



THIAGO DO PRADO RAMOS

**DIMENSIONAMENTO DE REDE SEM FIO DE UM CAMPUS
UNIVERSITÁRIO BASEADO NA ANÁLISE DE *LOGS* DE
CONEXÃO**

LAVRAS – MG

2023

THIAGO DO PRADO RAMOS

**DIMENSIONAMENTO DE REDE SEM FIO DE UM CAMPUS UNIVERSITÁRIO BASEADO
NA ANÁLISE DE *LOGS* DE CONEXÃO**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Ciência da Computação, área de concentração em Redes de Computadores e Sistemas Embarcados, para a obtenção do título de Mestre

Prof. DSc. Eric Fernandes de Mello Araújo
Orientador

Prof. DSc. Luiz Henrique Andrade Correia
Coorientador

**LAVRAS – MG
2023**

**Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca
Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).**

Ramos, Thiago do Prado.

Dimensionamento de rede sem fio de um campus universitário baseado na análise de logs de conexão / Thiago do Prado Ramos. - 2023.

63 p. : il.

Orientador: Prof. DSc. Eric Fernandes de Mello Araújo.

Coorientador: Prof. DSc. Luiz Henrique Andrade Correia.

Dissertação (mestrado acadêmico) – Universidade Federal de Lavras, 2023.

Bibliografia.

1. Dataset, . 2. Wi-Fi. 3. Dimensionamento de recursos.
I. Araújo, Eric Fernandes de Mello. II. Correia, Luiz Henrique Andrade. III. Título.

THIAGO DO PRADO RAMOS

**DIMENSIONAMENTO DE REDE SEM FIO DE UM CAMPUS UNIVERSITÁRIO BASEADO
NA ANÁLISE DE LOGS DE CONEXÃO**

**WIRELESS NETWORK SIZING OF A UNIVERSITY CAMPUS BASED IN CONNECTION
LOG ANALYSIS**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Ciência da Computação, área de concentração em Redes de Computadores e Sistemas Embarcados, para a obtenção do título de Mestre

APROVADA em 24 de Fevereiro de 2023.

Prof. DSc. Eric Fernandes de Mello Araújo	UFLA
Prof. DSc. Luiz Henrique Andrade Correia	UFLA
Prof. DSc. Demostenes Zegarra Rodrigues	UFLA
Prof. DSc. Thais Regina de Moura Braga Silva	UFV

Prof. DSc. Eric Fernandes de Mello Araújo
Orientador

Prof. DSc. Luiz Henrique Andrade Correia
Co-Orientador

**LAVRAS – MG
2023**

Agradeço a toda a minha família pelo carinho e apoio, que não mediram esforços para que eu chegasse até esta etapa na caminhada da minha vida. Este trabalho foi realizado com apoio do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Fundação de Amparo à Pesquisa de Minas Gerais (FAPEMIG) e Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

RESUMO

As redes de computadores estão cada vez mais heterogêneas e seus usuários acessam diversos tipos de aplicações usando diferentes tecnologias. O acesso às redes de computadores utilizando dispositivos móveis tem aumentado, fazendo uso massivo das conexões sem fio. Torna-se necessário conhecer os locais de conexão dos usuários dentro da rede para ser possível projetar e dimensionar redes com qualidade de serviço e melhorar a experiência do usuário. Com as informações de conexão pode-se melhorar o provisionamento de recursos, direcionar o tráfego de maneira mais eficaz, dependendo do tipo de serviço e localidade do usuário, economizar energia de dispositivos ou ainda conseguir prever problemas de segurança e desempenho da rede. Neste trabalho foi proposta a geração de uma base de dados (*dataset*) fundamentada nos arquivos de *logs* de conexão da rede Wi-Fi de um campus universitário, com aproximadamente 15 mil usuários. Esse *dataset* contém informações de um período de 4 meses, com intervalos de tempo divididos em 15 minutos, locais de conexão, tipos de usuários, quantidade de conexões, entre outros dados. Esse *dataset* foi anonimizado e serve como base para auxiliar na tomada de decisões quanto ao dimensionamento do número de pontos de acesso necessários. Assim, foi definida uma metodologia baseada em largura de banda mínima por usuário para cálculo do dimensionamento de recursos. Não foi avaliado atribuição de frequências, sobreposição de cobertura entre APs operando no mesmo canal de rádio e controle de potência de transmissão. Todas essas questões são resolvidas de forma automática pela controladora dos equipamentos da rede sem fio. Uma limitação encontrada se deve ao desafio do chamado efeito Ping Pong, que corresponde ao registro de uma sucessão de conexões/desconexões entre APs enquanto o usuário está parado. Para contornar esse problema foi realizado o agrupamento dos APs de um mesmo prédio/setor. Este trabalho se diferencia dos trabalhos anteriores da área pela criação e disponibilização da base de dados categorizada e com locais de conexão e por propor o dimensionamento de equipamentos baseado em largura de banda para o usuário baseado no número de conexões por local. Para definição da banda mínima reservada para cada usuário foi utilizado um guia dos serviços onlines típicos com a velocidade mínima de download necessária para o desempenho adequado de cada aplicativo descrito no texto. Os resultados indicam quais os locais que necessitam a adição de equipamentos. Também foram criados mapas georreferenciados interativos, onde foi possível localizar os locais com maior concentração de usuários no decorrer do tempo.

Palavras-chave: *Dataset*. Wi-Fi. Ponto de acesso. Dimensionamento de recursos.

ABSTRACT

Computer networks are increasingly heterogeneous and their users access different types of applications using different technologies. Access to computer networks using mobile devices has increased, making massive use of wireless connections. It becomes necessary to know the connection locations of users within the network to be able to design and scale networks with quality of service and improve the user experience. With the connection information, it is possible to improve resource provisioning, direct traffic more effectively, depending on the type of service and user location, save device energy or even predict security and network performance issues. In this work, it was proposed the generation of a database (*dataset*) based on the files of *logs* connection to the Wi-Fi network of a university campus, with approximately 15 thousand users. This *dataset* contains information for a period of 4 months, with time intervals divided into 15 minutes, connection locations, types of users, number of connections, among other data. This *dataset* was anonymized and serves as a basis to assist in making decisions regarding the dimensioning of the number of access points needed. Thus, a methodology based on the minimum bandwidth per user was defined for calculating resource sizing. Frequency allocation, coverage overlap between APs operating on the same radio channel and transmission power control were not evaluated. All these issues are automatically resolved by the controller of the wireless network equipment. A limitation found is due to the challenge of the so-called Ping Pong effect, which corresponds to the recording of a succession of connections/disconnections between APs while the user is stopped. To work around this problem, the APs of the same building/sector were grouped together. This work differs from previous works in the area by creating and making available a categorized database with connection locations and by proposing equipment sizing based on bandwidth for the user based on the number of connections per location. To define the minimum bandwidth reserved for each user, a guide of typical online services was used with the minimum download speed necessary for the adequate performance of each application described in the text. The result indicates which locations need the addition of equipment. Interactive georeferenced maps were also created, where it was possible to locate the places with the highest concentration of users over time.

Keywords: *Dataset*. Wi-Fi. Access Point. Resource sizing.

LISTA DE FIGURAS

Figura 1.1 – Métricas globais de rede sem fio	9
Figura 2.1 – Elementos de um rede 802.11	14
Figura 2.2 – Evolução do padrão 802.11	15
Figura 2.3 – Controladora WLAN	17
Figura 4.1 – Diagrama de conexão do Radius	31
Figura 4.2 – Conteúdo do arquivo de <i>Log</i>	32
Figura 4.3 – Relatório da controladora com associação do MAC	33
Figura 4.4 – Arquivo com MAC e local do AP	33
Figura 4.5 – <i>Log</i> de conexão anonimizado	35
Figura 4.6 – Mapa de exemplo dos locais com <i>Hotspot</i>	36
Figura 4.7 – Exemplo <i>dataframe</i> do local com latitude e longitude	37
Figura 4.8 – <i>DataFrame</i> com os dados da conversão do arquivo de <i>log</i>	38
Figura 4.9 – <i>DataFrame</i> após a limpeza de campos não utilizados	39
Figura 4.10 – <i>DataFrame</i> com MAC alterado para o local	39
Figura 4.11 – <i>Dataframe</i> com Latitude e Longitude	40
Figura 4.12 – <i>Dataframe</i> anonimizado	40
Figura 4.13 – Recorte <i>dataframe</i>	41
Figura 5.1 – Mapa de calor - 07/06/22 - 12:00	44
Figura 5.2 – Número de usuários conectados em 07/06/22 - 12:00	45
Figura 5.3 – Mapa de calor - 07/06/22 - 14:15	45
Figura 5.4 – Número de usuários conectados em 07/06/22 - 14:15	46
Figura 5.5 – Mapa de categoria	46
Figura 5.6 – Informações da categoria docente no local DEX	47
Figura 5.7 – Número de conexões x local	48
Figura 5.8 – Número de conexões x local x categoria - junho/22	49

SUMÁRIO

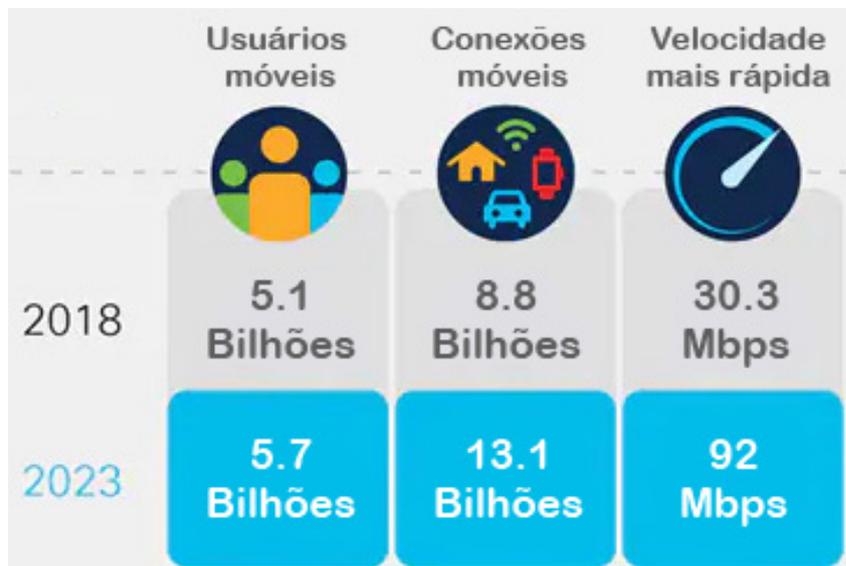
1	INTRODUÇÃO	9
1.1	Motivação	11
1.2	Definição do problema	11
1.3	Solução Proposta	12
1.4	Objetivos	12
1.5	Estrutura do Trabalho	12
2	REFERENCIAL TEÓRICO	13
2.1	Padrão IEEE 802.11	13
2.2	Gerenciamento centralizado	16
2.3	Estrutura de dados de Log de conexão	18
2.4	Dimensionamento de pontos de acesso	19
3	TRABALHOS RELACIONADOS	22
4	MATERIAIS E MÉTODOS	29
4.1	Controladora de rede sem fio	29
4.2	Arquivo de log de conexões	30
4.3	Tratamento de dados	34
4.4	Análise georreferenciada	35
4.5	Algoritmos para tratamento dos dados	37
4.6	Dimensionamento do número de APs	41
5	RESULTADOS	44
5.1	Mapa de calor - (<i>Heatmap</i>)	44
5.2	Mapa com pontos de dispersão	46
5.3	Análise gráfica do número conexões	47
5.3.1	Maior número de conexões	47
5.3.2	Número de acessos por categorias	48
5.4	Dimensionamento de recursos de rede	50
6	CONCLUSÃO	54
	REFERÊNCIAS	56
	APÊNDICE A	61

1 INTRODUÇÃO

Com a popularização dos dispositivos móveis (celulares, tablets e notebooks) cresceu a demanda por acesso de alta velocidade nas redes de dados sem fio. Apesar do grande volume de acesso à tecnologia 4G, o acesso às redes Wi-Fi baseadas em padrão 802.11 continua em franca expansão (JARDOSH et al., 2008). Faculdades, escolas e empresas vêm adotando a solução de redes Wi-Fi para conectividade e mobilidade de seus usuários, a fim de diminuir custos com redes cabeadas e oferecer ao usuário a possibilidade de manter-se online, para acesso de informações, troca de mensagens e *streaming* de vídeo, não importando o local (TANG; BAKER, 2000).

