a à teoria da informação em 1948. O conceito se aplica à predição de tráfego de rede porque o tráfego possuí características essencialmente aleatórias, como o IP de destino e a porta de origem. A entropia é expressa matematicamente como:

$$H(X) = -\sum_{i=1}^{n}(p_i)\log(p_i), \tag{4.2}$$

  no qual p_i é a probabilidade of i-enésimo termo da variável x.

  Quanto maior a entropia de uma característica maior a incerteza e consequentemente menos previsível se torna.

- **Aplicação da Transformada Wavelet**: A função *wavelet* decompõe os sinais no domínio da frequência sendo útil para o processamento de sinais no domínio do tempo. Chang *et al.* afirmam que a função é um método eficaz de análise tempo-frequência após a análise de Fourier [69]. Existem dois tipos principais de transformada *wavelet*, a *Discrete Wavelet Transform* (DWT) e a *Continuous Wavelet Transform* (CWT). A decomposição *wavelet* extrai componentes de baixa e alta

frequência pois ambas produzem de maneira satisfatória uma análise local da série em ambos os domínios. As componentes extraídas são a Componente Detalhada - (CD) e a Componente Aproximada (CA). CD é responsável pela geração de componentes lineares, enquanto CA gera componentes não lineares. As componentes extraídas podem ter informações que tornam as predições mais distintas. Alguns modelos de predição possuem singularidades que os tornam mais eficientes, dependendo de como o sinal é processado. O DWT geralmente fornece uma ferramenta rápida para remover o ruído de um sinal. Considerando um número limitado de coeficientes das componentes DWT, é possível realizar uma transformada inversa, obtendo um sinal com ruído reduzido. A técnica é útil na análise do tráfego de rede pois do ponto de vista da detecção de anomalias, o ruído pode representar um ataque ou tráfego de fundo.

- **Aplicação dos modelos preditivos** : Os modelos aplicados para realização da predição e validação são o ARIMA*(p,d,q)* e LSTM, baseado na Rede Neural Recorrente (RNN), sendo este uma técnica útil para detecção de anomalias, sistema de detecção de intrusão (IDS) ou outras aplicações de processamento de sinais, como reconhecimento de fala.

4.2 Conjunto de dados avaliado

O mecanismo proposto é avaliado utilizando um conjunto de dados com a captura de tráfego de rede contendo informações reais, coletado a partir da rede sem fio da Universidade Federal Fluminense através do protocolo NetFlow. Esta rede compreende uma infraestrutura com mais de 500 pontos de acesso atingindo mais de 5000 mil usuários simultâneos conectados diretamente [70]. Esta seção descreve o conjunto de dados utilizado e a proposta de predição de tráfego e detecção de anomalias através da análise de características categóricas utilizando entropia como medida de predição.

O conjunto de dados contém tráfego de uma semana da rede sem fio de um único campus da universidade, a Praia Vermelha. A coleta de dados foi realizada entre os dias 17 e 26 de abril de 2018 e gerou um arquivo com tamanho de 16 *Gigabytes*. Foi utilizado a linguagem Python para processar os dados e implantar modelos de predição em um computador equipado com processador Intel Quad Core i5 8265U com 1,60GHz, 8GB de RAM e armazenamento em disco SSD de 256GB utilizando Ubuntu 18.04 como sistema operacional.

4.3 Avaliação e resultados

A primeira etapa da avaliação é o processamento do conjunto de dados para cálculo da entropia de Shannon. Foram calculadas as entropias para uma janela com duração de uma hora para cada dia. Com isso, foram geradas amostras para o período analisado com 168 pontos. Em seguida, na segunda etapa da proposta, foi aplicada a transformada *wavelet* discreta - DWT, utilizando a *wavelet* de Haar por possuir tempo de execução menor em comparação com a *wavelet* de Daubechies [71]. A Figura 4.2 apresenta a série temporal original para a entropia IP de origem comparada com as componentes de sinal reconstruídas lineares e não lineares, após a aplicação da transformada inversa. Para a característica IP de origem, destaca-se que os valores de entropia apresentam principalmente um comportamento não linear.

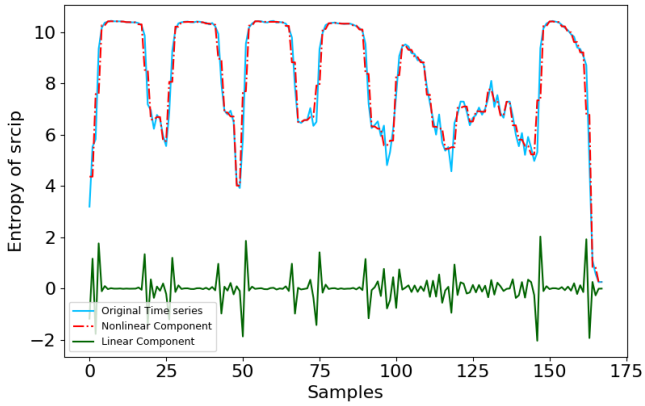
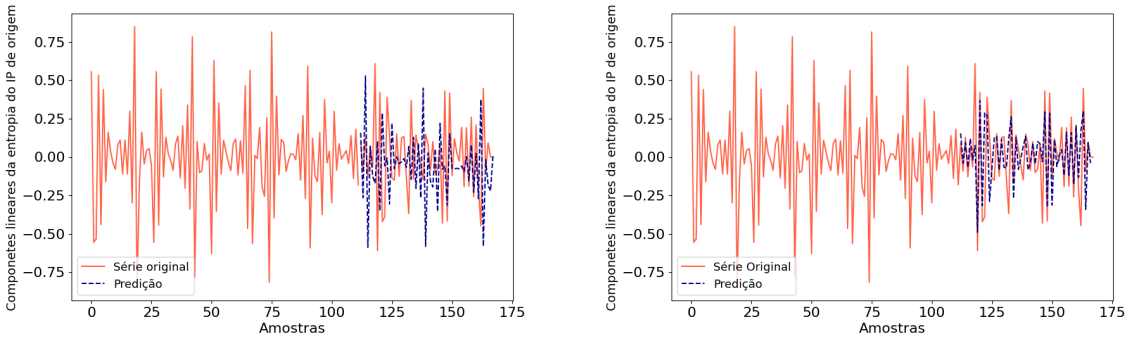