Segundo Systems (2018) mais de 70% da população global terá conectividade móvel em 2023 e os dispositivos móveis crescerão de 8,8 bilhões em 2018 para 13,1 bilhões em 2023. A velocidade do Wi-Fi de dispositivos móveis triplicará até 2023, passando da média de 30,3 Mbps em 2018 para 92 Mbps até 2023. A Figura 1.1 mostra a tendência do aumento de velocidade para 2023.

Figura 1.1 – Métricas globais de rede sem fio



Fonte: adaptada de Systems (2018)

Haverá quase 628 milhões de *hotspots* de Wi-Fi públicos em 2023, contra 169 milhões de *hotspots* em 2018, um aumento de aproximadamente quatro vezes. Os pontos de acesso Wi-Fi 6 crescerão 13 vezes de 2020 a 2023 e serão 11% de todos os pontos de acesso Wi-Fi públicos até 2023 (SYSTEMS, 2018). Esse crescimento está ligado aos avanços e o novo padrão de rede sem fio, denominado Wi-Fi 6 (protocolo 802.11ax). Ambientes densos com muitos dispositivos de conexão simultânea e conexões de Internet das Coisas (IoT - *Internet of Things*), como aeroportos, transporte público, varejo, saúde, cidades inteligentes e estádios, resultarão em casos de uso de Wi-Fi público.

Com essa taxa de crescimento, a demanda por mais serviços e o aumento de mais dispositivos na rede elevam o consumo de banda, tendo-se a necessidade de aumentar o provisionamento de recursos (GANJI; BUDZISZ; WOLISZ, 2013). Com o novo padrão Wi-Fi 802.11ax (Wi-Fi 6), é possível conectar mais usuários a um mesmo ponto de acesso. Esse padrão aumenta a eficiência, a flexibilidade e a escalabilidade, o que permite que as redes atuais e as novas tenham mais velocidade e maior capacidade com aplicações de próxima geração. Isso também irá proporcionar uma melhor experiência de navegação ao usuário nas redes Wi-Fi (CISCO, 2021).

Para que seja possível projetar e dimensionar redes com uma melhor experiência de navegação é essencial a implementação de qualidade de serviço (QoS). O termo QoS (*Quality of Service*) está relacionado à habilidade da rede em oferecer um melhor serviço para um determinado tráfego em relação a outros tráfegos concorrentes. Isto é possível utilizando garantia de banda, no gerenciamento e prevenção de congestionamento da rede, na limitação de banda para determinadas classes de tráfego e na priorização de tráfego através da rede (MELO, 2005). Segundo Startdust (1999), em um ambiente compartilhado de rede, a qualidade de serviço está relacionada à reserva de recursos ou priorização de tráfego. A priorização de tráfego pode ser interpretada como um método para oferecer uma forma de tratamento prioritária para determinado fluxo da rede, ou seja, a capacidade de um elemento de rede ter algum nível de garantia de que seus requisitos de serviço e tráfego serão satisfeitos.

A QoS está relacionada às métricas largura de banda, latência, perda de pacotes, entre outros. A análise dessas métricas podem indicar se a rede atende aos requisitos desejados pelos usuários, como, por exemplo, uma chamada de vídeo de alta definição. Entretanto, métricas de QoS não permitem avaliar a experiência dos usuários sobre o desempenho da rede. Em um serviço de *streaming*, um erro de codificação pode causar a reprodução de um vídeo em baixa qualidade. Mesmo que as métricas de QoS indiquem uma boa qualidade da rede, o baixo nível de satisfação dos usuários poderia indicar a existência de um problema que não é perceptível através de parâmetros de QoS. Desse modo, para considerar a opinião dos usuários em relação ao serviço oferecido, outras métricas foram criadas e denominadas de QoE. A Qualidade de Experiência (QoE - *Quality of Experience*) é uma métrica que representa o desempenho de um serviço a partir da percepção dos usuários, com ênfase na satisfação de uso. Em redes de computadores, a QoE mostra a percepção do usuário sobre a qualidade de um serviço. As métricas de QoE referem-se à percepção do usuário, podendo ser expressas, como por exemplo, excelente, bom, satisfatório, ruim e muito ruim. As métricas de QoS que afetam o nível de QoE em redes Wi-Fi podem ser nível de sinal, velocidade do Wi-Fi, latência, *jitter* e performance dos dispositivos conectados ao ponto de acesso (SCHIMUNECK et al., 2014). Outro fator que pode afetar a performance dessas métricas

é a localização dos pontos de acesso, a distância até o usuário e densidade de pontos de acesso na rede/usuários.

1.1 Motivação

A elaboração de um plano de implantação de pontos de acesso adequado é necessário devido aos recursos de *hardware* desses dispositivos, que posicionados de forma errada podem limitar sua capacidade de comunicação e disponibilidade. Com um posicionamento correto de pontos de acesso, é possível maximizar a cobertura e tempo de vida, bem como minimizar custos de operação e consumo de energia. O comportamento humano pode influenciar na melhoria e manutenção da rede, uma vez que a demanda de acesso dos usuários (como por exemplo, quantidade de conexões do dispositivo com o ponto de acesso) podem sugerir a localização e quantidade de equipamentos (RODRIGUEZ-LOZANO et al., 2017).

1.2 Definição do problema

Mesmo com novas tecnologias, como o novo padrão de Wi-Fi 802.11ax, que melhora a conexão do usuário com mecanismos que aumentam a qualidade de serviço e a experiência do usuário, problemas de conexão, áreas de sombra e sinal fraco podem ocorrer. O congestionamento da rede em determinados pontos de acesso, tem se tornado um grande desafio para o gerenciamento de rede, à medida que a densidade de usuários móveis aumenta muito em locais específicos. Melhorar a alocação de recursos de acordo com um padrão de qualidade é um fator essencial para satisfação do usuário e aumentar o desempenho da rede.

Dado que as redes Wi-Fi tem um limite de usuários por AP (*Access Point*), e que a largura de banda é limitada, a qualidade de experiência do usuário pode ser afetada por falta de recursos. O dimensionamento adequado da localização e da quantidade de pontos de acesso, está relacionado com a quantidade de usuários, horário de utilização, área de cobertura desses equipamentos e largura de banda disponível na rede. Considerando que a área de cobertura dos APs e a largura de banda da rede são conhecidos, o problema é conhecer a quantidade de usuários conectados em cada local da rede dentro de um intervalo de tempo, para dimensionar corretamente a quantidade de equipamentos por região.

E como realizar o dimensionamento da quantidade de equipamentos por local?

1.3 Solução Proposta

Neste trabalho a solução proposta é investigar, analisar e tratar os dados de conexão dos usuários da rede Wi-Fi de um campus universitário para o correto dimensionamento da rede. Para isso, uma base de dados (*dataset*) da rede do campus foi gerada com dados de conexão dos usuários, com o tempo de conexão, tipo de usuário, local de conexão, entre outros. De posse dessas informações foi possível definir localização dos APs mais utilizados e tipo de acesso para tomada de decisões, como, por exemplo, melhorar o desempenho da rede e gerenciar o dimensionamento de recursos.

Foram analisados parâmetros como: duração da sessão de conexão (tempo durante o qual o usuário permaneceu conectado a um ponto de acesso), distribuição do usuário entre pontos de acesso, utilização de pontos de acesso por região e quando e com que intensidade os usuários usam a rede.

1.4 Objetivos

O objetivo deste trabalho é propor uma técnica para redimensionamento do número de APs de uma rede Wi-Fi universitária com base na análise de demanda e em critérios de QoS. Será gerado um *dataset* com os dados de conexões de usuários da rede Wi-Fi e com essas informações será possível realizar o dimensionamento mínimo de equipamentos para cada local, baseado em um consumo de banda por usuário.

Os objetivos específicos são:

- a) Criar *dataset* anonimizado a partir de *logs* de conexão e outras informações;
- b) Criar uma ferramenta gráfica para visualização espacial e temporal dos acessos por AP;
- c) Melhorar a gestão de redes Wi-Fi, através das informações do *dataset* criado;
- d) Realizar o dimensionamento de equipamentos para cada local, baseado em quantidade de usuário e de consumo de banda mínimo;

1.5 Estrutura do Trabalho

Este trabalho está organizado como descrito a seguir. No Capítulo 2 são mostrados conceitos fundamentais, características e fatores que envolvem dimensionamento de recursos de pontos de acesso. Os trabalhos relacionados são apresentados no Capítulo 3, o qual mostra as soluções encontradas na literatura. No Capítulo 4 é apresentada a metodologia utilizada no desenvolvimento do trabalho, como os cenários utilizados, diferentes topologias nas simulações e métricas utilizadas. Os resultados obtidos são descritos no Capítulo 5 e no Capítulo 6 são apresentadas as considerações finais do trabalho e trabalhos futuros.

2 REFERENCIAL TEÓRICO

O objetivo desse capítulo é explorar os principais conceitos relacionados à tecnologia Wi-Fi, com ênfase no padrão 802.11, no gerenciamento centralizado (*wireless controller*), na estrutura de dados baseadas no registro das informações em arquivos(Log) e no dimensionamento de número de pontos de acesso.

O padrão 802.11 é uma especificação de rede sem fio que define as características e requisitos para a comunicação entre dispositivos em uma rede Wi-Fi. O gerenciamento centralizado permite a administração de vários pontos de acesso a partir de um único ponto central, o que simplifica a gestão da rede. Entender a estrutura de dados é essencial para manipulação dos arquivos de Log. Por fim, o dimensionamento do número de pontos de acesso é uma etapa importante na implementação de uma rede Wi-Fi, uma vez que deve ser feito de forma adequada para garantir a cobertura e a capacidade necessárias para os usuários.

Neste sentido, o referencial teórico irá explorar os principais aspectos relacionados a esses conceitos, a fim de fornecer um embasamento teórico para a pesquisa e, conseqüentemente, permitir uma análise mais precisa e fundamentada do objeto de estudo.

2.1 Padrão IEEE 802.11

As redes locais sem fio são sistemas para interconexão de computadores com cobertura em uma área com raio de alcance pouco maior que uma centena de metros. A rede pode ser organizada a partir de uma base central de controle ou de forma espontânea, sem controle central, sendo, neste caso, denominada rede *ad hoc* (ROCHOL, 2018).

Essas redes são muito utilizadas atualmente no processo de comunicação entre diversos dispositivos móveis. O padrão IEEE 802.11 define uma arquitetura para as redes sem fio baseada na divisão da área coberta pela rede em células. O 802.11, assim como os outros protocolos da família 802.x, especifica as camadas físicas e de controle de acesso ao meio (MAC - *Medium Access Control*) e foi criado para definir uma norma para redes sem fio (LIMA et al., 2002).

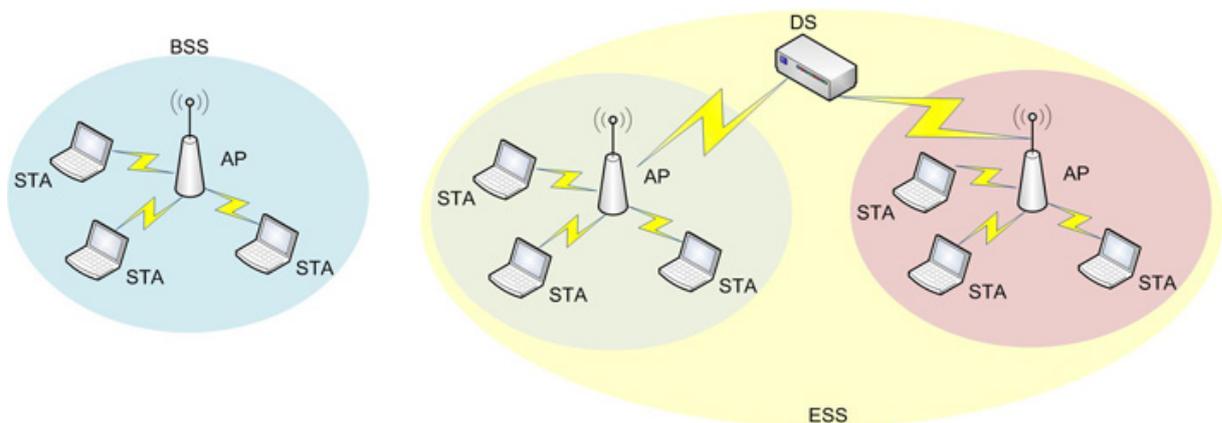
O núcleo básico de uma rede sem fio é composto por um conjunto de estações (STA - *Wireless LAN Stations*), cujo o acesso à rede é gerenciado por um AP (*access point*). As estações que se encontram na área de cobertura do AP formam o BSS (*Basic Service Set*) de um rede padrão IEEE 802.11 (ROCHOL, 2018).

Segundo Bonilha (2003) a topologia de uma rede sem fio (WLAN) IEEE 802.11 é composta por cinco elementos, descritos abaixo:

- a) BSS - *Basic Service Set* - Corresponde a uma célula de comunicação da rede sem fio. Nestas condições uma STA pode se movimentar de uma célula BSS para outra permanecendo conectada à rede. Este processo é denominado de *Roaming*;
- b) STA - *Wireless LAN Stations* - São os diversos clientes da rede;
- c) AP - *Access Point* - É o nó que coordena a comunicação entre as STAs dentro da BSS. Funciona como uma ponte de comunicação entre a rede sem fio e a rede convencional;
- d) DS - *Distribution System* - Corresponde ao backbone da WLAN, realizando a comunicação entre os APs;
- e) ESS - *Extended Service Set* - Conjunto de células.

Na Figura 2.1 é possível visualizar os os elementos de uma rede sem fio (WLAN) IEEE 802.11.

Figura 2.1 – Elementos de um rede 802.11



Fonte: do autor (2023)

O IEEE 802.11 definiu serviços que precisam ser oferecidos por uma rede sem fio para que tenha funcionalidades equivalentes as redes locais cabeadas. Os serviços básicos são de associação, desassociação e reassociação. Uma estação (STA), antes de começar a transmitir dados, precisa se conectar a um AP. Quando a estação se move para um área de cobertura de um novo BSS, ela faz uma reassociação com o novo AP. Ao final da sessão, a STA faz uma desassociação (ROCHOL, 2018).

Um dispositivo (smartphone, notebook, entre outros) com uma interface Wi-Fi escaneia periodicamente o ambiente para descobrir quais APs estão disponíveis. O dispositivo envia uma requisição de sondagem e espera as respostas dos APs ao redor, informando sobre a qualidade das conexões disponíveis. O dispositivo pode então enviar um pedido para se associar a um dos APs, ou para se reassociar a um novo se já estiver conectado à rede. O processo de reassociação é realizado quando um AP disponível apresenta uma conexão melhor que a atual. O dispositivo pode enviar uma solicitação de desassociação ao AP quando o sinal estiver muito fraco (FILHO, 2016).

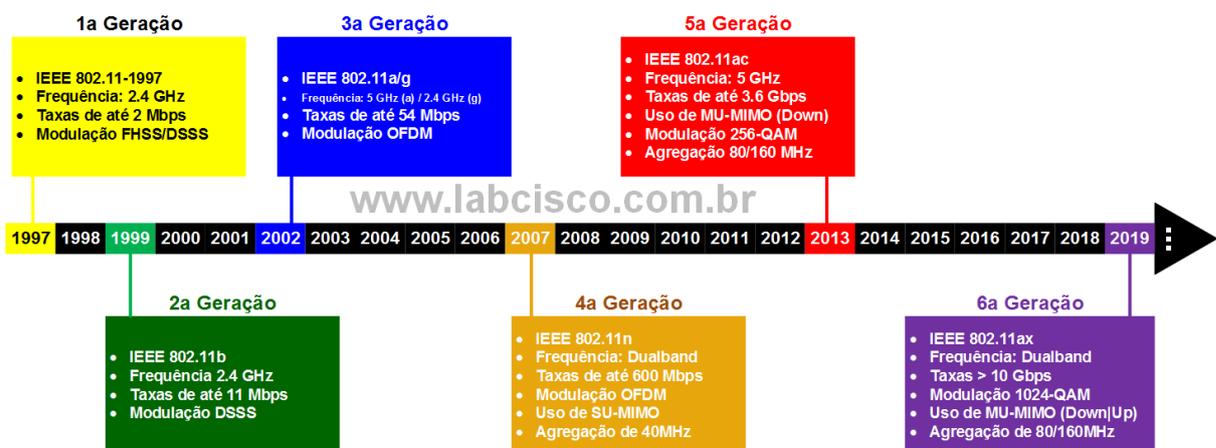
Introduzido pela primeira vez em 1999, os padrões IEEE 802.11 foram desenvolvidos principalmente tendo em mente o ambiente doméstico e de escritório para conectividade de rede local sem fio. O padrão 802.11b possui um velocidade máxima de transmissão de 11 Mbps por AP. Outras extensões como IEEE 802.11g e IEEE 802.11a, fornecem taxa de dados máxima de 54 Mbps por AP (SANCHES, 2011).

O padrão IEEE 802.11n, quando operando com 4 antenas no transmissor e no receptor, e utilizando a modulação 64-QAM (*Quadrature Amplitude Modulation*) e utilização de canais com 40 MHz de banda, permite chegar até os 600 Mbps (IEEE, 2009). O Padrão 802.11ac pode utilizar canais de 80 e 160 MHz e com velocidades que podem chegar até 1.800 Mbps com modulação QAM-256 (CISCO, 2018).

O novo Padrão 802.11ax (também conhecido como Wi-Fi 6) permite que os pontos de acesso suportem mais clientes em ambientes densos e forneçam uma melhor experiência para os usuários. A tecnologia OFDMA (*Orthogonal frequency-division multiple access*) permite que até mesmo pontos de acesso 802.11ax *Wave1* suportem oito fluxos espaciais e forneçam até 4.800 Mbps, dependendo da implementação do fabricante (CISCO, 2020). Vários dispositivos com várias necessidades de largura de banda podem ser atendidos simultaneamente, enquanto que no padrão anterior os dispositivos competiam entre si para enviar dados. Com o Wi-Fi 6, não há contenção, pois cada dispositivo é programado simultaneamente para transmitir dados em paralelo (ARUBA, 2020).

Na Figura 2.2 é possível visualizar a evolução dos padrão 802.11.

Figura 2.2 – Evolução do padrão 802.11



Fonte: adaptado de LabCisco (2023)

2.2 Gerenciamento centralizado

As implementações de rede sem fio (WLAN) tradicionais, consistem em pontos de acesso operando como nós separados e independentes que são configurados de forma autônoma. Esses equipamentos são gerenciados, operados e configurados de forma independente. À medida que a rede sem fio das empresas crescem, a capacidade de gerenciar a infraestrutura sem fio apresenta novos requisitos e desafios (NETGEAR, 2007).

Tradicionalmente, nas WLANs cada ponto de acesso opera como um nó separado, configurado de forma autônoma com configurações de canal e energia de um plano de RF (*Radio Frequency*) estático. Embora esses pontos de acesso autônomos escutem um ponto de acesso próximo operando no mesmo canal, o ponto de acesso autônomo não tem como determinar se o ponto de acesso adjacente faz parte da mesma rede ou de uma rede vizinha (CISCO, 2006).

As redes sem fio (Wi-Fi) são infraestruturas que provém mobilidade para acesso à internet e são muito utilizadas em ambientes universitários, onde a comunidade utiliza para fins acadêmicos e de pesquisa. Devido à sua importância as instituições realizam grandes investimentos para melhorar a cobertura, largura de banda, latência, segurança, consumo de energia e parâmetros de qualidade de serviço (QoS). As redes sem fio, com grande volume de equipamentos, são geralmente controladas e monitoradas por softwares centralizados que podem gerenciar os usuários e os recursos, como APs (RAJKUMAR et al., 2010).