Figura 4.2: Entropia do recurso IP de origem e decomposição usando DWT. Séries originais e componentes lineares e não lineares extraídas.

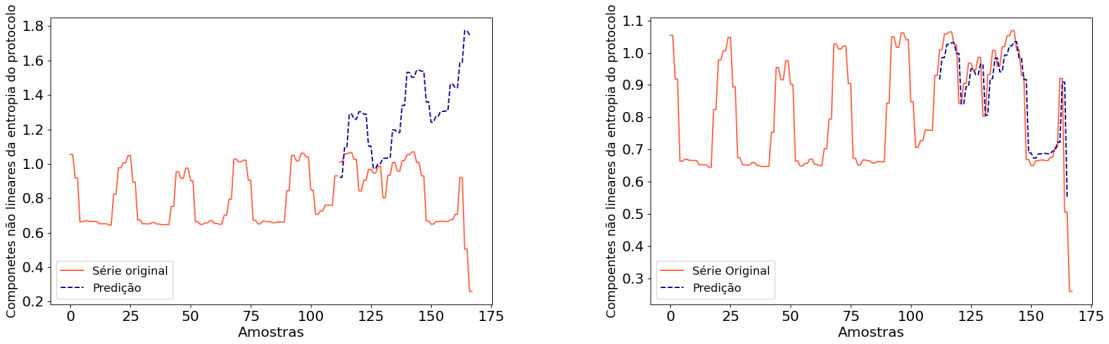


(a) Modelo ARIMA aplicado para a componente linear da entropia da cacaterística de IP de origem.

(b) Modelo LSTM aplicado para a componente near da entropia da cacaterística de IP de origem.

Figura 4.3: Comparação gráfica entre a aplicação dos modelos LSTM e ARIMA para a mesma componente linear.

O último passo da proposta é realizar a predição da entropia das características do tráfego da rede. Para esta etapa, foram treinados os modelos com 67% do conjunto de dados de entrada e para predição os seguintes 33%. Essa segmentação é equivalente ao treino de aproximadamente quatro dias de utilização da rede por parte dos usuários, dado janela de tempo de entropia de uma hora. A Figura 4.3 apresenta os métodos utilizados para prever a entropia com base nas componentes lineares. Por outro lado, Figure 4.4 apresenta uma comparação entre previsões utilizando componentes não lineares. É interessante notar, que no modelo ARIMA existe a partir do momento em que a série apresenta uma disruptura na ciclcidade enquanto o modelo LSTM consegue predizer forma mais acertiva, demonstrando assim uma superior vantagem do modelo de rede neural comparado com um modelo puramente estatístico.



(a) Modelo ARIMA aplicado para a componente não linear da entropia da cacaterística protocolo.

(b) Modelo LSTM aplicado para a componente não linear da entropia da cacaterística protocolo.

Figura 4.4: Comparação gráfica entre a aplicação dos modelos LSTM e ARIMA para a mesma componente não linear.

A métrica RMSE foi utilizada para o cálculo comparativo entre os modelos LSTM e ARIMA, sendo essa uma das métricas mais simples e eficientes para validação dos modelos de predição. Ao ocorrer uma diferença considerável entre esses valores e o quadrado é aplicado, tal diferença contribui para um alto peso no erro final do modelo de predição. Tanto para avaliação de predição das componentes lineares e não lineares, o modelo LSTM supera o ARIMA, pois o mesmo apresenta o menor valor de RMSE, conforme mostrado nas Tabelas 4.2 e 4.1. No entanto, a diferença entre os modelos de predição da componente linear é pequena, indicando que é possível, dependendo do espaço amostral, utilizar ambos. Como o modelo ARIMA é um método que exige a parametrização, foi implementada uma função que realiza testes para cada valor de p, d e q, e escolhe os melhores parâmetros de predição com base no Critério de Informação de Akaike (AIC). O AIC é um estimador de erro de predição no qual quanto menor mais simples e eficiente é o modelo. Embora

almejam-se os melhores parâmetros para o ARIMA, ele ainda apresenta um desempenho inferior em comparação ao LSTM. A Tabela 4.2 demonstra os resultados para a predição utilizando o ARIMA. Para cada característica, foram calculados o tempo de execução do código, incluído treinamento e predição. Cabe ressaltar que para ambos os casos, lineares e não lineares, os parâmetros p, d e q referentes ao modelo ARIMA, são os mesmos para as características com melhores resultados no tempo de execução e RMSE. Nos casos do modelo linear, os valores 2, 1, 1 foram estipulados para predição, enquanto para o modelo não linear, os valores escolhidos pelo estimador foram 0, 1, 0.

Tabela 4.1: Comparativo entre as componentes lineares e não lineares utilizando o modelo LSTM.

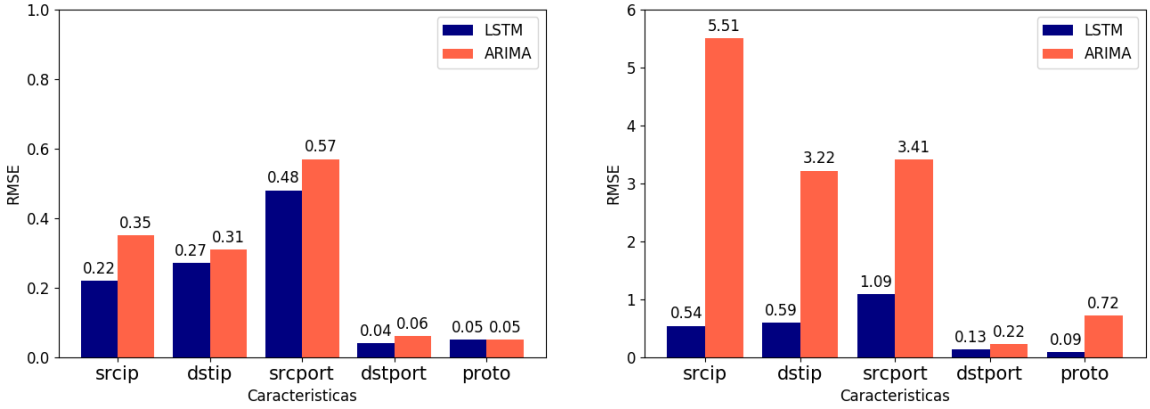
	Tempo decorrido	**RMSE**	**Perda Máxima**
srcip (linear)	13.98s	0.22	0.2037
dstip (linear)	12.93s	0.27	**0.1638**
srcport (linear)	12.94s	0.48	0.2159
dstport (linear)	**12.82s**	**0.04**	0.2360
proto (linear)	13.56s	0.05	0.2252
srcip (não linear)	13.73s	0.54	0.4240
dstip (não linear)	13.17s	0.59	0.6059
srcport (não linear)	13.29s	1.09	0.5938
dstport (não linear)	**13.13s**	0.13	0.3812
proto (não linear)	13.26s	**0.09**	**0.1951**

Tabela 4.2: Comparativo entre as componentes lineares e não lineares utilizando o modelo ARIMA.

	Tempo decorrido	**RMSE**	**AIC**	**Modelo**
srcip (linear)	**38s**	0.35	-55.07	2,1,1
dstip (linear)	42s	0.31	-40.0	3,1,1
srcport (linear)	40s	0.47	42.30	3,1,1
dstport (linear)	44s	0.06	-183.66	2,1,1
proto (linear)	44s	**0.05**	-408.29	2,1,1
srcip (não linear)	79s	5.51	43.23	0,1,0
dstip (não linear)	45s	3.22	71.23	1,1,1
srcport (não linear)	**39s**	3.41	175.36	1,1,1
dstport (não linear)	55s	**0.22**	-166.62	0,1,0
proto (não linear)	33s	0.72	356.17	0,1,0

Em relação à predição com LSTM, a Tabela 4.1 mostra o tempo de execução de cada recurso, o valor de RMSE e a perda máxima de cada época do modelo. É possível verificar que o tempo de execução do código foi significativamente melhor que o ARIMA e que os valores de RMSE também tiveram uma vantagem significativa. O componente linear

para o recurso de porta de destino (*dstport*) mostra os melhores resultados de tempo de execução e RMSE global. O recurso do tipo protocolo apresenta o melhor desempenho para a métrica RMSE, considerando os componentes não lineares.



(a) RMSE das características lineares.

(b) RMSE das características não lineares

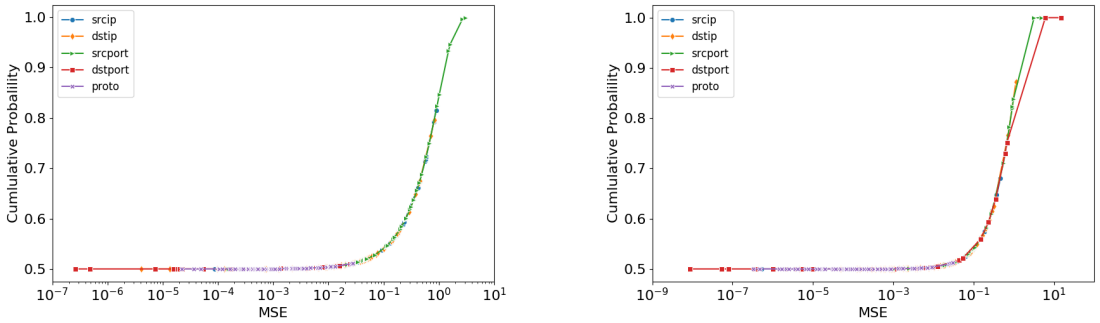
Figura 4.5: Os resultados de *Root Mean Square Error (RMSE)* para cada um dos modelos preditivos.