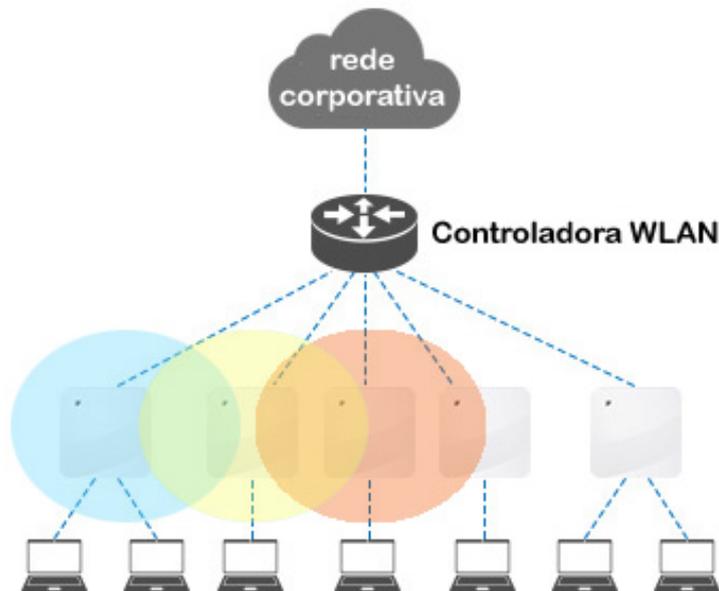
As redes locais sem fio visam atender às demandas de rede em constante mudança e podem ser gerenciadas a partir de uma base central de controle. Essa base central recebe o nome de controladora sem fio (*wireless controller*), que gerencia os pontos de acesso que permitem que dispositivos sem fio se conectem à rede. As controladoras facilitam o gerenciamento e segurança das redes sem fio (NETWORKS, 2015).

Uma controladora WLAN, monitora e gerencia pontos de acesso em massa e permite que dispositivos se conectem à rede sem fio. Como um dispositivo centralizador da rede sem fio, todos os pontos de acesso estão direta ou indiretamente conectados a ele. Assim é possível fornecer todos os serviços essenciais para reduzir o custo de implantação, simplificar o gerenciamento e adicionar várias camadas de segurança (NETGEAR, 2007).

Como a controladora WLAN conhece as características de RF dentro da infraestrutura de rede como um todo, ele pode detectar interferências entre APs próximos e reconfigurar suas opções de energia e canal automaticamente. Pelo mesmo princípio, quando o controlador detecta uma lacuna de cobertura devido a um AP instável, ele pode instruir os APs próximos a aumentar seus níveis de potência para

preencher a lacuna (NETGEAR, 2007). Na Figura 2.3 as cores mostram os sinais de RF que se sobrepõem e serão gerenciados pela controladora, para evitar interferências.

Figura 2.3 – Controladora WLAN



Fonte: adaptado de FSCommunity (2023)

Uma das principais vantagens das controladoras wireless é a possibilidade de gerenciamento centralizado de múltiplos APs, permitindo que os administradores de rede possam realizar configurações, atualizações e monitoramento de desempenho de maneira simplificada. Além disso, as controladoras wireless permitem a implementação de políticas de segurança de maneira eficiente, garantindo a proteção das informações trafegadas na rede sem fio (SMITH JAKE WOODHAMS, 2011).

Comparando a arquitetura de rede tradicional com a arquitetura de um rede com a controladora sem fio, tem-se as seguintes vantagens (HUAWEI, 2022):

- a) Implantação de configuração: Todas as configurações estão na controladora, não há necessidade de configurar cada AP individualmente;
- b) Ajuste inteligente de RF: O AP seleciona automaticamente o canal com menos interferência de acordo com o ambiente sem fio ao redor e ajusta a potência para melhorar a qualidade da rede sem fio;
- c) Gerenciamento e manutenção unificados: O gerenciamento centralizado faz com que os APs executem a mesma versão de *firmware* e, quando for necessário atualizar, todos os pontos de acesso são atualizados uniformemente;

- d) Balanceamento de carga do usuário: O balanceamento de carga otimiza a taxa de transferência para todos os clientes, otimizando constantemente as associações de usuários para fornecer a cada cliente uma taxa de transferência ideal, fazendo o roaming quando necessário.

A controladora também garante a continuidade da cobertura de sinal ao ajudar a rede sem fio a se recuperar de falhas de pontos de acesso. Se um ponto de acesso falha, a controladora é capaz de redirecionar o tráfego de rede sem fio para um ponto de acesso em funcionamento garantindo os requisitos de serviço da rede. Além disso, ela consegue reforçar a segurança da rede, impondo políticas de segurança em cada ponto de acesso e impedindo acessos não autorizados à rede sem fio, detectando e prevenindo a invasão de intrusos em caso de ataque de rede (RUCKUS, 2021). Essas informações de comunicação são registrados em arquivos de *logs* de conexão, que contêm informações sobre o horário, endereço MAC, identificação do AP conectado entre outras (FILHO, 2016).

2.3 Estrutura de dados de Log de conexão

A análise de dados trata-se de todo um processo, que inicia na definição do problema e passa pela extração de dados e sua análise e finaliza na obtenção da informação para resolução do problema inicialmente definido (BRIETZIG, 2022). Para Gil (2008) a análise de dados tem por objetivo organizar e sintetizar as informações para que seja possível conseguir as respostas ao problema proposto.

Existem diferentes tipos de análise de dados, como por exemplo a análise descritiva, a preditiva e a prescritiva. Na análise preditiva, os dados são utilizados para fazer projeções de cenários e identificar tendências a partir de determinados padrões. Na análise prescritiva, avalia-se as consequências que certas decisões podem trazer baseadas nas informações, essa análise auxilia na tomada de decisão. E a análise descritiva descreve certos eventos ou objetos que estão sendo analisados, comportamentos e tendências a partir de cenários realistas (SHARDA; DELEN; TURBAN, 2019). Nesta trabalho será utilizado a análise prescritiva.

O processo de análise de dados é composto por algumas etapas essenciais, como, definição do problema a ser resolvido, preparação e exploração dos dados, análise exploratória de dados, criação do modelo e apresentação dos resultados (DOMINGOS, 2019).

Para extração dos dados pode-se utilizar os arquivos de *log* de um sistema. O objetivo principal de um arquivo de *log* é registrar todas as atividades executadas e monitorar o status da infraestrutura de TI (DU et al., 2017). O arquivo de *log* do sistema também é usado como fonte elementar para identificar e solucionar problemas (MIZUTANI, 2013).

Para um arquivo de *log* ser capaz de prover as informações para tomada de decisões seja em dimensionamento, segurança, análise de mobilidade ou análise de padrões de comportamento, é essencial que os dados estejam estruturados e preparados para análise. Um arquivo de *log* são grupos de dados não estruturados que são coletados de diferentes fontes, como informações de sistema, dados de conexões, informações de configuração e de aplicativos (VALDMAN, 2001).

Os arquivos de *logs* do sistema normalmente são enormes e estão disponíveis em um formato não estruturado. Assim, há uma necessidade de realizar pré-processamento de informações para melhor compreensão e recuperação de informações significativas de dados de arquivos de *logs* complexos (DUAN et al., 2021).

Segundo os autores Bhanage, Pawar e Kotecha (2021), o arquivo de *log* coletado, na grande maioria das vezes, está sempre em um formato não estruturado, duplicado e ambíguo. O pré-processamento é muito importante antes de enviá-lo para análise. O pré-processamento inclui três etapas: 1) Filtragem: que remove os dados duplicados e ruidosos, 2) Análise: que converte o *log* não estruturado em um formato estruturado; e 3) Análise: que visualiza o *log* em um formato mais legível e compreensível.

Os autores Hassani et al. (2018) afirmaram que às vezes as mensagens do arquivo de *log* não são confiáveis. Elas podem conter erros como informações impróprias, falta de instruções, nível de informações insatisfatória, problemas de estrutura nos arquivos, problemas de tempo de execução e alterações na biblioteca de arquivos. Assim, antes de pré-processar os dados, é preciso validar sua qualidade.

Com os avanços na capacidade de armazenamento dos servidores e na automatização da coleta de dados, o tamanho dos bancos de dados aumentaram consideravelmente. Isso tem dificultado a extração de informações e a análise manual dos dados (KONONENKO; KUKAR, 2007).

Os administradores de rede analisavam os arquivos de *log* manualmente para compreender o comportamento do sistema. No entanto, devido ao aumento da complexidade e a quantidade de dados, o monitoramento da infraestrutura de TI vem exigindo o monitoramento automatizado do sistema (JAIN et al., 2009).

Então, para realizar o dimensionamento de quantidade de pontos de acesso na rede, é necessário realizar uma análise e fazer o pré-processamento dos arquivos de Log coletados.

2.4 Dimensionamento de pontos de acesso

Diversos parâmetros envolvidos no contexto físico das redes podem ser otimizados, isoladamente ou em combinação, buscando maximizar a cobertura e o tempo de vida dos equipamentos, bem como minimizar os custos de operação (número de dispositivos) e consumo de energia (BASAGNI et al., 2009).

Portanto, é imprescindível ao usuário se associar a um ponto de acesso adequado para melhorar sua qualidade da experiência (QoE). O problema de selecionar um nó de destino para um usuário é denominado problema de seleção de ponto de acesso (KOBAYASHI et al., 2018).

Um AP e seus usuários associados compartilham um recurso de largura de banda limitado e os APs podem ficar sobrecarregados, levando a congestionamento e baixo desempenho da rede. Por outro lado, um usuário pode se comunicar com sucesso com dois ou mais APs se existir uma controladora que os gerencie. Assim, é desejável possuir uma rede sem fio capaz de distribuir associações de clientes entre APs de maneira balanceada, de modo que nenhum equipamento seja indevidamente sobrecarregado. Redes sem fio que possuam esta capacidade, melhoraram consideravelmente o desempenho da rede (KBAR; MANSOOR, 2006).

Do ponto de vista técnico, o problema da seleção de AP é atribuir um usuário a um AP adequado, entre as redes de acesso disponíveis, para otimizar uma determinada função. As atuais políticas de associação são baseadas principalmente no comportamento do usuário egoísta (GERASIMENKO et al., 2013). Esse comportamento egoísta dos usuários pode ser comparado com uma estratégia de seleção baseada na intensidade do sinal. Considera-se a taxa de transmissão distinta de cada usuário e a influência da associação de cada usuário na taxa de transferência recebida de outro usuário (ARYAFAR et al., 2013).

Os usuários da rede têm necessidades diferentes, uma vez que os aplicativos possuem seus requisitos específicos de qualidade de serviço (QoS). Consequentemente, quando os usuários selecionarem um AP com a melhor taxa de transferência, esse AP pode ficar indisponível devido a sobrecarga do número de conexões. Os usuários que necessitam de melhor sinal e banda serão prejudicados, porque usuários com pouca necessidade de qualidade de banda estão conectados nesse AP (ÇEKEN; YARKAN; ARSLAN, 2010).

A noção de sustentabilidade é um fator significativo para avaliar o efeito das seleções de AP, bem como a utilidade mínima, porque os recursos de rede são limitados, especialmente em situações de congestionamento. O congestionamento da rede na seleção de AP tem se tornado um desafio importante e fundamental para o gerenciamento de rede, à medida que a densidade de usuários móveis aumenta muito em alguns locais específicos (KOBAYASHI et al., 2018).

À medida que os usuários se movem dentro da rede sem fio, o número de usuários atendidos por um ponto de acesso pode variar drasticamente, de um número muito baixo a um número muito alto. Quando um AP é requisitado a atender a um grande número de usuários, é provável que fique sobrecarregado e o congestionamento resultante degrade significativamente o serviço recebido pelos usuários dependentes desse AP. O congestionamento reduz a largura de banda experimentada pelo usuário (KUO; LIAO, 2008).