Os resultados do RMSE para as previsões LSTM e ARIMA são mostrados na Figura 4.5. O RMSE para os componentes lineares é apresentado na Figura 4.5(a) e as componentes não lineares são apresentados na Figura 4.5(b). É possível identificar que o modelo LSTM supera o ARIMA em ambos os cenários com grande diferença, principalmente na predição não linear. De fato, espera-se que o LSTM supere o ARIMA na predição desta componente, em função da alta capacidade de adequação das redes neurais para lidar com problemas não lineares. Para as componentes lineares os valores obtidos de RMSE para o modelo ARIMA foram tão baixo quantos do LSTM, demostrando que baseando-se nesta métrica ambos os modelos podem ser considerados para realizar predições. Ressalta-se ainda que o melhor desempenho de tempo para os modelos baseados em redes neurais, deve-se à implementação otimizada dos algoritmos das bibliotecas já disponíveis em Python[1].

Foram calculadas Função de Distribuição Acumulada (FDA) para efeito de comparação entre a probabilidade de as características analisadas alcançarem um Erro Quadrático Médio (MSE). O MSE é outra medida da qualidade dos modelos de predição, no qual valores mais baixos indicam maior precisão das previsões. A Figura 4.6 apresenta os valores da Função de distribuição acumulada do erro quadrático médio, para cada modelo de previsão. A componente linear da característica porta de destino utilizando LSTM apresenta

[1]https://pypi.org/

o menor erro de predição devido aos baixos valores de entropia gerado, indicando que uma variação nesse parâmetro pode ser considerada uma anomalia na rede.



(a) Probabilidade acumulada do modelo ARIMA para componentes lineares

(b) Probabilidade acumulada do modelo LSTM para componentes lineares

Figura 4.6: Probabilidade acumulada (FDA) do erro quadrático médio (MSE) para (a) FDA para as componentes lineares do modelo ARIMA, (b) FDA para as componentes lineares do modelo LSTM.

Capítulo 5

Conclusão e trabalhos futuros

As redes móveis propiciaram um grande avanço nas comunicações, pois suas propriedades físicas alcançam diversos ambientes. Outrossim, a comunicação de dados dessas redes modificou as relações humanas na última década, pois aplicações variadas foram desenvolvidas e outras aperfeiçoadas para viabilizar uma maior troca de dados em tempo real, como transmissão de vídeos, vídeo c,ivos, permitindo que a inteligência artificial seja aplicada no aperfeiçoamento de recursos energéticos por exemplo. Particularmente, a rede 5G possui avanços relevantes com relação às gerações anteriores como Serviços de Missão Crítica (MCS) e Banda larga móvel aprimorada (eMBB). Veículos autônomos e serviços de saúde remotos são exemplos de paradigmas de serviços de missão crítica, no qual a baixa latência e as altas taxas de transmissão são indispensáveis para efetividade, pois atrasos na chegada de pacotes podem ocasionar danos irrecuperáveis. Além disso, altas taxas de transmissão permitirão maiores demandas de vídeos com alta definição, gerando um tráfego acentuado nas redes, principalmente em função das redes de distribuição de conteúdo (CDN). Premissas como essas devem ser aplicadas para casos relativos à utilização de realidade virtual (*Virtual Reality* - VR), realidade aumentada (*Augumented Reality* - AR) ou realização de monitoramento e rastreamento de doenças infectocontagiosas através de sistemas de vigilância com câmeras infravermelhas, tal como a pandemia do SARS-CoV2 (COVID-19).

Neste trabalho foi analisada a proposta de geração de novas características a partir do cálculo de entropia das características categóricas do fluxo de rede, visando demonstrar o grau de aleatoriedade existente em sistema de comunicações para utilização em predição de anomalias e tráfego de rede. Desta forma a utilização de características categóricas, que são informações elementares em um fluxo mas não essencias para predição de tráfego, por exemplo, podem ser utilizadas após o cálculo de entropia como métricas promissoras. Esta

abordagem pode ser utilizada também em conjunto de dados com características sazonais, pois dependendo do uso, o tráfego de rede exibe alta dispersão estatística relacionada aos recursos de fluxo do modelo TCP/IP, como IP de origem e destino ou portas de origem e destino. O modelo estatístico ARIMA analisa as classificações temporais para entender os dados históricos e prevê o tráfego com base em médias móveis e regressão linear. Em contraste, o modelo de rede neural LSTM possuí a vantagem de ser mais eficiente no processamento do que o ARIMA. Os resultados mostram que a rede neural LSTM fornece baixo erro na predição de componentes lineares e não lineares da série temporal de entropia. Conclui-se que um método híbrido que considera ambos os modelos, ARIMA e LSTM é a melhor solução para prever o comportamento anormal de uma rede de grande escala, como as redes 5G.

A fundamentação teórica desta dissertação originou a produção das seguintes publicações:

- G. N. N. Barbosa, M. A. Lopez, D. S. V. Medeiros and D. M. F. Mattos, "An Entropy-based Hybrid Mechanism for Large-Scale Wireless Network Traffic Prediction,"2021 International Symposium on Networks, Computers and Communications (ISNCC), 2021;
- BARBOSA, G.; BEZERRA, G. M.; MEDEIROS, D. S. de; LOPEZ, M. A.; MATTOS, D. Segurança em Redes 5G: Oportunidades e Desafios em Detecção de Anomalias e Predição de Tráfego Baseadas em Aprendizado de Máquina. SBC, out 2021.;

Como trabalhos futuros, identifica-se a necessidade de aprimorar as técnicas de aprendizado de máquinas voltadas à predição de anomalias de maneira distribuída, onde cada dispositivo inteligente é capaz de treinar dados localmente, enviando para um servidor central os modelos de treinamento. Esta é uma das características do modelo de aprendizado federado, ou *federated learning*. Isto porque com o aperfeiçoamento dos dispositivos móveis e com a implementação das redes 6G e 7G, eles terão incremento na capacidade de processamento gráfico o que facilita inclusive a utilização de outras técnicas como de Redes Neurais Convolucionais. Ainda como trabalho futuro, pretende-se aplicar um classificador de máquina de vetor de suporte de uma classe (OC-SVM) para detecção de anomalias em fluxos de redes com foco na análise em tempo real.

Referências

[1] BARBOSA, G.; BEZERRA, G. M.; MEDEIROS, D. S. de; LOPEZ, M. A.; MATTOS, D. *Segurança em Redes 5G: Oportunidades e Desafios em Detecção de Anomalias e Predição de Tráfego Baseadas em Aprendizado de Máquina*. SBC, out 2021. 145–189 p. Disponível em: <http://dx.doi.org/10.5753/sbc.7165.8.4>.

[2] KHAN, R.; KUMAR, P.; JAYAKODY, D. N. K.; LIYANAGE, M. A survey on security and privacy of 5G technologies: Potential solutions, recent advancements, and future directions. *IEEE Communications Surveys & Tutorials*, IEEE, v. 22, n. 1, p. 196–248, 2019.

[3] DUTTA, A.; HAMMAD, E. 5g security challenges and opportunities: A system approach. In: *2020 IEEE 3rd 5G World Forum (5GWF)*. [S.l.: s.n.], 2020. p. 109–114.

[4] AHMAD, I.; KUMAR, T.; LIYANAGE, M.; OKWUIBE, J.; YLIANTTILA, M.; GURTOV, A. Overview of 5G security challenges and solutions. *IEEE Communications Standards Magazine*, v. 2, n. 1, p. 36–43, 2018.

[5] YAO, J.; HAN, Z.; SOHAIL, M.; WANG, L. A robust security architecture for SDN-based 5G networks. *Future Internet*, Multidisciplinary Digital Publishing Institute, v. 11, n. 4, p. 85, 2019.

[6] SCOTT-HAYWARD, S.; NATARAJAN, S.; SEZER, S. A survey of security in software defined networks. *IEEE Communications Surveys & Tutorials*, IEEE, v. 18, n. 1, p. 623–654, 2015.

[7] ALSAEEDI, M.; MOHAMAD, M. M.; AL-ROUBAIEY, A. A. Toward adaptive and scalable openflow-sdn flow control: A survey. *IEEE Access*, v. 7, p. 107346–107379, 2019.

[8] CHOWDHURY, M. Z.; SHAHJALAL, M.; AHMED, S.; JANG, Y. M. 6G wireless communication systems: Applications, requirements, technologies, challenges, and research directions. *IEEE Open Journal of the Communications Society*, v. 1, p. 957–975, 2020.

[9] DOGRA, A.; JHA, R. K.; JAIN, S. A survey on beyond 5G network with the advent of 6G: Architecture and emerging technologies. *IEEE Access*, v. 9, p. 67512–67547, 2021.

[10] ALWIS, C. D.; KALLA, A.; PHAM, Q.-V.; KUMAR, P.; DEV, K.; HWANG, W.-J.; LIYANAGE, M. Survey on 6G frontiers: Trends, applications, requirements, technologies and future research. *IEEE Open Journal of the Communications Society*, v. 2, p. 836–886, 2021.

[11] SUN, Y.; LIU, J.; WANG, J.; CAO, Y.; KATO, N. When machine learning meets privacy in 6G: A survey. *IEEE Communications Surveys Tutorials*, v. 22, n. 4, p. 2694–2724, 2020.

[12] NETO, H. N. C.; MATTOS, D. M. F.; FERNANDES, N. C. Privacidade do usuário em aprendizado colaborativo: Federated learning, da teoria à prática. *Minicursos do Simpósio Brasileiro de Segurança de Informação e de Sistemas Computacionais - SBSeg*, v. 20, p. 142–195, 2020.

[13] GIORDANI, M.; POLESE, M.; MEZZAVILLA, M.; RANGAN, S.; ZORZI, M. Toward 6G networks: Use cases and technologies. *IEEE Communications Magazine*, v. 58, n. 3, p. 55–61, 2020.

[14] ZHANG, S.; WANG, Y.; ZHOU, W. Towards secure 5G networks: A survey. *Computer Networks*, Elsevier, v. 162, p. 106871, 2019.

[15] LOPEZ-MARTIN, M.; CARRO, B.; SANCHEZ-ESGUEVILLAS, A.; LLORET, J. Conditional variational autoencoder for prediction and feature recovery applied to intrusion detection in IoT. *Sensors*, v. 17, n. 9, 2017. ISSN 1424-8220. Disponível em: <https://www.mdpi.com/1424-8220/17/9/1967>.