Normalmente, há duas partes de um projeto de dimensionamento de rede sem fio: seleção de locais de AP e atribuição de canais de rádio aos APs. O projeto geralmente é baseado em medições de intensidade de sinal, quantidade de APs e na consideração de problemas de propagação de rádio. A primeira parte do processo, seleção dos locais AP, deve ser feita para fornecer cobertura completa do espaço-alvo sem sobreposição indevida de cobertura. A segunda parte é a quantidade de APs e as características do ambiente de propagação de rádio no qual a rede sem fio está sendo implantada, que pode ser difícil, mas é importante em um projeto de rede sem fio (HILLS, 2001).

O projeto de construção de um rede sem fio é medir informações suficientes para determinar o número e a localização dos pontos de acesso que fornecem cobertura adequada aos usuários. Basicamente, o procedimento envolve a implantação de APs temporários em locais preliminares, um único AP por vez ou uma rede sem fio inteira é construída temporariamente com base na opinião e nas experiências de um projetista. Em seguida, a cobertura e/ou Qualidade de Serviço (QoS) é examinada usando a medição na pesquisa de campo (*site survey*). Com base nos resultados, as posições e configurações dos APs são alteradas ou novos APs são introduzidos e então, novamente, uma nova pesquisa de campo é repetida para encontrar uma solução aceitável de forma iterativa (ZVANOVEC; PECHAC; KLEPAL, 2003).

Em redes sem fio o equilíbrio ideal entre o número de usuários, o número de APs e a quantidade de largura de banda disponível são componentes críticos para um bom desempenho da rede (AGGARWAL; KOKSAL; SCHNITER, 2013). Para conseguir ter um bom desempenho da rede é preciso que seja realizada análise de dados para auxiliar o processo de tomada de decisões.

Após a apresentação dos conceitos fundamentais para o entendimento do projeto, no próximo capítulo são apresentados os trabalhos relacionados presentes na literatura.

3 TRABALHOS RELACIONADOS

Nesta seção são apresentados os trabalhos relacionados a essa pesquisa. Diversos trabalhos presentes na literatura buscam realizar a melhor forma de projetar rede WLAN para dimensionamento de recursos de forma eficiente.

O trabalho desenvolvido por Bahri e Chamberland (2005) aborda o problema de projeto de rede sem fio com garantias de desempenho. Este problema consiste em selecionar conjuntamente a localização dos pontos de acesso, bem como a potência e o canal de cada AP. As garantias de desempenho ao que os autores se referem, são garantias de cobertura e largura de banda mínima. O objetivo é minimizar o número de APs a serem instalados na rede. Isso ocorre porque o custo da LAN (incluindo a rede de acesso sem fio e a rede de distribuição), bem como o custo do gerenciamento da rede, dependem do número de APs. Para isso propuseram um modelo de otimização para selecionar a localização, bem como a potência e o canal de cada AP. Foi apresentada uma formulação matemática do problema, uma heurística para gerar soluções iniciais e um algoritmo de busca tabu para melhorar essa solução inicial. Para conseguir otimizar o número de equipamentos e aumentar o número de estações móveis conectadas foi preciso respeitar uma largura de banda mínima por estação de 500 kbps, que é considerada muito pequena para demanda atual.

Segundo Hills (2001), uma rede sem fio de grande escala deve ser projetada para que a cobertura de rádio cubra todo o campus e que sua capacidade seja adequada para suportar a carga de tráfego esperada. No planejamento da instalação de rede sem fio com vários APs, deve-se tomar cuidado para garantir que a cobertura de rádio adequada seja fornecida em toda a área de serviço, localizando cuidadosamente os APs. Para o correto dimensionamento do projeto de WLAN foram utilizados dois componentes: seleção da localização do AP e atribuição de frequências de rádio aos APs. Eles demonstraram que o leiaute deve ser baseado em medições, que envolvem testes extensivos e considerações dos problemas de propagação de rádio quando a área de serviço é grande, como um campus inteiro. O desempenho da rede também é um problema, já que há uma quantidade finita de largura de banda disponível e ela deve ser compartilhada pelos APs e computadores móveis. Foi desenvolvido uma técnica para seleção inicial de localizações de ponto de acesso. Essa seleção inicial é provisória e deve ser modificada na segunda etapa do processo. Os APs são temporariamente instalados nos locais selecionados e as áreas de cobertura e as sobreposições entre as áreas de cobertura são medidas. Normalmente, lacunas de cobertura e/ou sobreposições excessivas são encontradas. Com base nos resultados da medição, as localizações AP são ajustadas conforme necessário, mais medições são feitas, mais ajustes são feitos e assim por diante, até que um projeto aceitável é encontrado. O processo é iterativo. Pode ser necessário repetir esse ciclo de projeto-teste-reprojeto várias vezes para encontrar uma solução aceitável.

Depois que as localizações finais do AP foram selecionadas, um mapa de cobertura da área de projeto foi criado e as frequências são atribuídas aos APs de forma a minimizar a sobreposição de cobertura co-canal. Em seguida, é feito um conjunto completo de medições de cobertura (auditoria) para todo o edifício com os APs operando nas frequências selecionadas, e os resultados dessas medições são documentados. Neste ponto, o projeto é considerado completo e o mapa de cobertura é atualizado para refletir as localizações finais de AP, medições de cobertura e atribuições de frequência. Esse método é bem trabalhoso por requer que sejam instalados os APs de forma provisória e depois reinstalados diversas vezes até seu perfeito funcionamento, gerando gastos em infraestrutura e mão de obra.

Segundo os autores Gibney, Klepal e Pesch (2010), uma implantação de rede local sem fio (WLAN) deve-se considerar muitas métricas de projeto, como, configuração do ambiente, restrições de cobertura, atribuição de frequência, interferência de co-canal, demandas do usuário na WLAN, tipos de aplicativos, número de pontos de acesso e sua posição ao realizar um projeto e o uso de um ferramenta para esse dimensionamento. Assim, foi desenvolvida uma ferramenta de modelagem, projeto e avaliação de rede para auxiliar os projetistas ao realizar a tarefa de dimensionamento de rede WLAN. A ferramenta desenvolvida incorpora os componentes necessários para otimizar automaticamente o posicionamento do ponto de acesso para atender às demandas específicas do local. Um componente chave da ferramenta é a incorporação de um algoritmo de otimização baseado em elementos de inteligência artificial.

Para testar a ferramenta desenvolvida, foi realizado um projeto piloto em um edifício de 125 x 67 metros com um grande número de salas de aula, escritórios e laboratórios. Foi estimado 500 usuários com conectividade sem fio e cada usuário executou aplicativos operando em alto uso. Um total de 2.651 posições de pontos de acesso candidatos foram gerados usando um algoritmo padrão e usados como entrada para os algoritmos de otimização. O número mínimo de pontos de acesso para atender às demandas do usuário é de 25, calculado com o número de usuários e *throughput* máximo disponível por ponto de acesso. A grande vantagem da abordagem foi o nível de escalabilidade que pode ser alcançado. O algoritmo do agente produz uma solução melhor do que o padrão e tem uma satisfação de demanda do usuário maior com um número significativamente menor de pontos de acesso. Com essa distribuição de recursos se tem uma rede mais equilibrada com bordas mais suaves entre as células de cobertura do ponto de acesso.

Para definição do projeto de rede sem fio, os autores Mateus, Loureiro e Rodrigues (2001) focaram seu trabalho em duas grandes áreas: determinar a localização das estações base (BS) e atribuir os canais de frequência para essas estações. O trabalho seleciona um conjunto de BS para fornecer a melhor cobertura de área de demanda e maximizar o nível de sinal e a prioridade de atendimento da área física. A atribuição de canal adequada reduz a interferência de sinal e melhora o rendimento da rede.

Foi proposto um modelo matemático, implementado utilizando a linguagem AMPL integrada ao CPLEX (pacotes de software usados para resolver problemas de otimização). Foi necessário medir/prever o sinal em cada ponto da área de demanda, localizar as BS e então atribuir um canal de frequência. Com dimensionamento proposto, o ambiente não satura quando instalados vários BS, pois há uma redução da área de cobertura e controle de frequências. Desta forma, se reduz o número de usuários atendidos por cada BS e, conseqüentemente, uma melhora do *throughput*.

Os autores Gondran et al. (2007) propuseram uma solução para o problema de implantação em larga escala de redes locais sem fio (WLAN) através de um modelo que integra todos os parâmetros úteis para estimar a localização e as configurações de parâmetros dessa rede, otimizando os custos e a taxa de transferência da rede. Os parâmetros utilizados foram a localização do ponto de acesso, o padrão do ponto de acesso, a orientação da antena, a potência emitida e o canal de frequência da antena. Todos esses itens são definidos neste modelo que reúne recursos de planejamento de célula e frequência (mistura de Planejamento Automático de Célula (ACP) e Planejamento Automático de Frequência (AFP)). A vantagem deste modelo é que o cálculo de avaliação de rede é feito de uma única maneira, desde a potência recebida até o *throughput* real, através da relação SNR (relação sinal e ruído), incluindo todos os canais de frequência disponíveis do sistema WLAN e cálculo de múltiplas interferências. No final, o problema de otimização foi minimizar o custo econômico de implantação e a taxa de *throughput*.

Em cenários de alta densidade, as redes locais sem fio (WLANs) encontram sérias interferências. O alinhamento de interferência é uma maneira eficaz de controlar a interferência e melhorar a capacidade do sistema no canal de interferência. No entanto, os canais de interferência interagem uns com os outros nas formas de planejamento de rede existentes e a taxa de transferência da rede não pode ser melhorada significativamente por meio do alinhamento de interferência. Para controlar a interferência e melhorar o *throughput* da rede de forma mais eficiente, os autores Peng et al. (2019) propuseram um esquema de planejamento de rede baseado no alinhamento de interferência em WLANs de alta densidade. Toda a rede é dividida em uma pluralidade de sub-regiões, e as sub-regiões vizinhas usam diferentes canais. Cada sub-região é considerada como um canal de interferência e o alinhamento é utilizado para controlar a interferência em uma única sub-região. Enquanto isso, a interferência entre sub-regiões é controlada pela alocação de canais. O tamanho apropriado das sub-regiões e a alocação ótima do canal serão selecionados de acordo com o princípio de maximizar o grau de liberdade de toda a rede. O grau de liberdade descreve o número de mensagens transmitidas ao mesmo tempo, portanto, dessa forma, tornará as WLANs de alta densidade mais competitivas. Nas simulações, a forma de planejamento de rede proposta é comparada à forma tradicional em diferentes escalas das WLANs, e os resultados mostram que a forma proposta melhora o grau de liberdade média em WLANs de alta densidade.

Os autores Wang et al. (2019) usaram o *ELK Stack (Elasticsearch, Logstash e Kibana)* como um software integrado para análise de arquivos de *log* de uma rede Wi-Fi. Realizaram a coleta, filtragem e armazenamento de dados e exibiram as informações como um painel completo que os dados podem ser facilmente analisados. Nesse painel é possível inspecionar, descobrir a causa da falha, monitorar em tempo real, detectar problemas, ter acesso a análise histórica (que pode identificar tendências e padrões), facilitar o rastreamento do acordo do nível de serviço (SLA), ter estatísticas em tempo real e gráficos intuitivos no processamento de dados para correto dimensionamento de equipamentos.