[16] Ariyaluran Habeeb, R. A.; NASARUDDIN, F.; GANI, A.; Targio Hashem, I. A.; AHMED, E.; IMRAN, M. Real-time big data processing for anomaly detection: A survey. *International Journal of Information Management*, v. 45, p. 289–307, 2019. ISSN 0268-4012. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0268401218301658>.

[17] AHMAD, S.; LAVIN, A.; PURDY, S.; AGHA, Z. Unsupervised real-time anomaly detection for streaming data. *Neurocomputing*, v. 262, p. 134–147, 2017. ISSN 0925-2312. Online Real-Time Learning Strategies for Data Streams. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0925231217309864>.

[18] ASSIS, M. V. O. D.; HAMAMOTO, A. H.; ABRãO, T.; PROENçA, M. L. A game theoretical based system using holt-winters and genetic algorithm with fuzzy logic for DoS/DDoS mitigation on sdn networks. *IEEE Access*, v. 5, p. 9485–9496, 2017.

[19] TARTAKOVSKY, A. G.; POLUNCHENKO, A. S.; SOKOLOV, G. Efficient computer network anomaly detection by changepoint detection methods. *IEEE Journal of Selected Topics in Signal Processing*, v. 7, n. 1, p. 4–11, 2013.

[20] BANG, J. ho; CHO, Y.-J.; KANG, K. Anomaly detection of network-initiated LTE signaling traffic in wireless sensor and actuator networks based on a hidden semi-markov model. *Computers and Security*, v. 65, p. 108–120, 2017. ISSN 0167-4048. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0167404816301614>.

[21] YU, S.-Z. Hidden semi-markov models. *Artificial Intelligence*, v. 174, n. 2, p. 215–243, 2010. ISSN 0004-3702. Special Review Issue. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0004370209001416>.

[22] FU, Y.; WANG, S.; WANG, C.-X.; HONG, X.; MCLAUGHLIN, S. Artificial intelligence to manage network traffic of 5g wireless networks. *IEEE Network*, v. 32, n. 6, p. 58–64, 2018.

[23] LU, N.; LI, D.; SHI, W.; VIJAYAKUMAR, P.; PICCIALLI, F.; CHANG, V. An efficient combined deep neural network based malware detection framework in 5G environment. *Computer Networks*, v. 189, p. 107932, 2021. ISSN 1389-1286. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1389128621000785>.

[24] LUO, C.; JI, J.; WANG, Q.; CHEN, X.; LI, P. Channel state information prediction for 5G wireless communications: A deep learning approach. *IEEE Transactions on Network Science and Engineering*, v. 7, n. 1, p. 227–236, 2020.

[25] ZHU, G.; ZAN, J.; YANG, Y.; QI, X. A supervised learning based QoS assurance architecture for 5G networks. *IEEE Access*, v. 7, p. 43598–43606, 2019.

[26] ALAWE, I.; KSENTINI, A.; HADJADJ-AOUL, Y.; BERTIN, P. Improving traffic forecasting for 5G core network scalability: A machine learning approach. *IEEE Network*, v. 32, n. 6, p. 42–49, 2018.

[27] SEDJELMACI, H. Cooperative attacks detection based on artificial intelligence system for 5G networks. *Computers and Electrical Engineering*, v. 91, p. 107045, 2021. ISSN 0045-7906. Disponível em: <https://www.sciencedirect.com/science/article/pii/S004579062100063X>.

[28] MAIMO, L. F.; GOMEZ, A. L. P.; CLEMENTE, F. J. G.; PEREZ, M. G.; PEREZ, G. M. A self-adaptive deep learning-based system for anomaly detection in 5G networks. *IEEE Access*, v. 6, p. 7700–7712, 2018.

[29] LOBATO, A. G. P.; LOPEZ, M. A.; CARDENAS, A. A.; DUARTE, O. C. M. B.; PUJOLLE, G. A fast and accurate threat detection and prevention architecture using stream processing. *Concurrency and Computation: Practice and Experience*, e6561, 2021.

[30] MEDEIROS, D.; NETO, H. C.; ANDREONI, M.; MAGALHãES, L.; SILVA, E.; BORGES, A.; FERNANDES, N.; MENEZES, D. Análise de dados em redes sem fio de grande porte: Processamento em fluxo em tempo real, tendências e desafios. In: _____. [S.l.]: Sociedade Brasileira de Computação, 2019. p. 142–195. ISBN 2177-9384.

[31] KAELBLING, L. P.; LITTMAN, M. L.; MOORE, A. W. Reinforcement learning: A survey. *Journal of artificial intelligence research*, v. 4, p. 237–285, 1996.

[32] GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep Learning*. [S.l.]: MIT Press, 2016. `http://www.deeplearningbook.org`.

[33] TRINH, H. D.; ZEYDAN, E.; GIUPPONI, L.; DINI, P. Detecting mobile traffic anomalies through physical control channel fingerprinting: A deep semi-supervised approach. *IEEE Access*, v. 7, p. 152187–152201, 2019.

[34] JIANG, W.; SCHOTTEN, H. D. Neural network-based fading channel prediction: A comprehensive overview. *IEEE Access*, v. 7, p. 118112–118124, 2019.

[35] RAMAKRISHNAN, N.; SONI, T. Network traffic prediction using recurrent neural networks. In: *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*. [S.l.: s.n.], 2018. p. 187–193.

[36] ZHANG, C.; PATRAS, P.; HADDADI, H. Deep learning in mobile and wireless networking: A survey. *IEEE Communications Surveys Tutorials*, v. 21, n. 3, p. 2224–2287, 2019.

[37] LOPEZ, M. A.; MATTOS, D. Resumo de grandes volumes de dados com filtro de bloom: Uma abordagem eficiente para aprendizado profundo com redes neurais convolucionais em fluxos de rede. In: *Anais do XXXIX Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*. Porto Alegre, RS, Brasil: SBC, 2021. p. 532–545. ISSN 2177-9384. Disponível em: <https://sol.sbc.org.br/index.php/sbrc/article/view/16745>.

[38] BOCHIE, K.; GILBERT, M.; GANTERT, L.; BARBOSA, M.; MEDEIROS, D.; CAMPISTA, M. Aprendizado profundo em redes desafiadoras: Conceitos e aplicações. In: *Minicursos do XXXVIII Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*. SBC, 2020. p. 140–189. Disponível em: <https://doi.org/10.5753/sbc.5033.7.4>.

[39] WU, D.; NEKOVEE, M.; WANG, Y. Deep learning-based autoencoder for m-user wireless interference channel physical layer design. *IEEE Access*, v. 8, p. 174679–174691, 2020.

[40] WANG, Y.; MASOUD, N.; KHOJANDI, A. Real-time sensor anomaly detection and recovery in connected automated vehicle sensors. *IEEE Transactions on Intelligent Transportation Systems*, v. 22, n. 3, p. 1411–1421, 2021.

[41] BOUKERCHE, A.; TAO, Y.; SUN, P. Artificial intelligence-based vehicular traffic flow prediction methods for supporting intelligent transportation systems. *Computer Networks*, v. 182, p. 107484, 2020. ISSN 1389-1286. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1389128620311567>.

[42] CHAKRABORTY, P.; CORICI, M.; MAGEDANZ, T. A comparative study for time series forecasting within software 5G networks. In: *2020 14th International Conference on Signal Processing and Communication Systems (ICSPCS)*. [S.l.: s.n.], 2020. p. 1–7.

[43] YANG, H.; LI, X.; QIANG, W.; ZHAO, Y.; ZHANG, W.; TANG, C. A network traffic forecasting method based on sa optimized arima–bp neural network. *Computer Networks*, v. 193, p. 108102, 2021. ISSN 1389-1286. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1389128621001821>.

[44] SONE, S. P.; LEHTOMäKI, J. J.; KHAN, Z. Wireless traffic usage forecasting using real enterprise network data: Analysis and methods. *IEEE Open Journal of the Communications Society*, v. 1, p. 777–797, 2020.

[45] KROMKOWSKI, P.; LI, S.; ZHAO, W.; ABRAHAM, B.; OSBORNE, A.; BROWN, D. E. Evaluating statistical models for network traffic anomaly detection. In: *2019 Systems and Information Engineering Design Symposium (SIEDS)*. [S.l.: s.n.], 2019. p. 1–6.

[46] HANBANCHONG, A.; PIROMSOPA, K. SARIMA based network bandwidth anomaly detection. In: *2012 Ninth International Conference on Computer Science and Software Engineering (JCSSE)*. [S.l.: s.n.], 2012. p. 104–108.

[47] DWIVEDI, R. K.; PANDEY, S.; KUMAR, R. A study on machine learning approaches for outlier detection in wireless sensor network. In: *2018 8th International Conference on Cloud Computing, Data Science Engineering (Confluence)*. [S.l.: s.n.], 2018. p. 189–192.

[48] AHMAD, A.; HARJULA, E.; YLIANTTILA, M.; AHMAD, I. Evaluation of machine learning techniques for security in SDN. In: *2020 IEEE Globecom Workshops (GC Wkshps*. [S.l.: s.n.], 2020. p. 1–6.

[49] REDDY, K.; THILAGAM, P. Naïve bayes classifier to mitigate the DDoS attacks severity in ad-hoc networks. *International Journal of Communication Networks and Information Security*, v. 12, p. 221–226, 08 2020.

[50] SWARNKAR, M.; HUBBALLI, N. OCPAD: One class naive bayes classifier for payload based anomaly detection. *Expert Systems with Applications*, v. 64, p. 330–339, 2016. ISSN 0957-4174. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0957417416303839>.

[51] CHADZA, T.; KYRIAKOPOULOS, K. G.; LAMBOTHARAN, S. Analysis of hidden markov model learning algorithms for the detection and prediction of multi-stage network attacks. *Future Generation Computer Systems*, v. 108, p. 636–649, 2020. ISSN 0167-739X. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0167739X1932847X>.

[52] ZHAN, M.; LI, Y.; YANG, X.; CUI, W.; FAN, Y. NSAPs: A novel scheme for network security state assessment and attack prediction. *Computers and Security*, v. 99, p. 102031, 2020. ISSN 0167-4048. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0167404820303047>.

[53] BERNARDINO, E. D.; BROGI, G. Hidden markov models for advanced persistent threats. *International Journal of Security and Networks*, v. 14, p. 181, 01 2019.

[54] SIVANATHAN, A.; GHARAKHEILI, H. H.; LOI, F.; RADFORD, A.; WIJENAYAKE, C.; VISHWANATH, A.; SIVARAMAN, V. Classifying iot devices in smart environments using network traffic characteristics. *IEEE Transactions on Mobile Computing*, v. 18, n. 8, p. 1745–1759, 2019.