Foi proposto um modelo para o problema de dimensionamento de rede local sem fio (WLAN) com garantias de desempenho (cobertura e largura de banda mínima) pelos autores Kbar e Mansoor (2006). A proposta consiste em selecionar a localização dos pontos de acesso (APs), bem como a potência e o canal de cada AP através de uma heurística para gerar soluções iniciais e um algoritmo de busca tabu para encontrar soluções para instâncias do problema. Os resultados obtidos demonstram que boas soluções podem ser encontradas com o algoritmo proposto em um tempo razoável.

Em Kamenetsky e Unbehaun (2022), foi investigado as melhorias de desempenho que podem ser obtidas otimizando o posicionamento dos pontos de acesso sem fio para gerar melhor qualidade de serviço ao usuário. Foi proposto um novo algoritmo de poda, que tem convergência garantida e independente das condições iniciais, pode ser usado para fornecer uma solução que pode servir como ponto de partida para algoritmos de busca mais complexos. Simulações numéricas em um ambiente, mostraram que dos custos de infraestrutura (ou seja, número de APs) podem ser economizado em comparação com uma rede dispersa regular, onde as posições de AP são selecionadas uniformemente dentro do ambiente. Embora esses algoritmos garantam o melhor desempenho possível para uma determinada estrutura de rede, eles não melhoram a topologia em si, ou seja, não podem compensar APs mal posicionados ou corrigir lacunas de cobertura.

Os autores Zhong et al. (2015) realizaram um estudo baseado em simulações para testar como a distância entre dois APs afeta a largura de banda de uma rede sem fio e consequentemente o dimensionamento de equipamentos. Verificou-se que em uma implantação com sobreposição de serviços (*Overlapping Basic Service Set - OBSS*), vários efeitos, como supressão de link, amplificação de interferência, co-canal, transmissões gerando impasse temporário de APs vizinhos e nós ocultos/expostos, degradam significativamente o desempenho da taxa de transferência. Em tais implantações, apenas aumentar a concentração de APs em uma área pode não necessariamente trazer melhorias de desempenho, por isso as colocações de APs devem ser cuidadosamente gerenciadas para mitigar os efeitos mencionados acima.

Nos artigos analisados, os autores utilizaram diversas técnicas para correto dimensionamento de equipamentos em redes Wi-Fi. Dentre esses pode-se relatar:

- a) Localização de APs, baseados em frequência, potencia e cobertura de sinal;
- b) Análise de Log de rede Wi-Fi para gerenciamento e dimensionamento;
- c) Dimensionamento de equipamentos baseado em cobertura e largura de banda mínima;
- d) Posicionamento de APs para melhorar a qualidade de serviço.

No Quadro 3.1 está apresentado um resumo dos trabalhos relacionados.

Quadro 3.1 – Resumo dos trabalhos relacionados

Artigo	Objetivo	Método	Ano
On the wireless local area network design problem with performance guarantees	Dimensionamento de projeto WLAN	Localização, potência e frequência do AP	2005
Large-scale wireless LAN design	Dimensionamento de projeto WLAN	Localização e frequência do AP	2001
Wireless local area network modeling tool for scalable indoor access point placement optimization	Dimensionamento de projeto WLAN	Ferramenta de modelagem	2010
Optimal Network Design for Wireless Local Area Network	Dimensionamento de projeto WLAN	Localização e frequência do AP	2001
Wireless LAN planning: a didactical model to optimise the cost and effective payback.	Dimensionamento de projeto WLAN	Localização, potência e frequência do AP	2007
Distributed resources management in wireless LANs	Dimensionamento de projeto WLAN	Localização, potência e frequência do AP	2006
On indoor position location with wireless LANs	Melhorar desempenho da rede	Controle de interferência	2019
The Implementation of Wi-Fi Log Analysis System with ELK Stack	Melhorar desempenho da rede	Análise de dados	2019
Coverage planning for outdoor wireless LAN systems	Melhorar desempenho da rede	Localização de AP	2022
Issues and challenges in dense WiFi networks	Melhorar desempenho da rede	Localização de AP	2015
Dimensionamento de rede sem fio de um campus universitário baseado na análise de LOGs de conexão	Dimensionamento e melhora de desempenho de rede	Dimensionamento de AP e throughput	2023

Este trabalho se difere dos citados acima, principalmente por que não se concentrou em prever a localização de instalação do ponto de acesso ou controlar as frequências dos canais, mas sim utilizar a informação do local da conexão do usuário para realizar o dimensionamento de equipamentos e aumentar a qualidade de serviço e qualidade de experiência ao usuário. Foi realizado a coleta e tratamento (anonimização, mascaramento e remoção de dados sensíveis) dos dados das conexões dos usuários da rede Wi-Fi do Campus, relacionamento dos dados de usuários com outras fontes de dados para obter dados de categoria de usuários, relacionamento dos dados de local com coordenadas para visualização geográfica em mapas, apresentação das análises georreferenciadas, geração de um *dataset* público e dimensionamento de equipamentos por local.

4 MATERIAIS E MÉTODOS

Este capítulo apresenta os materiais e métodos utilizados para o desenvolvimento deste trabalho, tais como a base de dados das conexões Wi-Fi, o processo utilizado para tratamento dos dados, as estratégias utilizadas para a extração das informações, as ferramentas utilizadas para a representação gráfica do ambiente e o método para o dimensionamento de recursos.

Os arquivos de *logs* de conexão de uma rede sem fio de um campus universitário foram analisados durante um período de quatro meses. Esta rede possui aproximadamente 350 pontos de acesso de alto desempenho, com gerenciamento centralizado, sendo utilizada por alunos de graduação e pós-graduação, técnicos administrativos e professores, com um total superior a quinze mil usuários.

Com o resultado desta análise foi criado um *dataset*, com o qual é avaliado o número de conexões dos usuários dentro da rede, sendo possível classificá-los em categorias e contabilizar o número de conexões por local a fim de realizar o dimensionamento de equipamentos baseados em uma largura de banda mínima por usuário.

Uma base de dados com as informações dos arquivos de *logs* foi construída, utilizando algoritmos em Python e consultas SQL em banco de dados Postgres para tratamento dos dados. Como existem dados sensíveis (*logins*) dos usuários na base de dados, foi necessário remover essas informações. A base de dados gerada foi totalmente anonimizada e seus dados mascarados para ser utilizada neste trabalho.

4.1 Controladora de rede sem fio

O campus analisado possui uma controladora do fabricante Ruckus Networks (ZoneDirector) para gerenciamento da rede sem fio. Com a utilização da controladora, é possível centralizar a implantação da rede, as operações, o gerenciamento e monitoramento de todos os pontos de acesso sem fio da rede. A controladora possui funcionalidades como descobrir novos pontos de acesso, configurá-los e ajustar automaticamente os níveis de potência de transmissão e as atribuições de canais de rádio frequência (RF), conforme a necessidade, para evitar ruídos, impedir a interferência de APs adjacentes e permitir cobertura redundante na eventualidade de uma falha de AP. Isso significa que a controladora otimiza o desempenho da rede sem fio com pouca intervenção humana.

A coleta de dados foi realizada por meio dos arquivos de *logs* de conexão que são gerados pelas conexões nos APs, gerenciados pela controladora *Zone Director*. Foram utilizadas informações de mais de trezentos e cinquenta pontos de acesso instalados no Campus que são armazenados em um servidor Radius (RADIUS, 2022).

4.2 Arquivo de log de conexões

O arquivo de *log* contém dados de quatro meses de conexão dos seus usuários (docentes, discentes, servidores administrativos e usuários de setores) em diversos locais como biblioteca, pavilhões de aulas, cantina, departamentos e setores. Toda associação dos usuários aos APs da rede do campus são gravadas em um servidor Radius (RADIUS, 2022) que armazena diversas informações tais como: endereço MAC (*Media Access Control*), login do usuário, quantidade de bits trafegados, ponto de acesso que o dispositivo está conectado, data e hora, tempo de conexão, entre outros.

O endereço MAC é um identificador exclusivo associado a uma interface de rede de um dispositivo que é utilizado como um endereço físico em uma rede Wi-Fi. As redes Wi-Fi são compostas por conjuntos de *access points* aos quais um dispositivo pode se conectar usando seu endereço MAC.

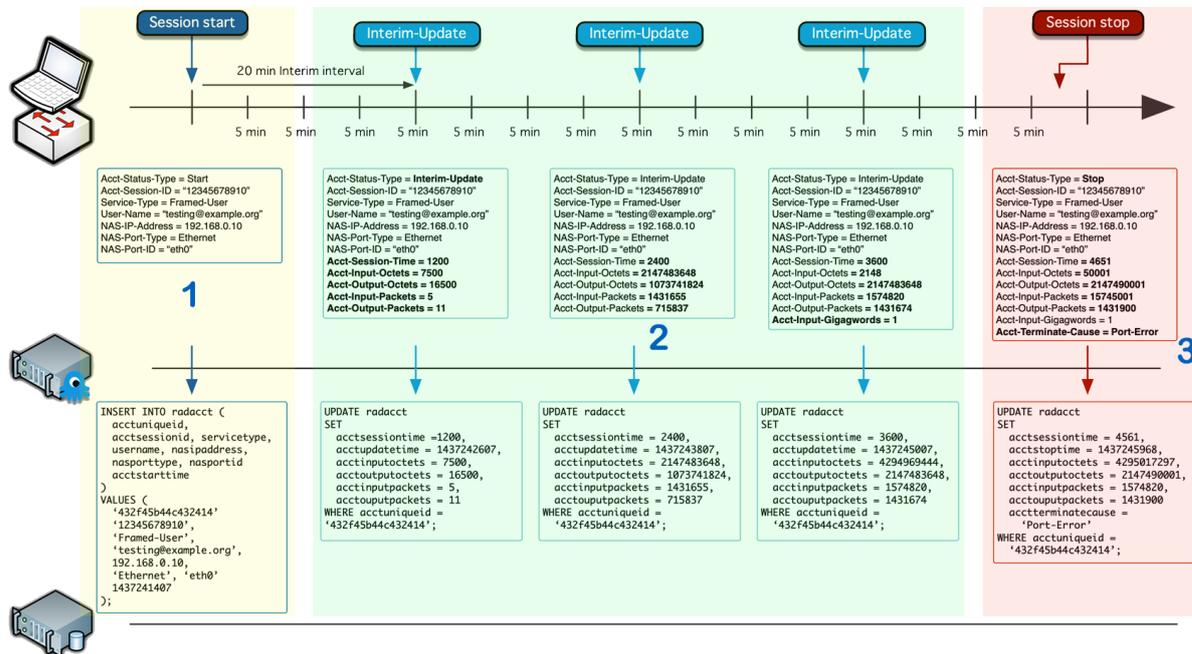
O servidor Radius fornece gerenciamento centralizado de autenticação, autorização e contabilização para usuários que conectam-se à rede e utilizam um serviço de rede (FREERADIUS, 2022). O módulo de autenticação é responsável por autorizar usuários ou dispositivos antes da concessão de acesso a rede. O módulo de autorização gerencia se outros usuários ou dispositivos podem usar determinados serviços providos pela rede e o módulo de contabilização (*accounting*) coleta dados para fins estatísticos e monitoramento de rede.

Quando o acesso é concedido, o servidor de acesso à rede (NAS) envia um pacote de solicitação ao Radius, que significa que o acesso do usuário à rede foi iniciado. Este pacote contém o identificador do usuário, o ponto de acesso e o endereço de rede, bem como um identificador de sessão exclusiva. Esse processo é chamado de *START*, conforme mostrado em 1 da Figura 4.1. O NAS também pode enviar um pacote de solicitação de contabilidade conhecido *INTERIM UPDATE*, mostrado em 2 da Figura 4.1, que consiste na duração da sessão e no uso de dados atual. Este pacote atualiza o status da sessão do usuário para o servidor Radius.

Assim que o acesso do usuário ao servidor Radius cessar, o NAS enviará um pacote de solicitação de contabilidade adicional, conhecido como registro *STOP*, conforme apresentado em 3 da Figura 4.1. Este registro fornece informações abrangentes da sessão, como tempo total, dados e pacotes transferidos, o motivo da desconexão e qualquer material adicional relevante para a sessão do usuário.

Finalmente, o servidor Radius responde com uma confirmação de resposta de contabilização. Nesse ponto, o processo de contabilização de um determinado usuário está concluído e todas as informações relacionadas a essa sessão são armazenadas com segurança no servidor (RADIUS, 2021). Todo o processo pode ser observado na Figura 4.1.

Figura 4.1 – Diagrama de conexão do Radius



Copyright © 2020 - Network RADIUS SARL

Fonte: adaptado de Radius (2021)

Os arquivos de *logs* de conexão são gerados diariamente e cada arquivo contém em média 17.000.000 (dezesete milhões) de linhas de dados. Dentre essa linhas temos os registros de "START", "STOP" ou "UPDATE", com um tamanho aproximado de 500 MB cada arquivo.

O conjunto de dados utilizado para a análise foi obtido com a junção de todos os arquivos gerados entre os meses de junho a setembro de 2022 equivalente ao período letivo do primeiro semestre de 2022. Foram processados mais de 120 arquivos com informações de conexão, totalizando mais de 60 GB de dados. Cada registro inclui um grande número de parâmetros e referem-se as conexões realizadas pelos usuários nos mais de trezentos e cinquenta pontos de acesso espalhados pelo Campus da Universidade.

O formato do arquivo *log* de conexão é apresentado na Figura 4.2:

Figura 4.2 – Conteúdo do arquivo de *Log*

```

Tue May 24 10:45:06 2022
User-Name = "thiagoramos"
Acct-Status-Type = Start
Acct-Authentic = Remote
Framed-IP-Address = 172.16.25.126
Calling-Station-Id = "48-2C-A0-AC-34-FF"
NAS-IP-Address = 192.168.14.2
NAS-Port = 144
Called-Station-Id = "2C-5D-93-06-D1-88:UFLA+"
NAS-Port-Type = Wireless-802.11
NAS-Identifier = "2C-5D-93-06-D1-88"
Connect-Info = "CONNECT 802.11g/n"
Acct-Session-Id = "5DABCFC6-0106AAA8"
Acct-Multi-Session-Id = "2c5d9306d188482ca0ac34ff5ddf53fd9df4"
Ruckus-SSID = "UFLA+"
Event-Timestamp = "May 24 2022 10:45:06 BRST"
Acct-Unique-Session-Id = "8570501e9ff9f3ba"
Timestamp = 1653399906
Request-Authenticator = Verified

```

Fonte: do autor (2022)

A descrição dos campos do arquivo de *Log* do Radius, gerado pela conexão dos usuários nos APs gerenciados pela controladora Ruckus é apresentada a seguir:

- a) **User-name:** Nome do usuário (login);
- b) **Acct-status-Type:** Define a mensagem de início ou término de sessão;
- c) **Acct-Authentic:** Indica se o servidor Radius é local ou remoto;
- d) **Framed-IP-Address:** IP de origem do cliente;
- e) **Calling-Station-ID:** Endereço MAC do dispositivo cliente;
- f) **NAS-IP-Address:** IP do dispositivo (servidor);
- g) **NAS-PORT:** Tipo de porta virtual utilizada;
- h) **Called-Station-Id:** Endereço MAC do AP e o nome da rede que o cliente conectou;
- i) **NAS-Port-Type:** Tipo de conexão (Wi-Fi, Ethernet, etc);
- j) **NAS-Identifier:** Endereço MAC do AP que o cliente conectou;
- k) **Connect-Info:** Informações do padrão de conexão (exemplo: 802.11ac);
- l) **Acct-Session-Id:** Este atributo é usado para identificar exclusivamente uma sessão para que os registros de parada e início possam ser intercalados e registrados com precisão;

- m) **Acct-Multi-Session-Id:** Este atributo é um ID de contabilização exclusivo para facilitar a vinculação de várias sessões relacionadas em um arquivo de *log*;
- n) **Ruckus-SSID:** Nome da rede que o cliente conectou;
- o) **Event-Timestamp:** Data e hora da conexão;
- p) **Acct-Unique-Session-Id:** Utiliza um *hash* MD5 que é usado como o valor para o ID de sessão exclusivo;
- q) **Timestamp:** Este atributo é incluído em um pacote *Accounting-Request* para registrar a hora em que esse evento ocorreu (em segundos).

O arquivo de *log* não possui explicitamente a informação em qual AP o usuário realizou a conexão. Esse arquivo possui o campo *NAS-Identifier* que tem o MAC da interface de rede sem fio que o usuário se conectou. Então, foi preciso extrair da controladora a informação do local do AP. Porém, na listagem da controladora é exibido somente o MAC da interface LAN, conforme Figura 4.3:

Figura 4.3 – Relatório da controladora com associação do MAC

MAC Address	Device Name	Model	Status
2c:5d:93:06:b9:c0	ALMOXARIFADO_01	zf7372	Connected

Fonte: do autor (2022)

Diante disso, foi preciso gerar um arquivo que relacionasse os endereços MACs das interfaces da rede sem fio e o endereço MAC da interface LAN de cada AP com sua localização dentro do Campus. A Figura 4.4, apresenta o MAC das interfaces de rede do AP denominado almoxarifado.

Figura 4.4 – Arquivo com MAC e local do AP

MAC Address,Device Name
2c-5d-93-06-b9-c0,ALMOXARIFADO
2c-5d-93-06-b9-c8,ALMOXARIFADO
2c-5d-93-06-b9-cc,ALMOXARIFADO

Fonte: do autor (2022)

Na Figura 4.3 temos a informação que o MAC 2C:5D:93:06:B9:C0 pertence a interface de rede LAN do AP denominado Almoxarifado. O arquivo de *log* mostra conexões de usuários nos MACs 2C:5D:93:06:B9:C8 e 2C:5D:93:06:B9:CC que são referentes a interface de rede sem fio do AP. A informação dos endereços MAC das interfaces sem fio foram obtidas através de conexão via SSH com o

AP. Com a criação do arquivo de relacionamento foi possível associar todos os endereços MACs de cada equipamento com o seu local de instalação.

4.3 Tratamento de dados

A base de dados utilizada precisou ser tratada, devido à incompletude dos dados e à granularidade temporal das informações obtidas na universidade onde os dados foram obtidos.

Após coletar e armazenar os dados de conexão da rede Wi-Fi, provenientes do servidor Radius, é necessário realizar o tratamento dos dados (anonimização, mascaramento e remoção de dados com erro/não utilizados).

O primeiro desafio ocorre da necessidade de anonimização dos dados do usuário, essencial para garantir a privacidade do usuário (MENESES; MOREIRA, 2012).

A anonimização dos dados é realizada para preservar a privacidade do usuário, conforme prevê o regulamento sobre privacidade do Comitê de Ética em Pesquisa com Seres Humanos designado pela portaria 729/10. O pedido para utilização dos dados foi cadastrado sobre o número CAAE: 45215321.3.0000.5148 e aprovado pelo parecer número 4.996.908. Alguns campos utilizados pelo servidor Radius, não serão utilizados nessa pesquisa, sendo esses removidos do arquivo de *log* após a anonimização.

O campo *User-Name* que contém o *login* do usuário foi modificado de acordo com sua categoria. As categorias foram definidas da seguinte forma: docentes, discentes de graduação, discentes de pós-graduação, técnicos administrativos, funcionários terceirizados e setores administrativos. Assim, um *login* que se encaixa na categoria discente de graduação, foi alterado para *discente_graduacao_01* e assim sucessivamente.

O campo *Calling-Station-ID* é onde é gravado o MAC do dispositivo de cada conexão ao AP no arquivo de *log*. Apesar desse campo conter dados do MAC de um dispositivo específico é relativamente baixa a possibilidade do usuário ser identificado pelo MAC. Assim, foi mantido esse campo no *dataset* final. Com essa informação é possível saber o número de dispositivos diferentes que o usuário se conectou na rede ou que tipo de dispositivo (celular, tablet, notebook) está realizando a conexão.

Na Figura 4.5 é apresentado um exemplo do arquivo de *log* de conexão após anonimização e remoção de campos não utilizados.

Figura 4.5 – Log de conexão anonimizado

```
Tue May 24 10:45:06 2022
User-Name = "discente_pos_graduacao_114"
Acct-Status-Type = Start
NAS-Identifier = "2C-5D-93-06-D1-88"
Event-Timestamp = "May 24 2022 10:45:06 -03"
```

Fonte: do autor (2022)

Esses dados correspondem às conexões de diferentes usuários no AP associado em diversos intervalos de tempo. É então necessário um primeiro pré-processamento para identificar e corrigir os eventuais erros existentes na base de dados. Após a correção são gerados indicadores representativos das conexões dos usuários no local ao longo do dia.

Para realizar o tratamento de dados (anonimização, limpeza e filtragem) foi necessário converter os arquivos de *log* em arquivos CSV para serem manipulados no Python e em um banco de dados PostgreSQL. Devido ao grande número de registros e o hardware utilizado, um longo período de computação (tempo de processamento) foi necessário para visualização dos resultados. Em consequência desse tempo de processamento foi mais eficiente converter o arquivo diário de *log* para CSV, fazer o tratamento dos dados em arquivos diários individuais e depois agrupar todos os arquivos CSV, gerando um arquivo único.

4.4 Análise georreferenciada

Para que seja possível visualizar graficamente o local de acesso dos usuários, foi utilizado o OpenStreetMap (OSM), que é um projeto de mapeamento colaborativo para criar um mapa livre e editável do mundo (OPENSTREETMAP, 2021). Todos os mapas, dados descritivos, e metadados ofertados pelo OSM são dados abertos, disponíveis sob a licença *Open Database License* e sua base de dados é sustentada por colaboradores e pela comunidade. Com a utilização dos mapas do OSM é possível a visualização da conexão dos usuários em intervalos de tempo em determinado local e data, apresentados de forma geoespacial.