[55] YUE, C.; JIN, R.; SUH, K.; QIN, Y.; WANG, B.; WEI, W. Linkforecast: Cellular link bandwidth prediction in lte networks. *IEEE Transactions on Mobile Computing*, v. 17, n. 7, p. 1582–1594, 2018.

[56] FENG, J.; CHEN, X.; GAO, R.; ZENG, M.; LI, Y. Deeptp: An end-to-end neural network for mobile cellular traffic prediction. *IEEE Network*, v. 32, n. 6, p. 108–115, 2018.

[57] FRANK, L. R.; OLIVEIRA, R. M. D.; VIEIRA, A. B.; SILVA, E. F. Improving a smart environment with wireless network user load prediction. In: *2021 IEEE Symposium on Computers and Communications (ISCC)*. [S.l.: s.n.], 2021. p. 1–6.

[58] MEI, L.; GOU, J.; CAI, Y.; CAO, H.; LIU, Y. Realtime mobile bandwidth and handoff predictions in 4g/5g networks. *Computer Networks*, v. 204, p. 108736, 2022. ISSN 1389-1286. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1389128621005879>.

[59] BENSLIMEN, Y.; SEDJELMACI, H.; MANENTI, A.-C. Attacks and failures prediction framework for a collaborative 5G mobile network. *Computing*, v. 103, n. 6, p. 1165–1181, Jun 2021. ISSN 1436-5057. Disponível em: <https://doi.org/10.1007/s00607-020-00893-8>.

[60] WANG, J.; TANG, J.; XU, Z.; WANG, Y.; XUE, G.; ZHANG, X.; YANG, D. Spatiotemporal modeling and prediction in cellular networks: A big data enabled deep learning approach. In: *IEEE INFOCOM 2017 - IEEE Conference on Computer Communications*. [S.l.: s.n.], 2017. p. 1–9.

[61] NIE, L.; JIANG, D.; YU, S.; SONG, H. Network traffic prediction based on Deep Belief Network in wireless mesh backbone networks. In: *2017 IEEE Wireless Communications and Networking Conference (WCNC)*. [S.l.: s.n.], 2017. p. 1–5.

[62] L. Wang; Q. Li; Y. Jiang; X. Jia; J. Wu. Woodpecker: Detecting and mitigating link-flooding attacks via sdn. In: *Computer Networks*. [S.l.]: Elsevier, 2018.

[63] Vaclav Bartos; Martin Zadnik; Sheikh Mahbub Habib; Emmanouil Vasilomanolakis. Network entity characterization and attack prediction. In: *Future Generation Computer Systems*. [S.l.]: Elsevier, 2019.

[64] BILAL, A.; TARIK, T.; VAJDA, A.; MILOUD, B. Dynamic cloud resource scheduling in virtualized 5g mobile systems. In: *2016 IEEE Global Communications Conference (GLOBECOM)*. [S.l.: s.n.], 2016. p. 1–6.

[65] DIAS, R.; MAURICIO, L. A.; ARAGãO, M. V. P. de. Dasrs rest: Um algoritmo eficiente de detecção de anomalias em tempo real para data centers. In: *Anais do XXXVIII Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*. Porto Alegre, RS, Brasil: SBC, 2020. p. 672–685. ISSN 2177-9384. Disponível em: <https://sol.sbc.org.br/index.php/sbrc/article/view/12317>.

[66] MATTOS, D. M. F.; VELLOSO, P. B.; DUARTE, O. C. M. B. An agile and effective network function virtualization infrastructure for the internet of things. *Journal of Internet Services and Applications*, v. 10, n. 1, p. 6, Mar 2019. ISSN 1869-0238.

[67] MEDEIROS, D. S. V.; NETO, H. N. C.; LOPEZ, M. A.; MAGALHÃES, L. C. S.; FERNANDES, N. C.; VIEIRA, A. B.; SILVA, E. F.; MATTOS, D. M. F. A survey on data analysis on large-scale wireless networks: online stream processing, trends, and challenges. *Journal of Internet Services and Applications*, v. 11, n. 1, p. 6, Oct 2020. ISSN 1869-0238.

[68] TANG, J.; CHEN, X.; HU, Z.; ZONG, F.; HAN, C.; LI, L. Traffic flow prediction based on combination of support vector machine and data denoising schemes. *Physica A: Statistical Mechanics and its Applications*, v. 534, p. 120642, 2019. ISSN 0378-4371. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0378437119302262>.

[69] Zihan Chang; Yang Zhang; Wenbo Chen. Electricity price prediction based on hybrid model of adam optimized lstm neural network and wavelet transform. In: *Energy*. [S.l.]: Elsevier, 2019.

[70] REIS, L. H. A.; MAGALHÃES, L. C. S.; MEDEIROS, D. S. V. de; MATTOS, D. M. An unsupervised approach to infer quality of service for large-scale wireless networking. *Journal of Network and Systems Management*, Springer, v. 28, n. 4, p. 1228–1247, 2020.

[71] KANUNGO, A.; MITTAL, M.; DEWAN, L. Comparison of haar and daubechies wavelet based denoising for speed control of dc motor. In: *2020 First IEEE International Conference on Measurement, Instrumentation, Control and Automation (ICMICA)*. [S.l.: s.n.], 2020. p. 1–4.