Após análise do mapa da instituição de ensino, constatou-se que o mesmo estava incompleto, faltando alguns prédios e ruas. Como essas informações são de extrema importância para estudo dos locais de conexão dos usuários e o mapeamento do grupo da rede Wi-Fi, foram realizadas as correções e inserções no OSM de prédios, inserção de novas avenidas e mapeamento dos locais onde a Universidade possui *Hotspot*, como mostrado na Figura 4.6.

Figura 4.6 – Mapa de exemplo dos locais com *Hotspot*

Fonte: adaptado de OpenStreetMap (2021)

Nos prédios em que a Universidade possui *Hotspot* existem diversos APs instalados para cobertura de todo o prédio. A conexão do usuário é baseada no MAC da interface sem fio do AP, mas no arquivo de *log* temos diversas conexões e desconexões em pontos de acesso que se sobrepõem, mesmo o usuário estando sem movimentação no local. Essa característica é chamada de efeito Ping-pong (DANALET; FAROOQ; BIERLAIRE, 2014). Esse efeito pode gerar diferenças no processamento dos dados, devido a um usuário com conexão/desconexão em diversos APs, em períodos de tempo muito curtos.

Os autores Meneses e Moreira (2012) utilizaram taxas de conexão para agregar APs de um mesmo prédio. Neste trabalho, foram agrupados todos os equipamentos de ponto de acesso localizados no mesmo prédio, considerando-se como somente um ponto de acesso. Assim, qualquer registro de conexão no conjunto de equipamentos que fazem parte do mesmo grupo, será visualizado como uma conexão ao *hotspot* do prédio em questão, independente do prédio possuir um ou mais APs.

Conforme citado anteriormente, o arquivo de *log* somente possui o endereço MAC da interface sem fio de conexão do usuário. Para relacionar esse endereço MAC ao local físico de instalação desse AP foi necessário extrair de cada ponto de acesso os endereços MACs das interfaces sem fio e gerar um CSV com os campos do endereço MAC e local. Também foi necessário relacionar com outro arquivo que contém as informações de local e coordenadas geográficas (latitude e longitude), conforme exemplo apresentado na Figura 4.7. Para o relacionamento de endereços MACs com o local, foram mapeados 1.068 endereços MACs. O Campus possui aproximadamente 350 APs instalados em 62 locais distintos.

Figura 4.7 – Exemplo *dataframe* do local com latitude e longitude

	local	Latitude	Longitude
0	ALMOXARIFADO	-21.22716	-44.98301
1	ALOJAMENTOINT	-21.23305	-44.99393
2	ALOJAMENTO	-21.22942	-44.99225
3	BIBLIOTECA	-21.22797	-44.97683
4	CANTINA	-21.22712	-44.97742
...
57	SALAOCONVENCOES	-21.22973	-44.97364
58	SINDICATO	-21.22668	-44.97698
59	TRANSPORTE	-21.22798	-44.97302
60	TVU	-21.22781	-44.98117
61	USINA	-21.22481	-44.97278

Fonte: do autor (2022)

Para visualização da distribuição das conexões de rede sem fio foi criado um mapa (*heatmap*) das regiões com mais acessos, denominadas regiões quentes e as regiões com menor número de acesso, definidas como regiões frias. Assim, será exibido no mapa essas regiões, baseadas em quantidade, tempo e local das conexões. A cor azul significa a região mais fria (menor número de conexões) e a cor vermelha a região mais quente (maior número de conexões). O mapa de calor será mostrado na Figura 5.1 da Seção 5.1.

Esse tipo de representação contrasta com um mapa de dispersão, que é essencialmente apenas um gráfico de dispersão de pontos sobrepostos no topo de um mapa. Os mapas de dispersão são mais difíceis de visualizar a quantidade de conexões devido ao *overplotting* quando se tem muitos pontos de dados no mesmo local/área.

4.5 Algoritmos para tratamento dos dados

Conforme a Seção 4.2, foram analisados os dados presentes nos arquivos de conexão do servidor Radius e a partir desses dados foi possível sequenciar a localização das conexões dos usuários ao longo do tempo. Com o mapeamento dos *hotspots* será possível visualizar em que local do mapa o usuário está conectado.

Para conseguir realizar o dimensionamento de recursos foi utilizado o banco de dados PostgreSQL, a linguagem de programação Python e as bibliotecas Pandas, Plotly, Numpy, Folium e Seaborn.

Segundo McKinney (2019), Python é uma das linguagens mais importantes em ciência de dados e aprendizado de máquina (*machine learning*). Essa linguagem possui diversas bibliotecas para auxílio manipulação de dados e criação de mapas e gráficos. A biblioteca Numpy oferece código aglutinador para estrutura de dados, os algoritmos e a biblioteca necessários à maioria das aplicações científicas que envolvam numéricos em Python (NUMPY, 2022). Pandas é uma ferramenta de análise e manipulação de dados de código aberto, rápida, poderosa, flexível e fácil de usar (PANDAS, 2022).

A biblioteca de gráficos Plotly cria gráficos interativos com qualidade de publicação. Pode-se fazer gráficos de linha, gráficos de dispersão, gráficos de área, gráficos de barras, barras de erro, gráficos de caixa, histogramas, mapas de calor, subplots, eixos múltiplos, gráficos polares e gráficos de bolhas (PLOTLY, 2022). Seaborn é uma biblioteca de visualização de dados Python baseada em matplotlib. Ele fornece uma interface de alto nível para desenhar gráficos estatísticos atraentes e informativos (SEABORN, 2022). Com a biblioteca Folium pode-se criar um mapa de folheto interativo. Ele permite a vinculação de dados a um mapa e integração com OpenStreetMap, Mapbox e Stamen (FOLIUM, 2022).

Para realizar o tratamento dos dados foram desenvolvidos diversos códigos, como descrito a seguir. Os algoritmos estão demonstrados no Apêndice 6.

1) **log2csv**: Responsável por converter os arquivos de *log* para *Dataframe*. O código procura em uma pasta todos os arquivos de *log* iniciados com o nome *detail*, converte os dados para um *dataframe*, cria uma nova coluna chamada DATE interpretando os dados da coluna "Event Timestamp", conforme exemplo da Figura 4.8. Depois realiza a limpeza dos campos que não serão utilizados e gera um arquivo CSV para cada arquivo de *log*, conforme exemplo da Figura 4.9. O campo *user-name* está com os dados borrados porque exibe dados de login do usuário e nessa fase esses campos ainda não foram anonimizados.

Figura 4.8 – *DataFrame* com os dados da conversão do arquivo de *log*

	date	User-Name	Acct-Status-Type	Acct-Authentic	Framed-IP-Address	Calling-Station-Id	NAS-IP-Address	NAS-Port	Called-Station-Id	NAS-Port-Type	...	Acct-Output-Packets	Ruckus-Station-RSSI	Event-Timestamp	Tmp-String-9
1	2021-05-06 00:00:04	[REDACTED]	Interim-Update	Remote	172.16.59.17	98-39-8E-D8-A6-21	192.168.14.3	42	84-18-3A-0B-15-58:UFLA+	Wireless-802.11	...	187	11	May 6 2021 00:00:04-03	ai:
2	2021-05-06 00:00:11	[REDACTED]	Interim-Update	Remote	172.16.23.185	04-D9-F5-68-BC-03	192.168.14.3	42	F0-3E-90-3A-3E-A8:UFLA+	Wireless-802.11	...	2225973	21	May 6 2021 00:00:11-03	ai:
3	2021-05-06 00:00:14	[REDACTED]	Interim-Update	Remote	172.16.38.134	F8-1F-32-AD-E8-F3	192.168.14.3	13	2C-5D-93-06-C6-98:UFLA+	Wireless-802.11	...	609510	45	May 6 2021 00:00:14-03	ai:
4	2021-05-06 00:00:17	[REDACTED]	Start	Remote	172.16.62.127	FE-02-75-EE-12-12	192.168.14.3	13	2C-5D-93-07-6B-78:UFLA+	Wireless-802.11	...	NaN	NaN	May 6 2021 00:00:17-03	ai:
5	2021-05-06 00:00:17	[REDACTED]	Stop	Remote	172.16.62.127	FE-02-75-EE-12-12	192.168.14.3	89	2C-5D-93-06-AF-BC:UFLA+	Wireless-802.11	...	366	10	May 6 2021 00:00:17-03	ai:

5 rows x 29 columns

Fonte: do autor (2022)

Figura 4.9 – *DataFrame* após a limpeza de campos não utilizados

	date	User-Name	Acct-Status-Type	NAS-Identifier
1	2021-05-06 00:00:04	[REDACTED]	Interim-Update	84-18-3A-0B-15-58
2	2021-05-06 00:00:11	[REDACTED]	Interim-Update	F0-3E-90-3A-3E-A8
3	2021-05-06 00:00:14	[REDACTED]	Interim-Update	2C-5D-93-06-C6-98
4	2021-05-06 00:00:17	[REDACTED]	Start	2C-5D-93-07-6B-78
5	2021-05-06 00:00:17	[REDACTED]	Stop	2C-5D-93-06-AF-BC

Fonte: do autor (2022)

2) **Convert_mac_local**: Realiza a leitura dos arquivos .CSV originados do algoritmo **Logs2csv** e do arquivo de APs (esse arquivo contém os MACs e locais dos APs). Analisa a coluna `NAS_Identifier` que contém o MAC do AP que o usuário conectou e faz a troca do MAC pelo local, fazendo o tratamento de caracteres em maiúsculo/minúsculo. Depois gera um arquivo CSV para cada arquivo lido. Na Figura 4.10 é demonstrado um exemplo do *dataframe* com a conversão do MAC em local.

Figura 4.10 – *DataFrame* com MAC alterado para o local

	date	User-Name	Acct-Status-Type	NAS-Identifier
0	2021-05-06 00:00:04	[REDACTED]	Interim-Update	PORTARIA
1	2021-05-06 00:00:11	[REDACTED]	Interim-Update	DMV
2	2021-05-06 00:00:14	[REDACTED]	Interim-Update	ALOJAMENTO
3	2021-05-06 00:00:17	[REDACTED]	Start	DCA
4	2021-05-06 00:00:17	[REDACTED]	Stop	DEG

Fonte: do autor (2022)

3) **Create_BD**: Realiza a leitura dos arquivos .CSV originados do script `convert_mac_local` e popula uma tabela no banco de dados Postgres, para posteriormente realizar o agrupamento de todas as conexões.

4) **View_BD**: Esse código é executado em SQL em um banco de dados Postgres. Nele é inserido os dados de todos os arquivos .CSV gerados pelo script `convert_mac_local` e realizado o agrupamento das conexões dos usuários em determinado local em intervalo de tempo de 15 minutos.

5) **Create_csv_from_bd**: Cria um arquivo CSV a partir do banco de dados Postgres, onde os arquivos foram agrupados e as conexões dos usuários foram separados em intervalos de tempo de 15 minutos.

6) **Insert_lat_long**: Realiza a leitura do arquivo CSV originado pelo script `create_csv_from_bd` e do arquivo `local_lat_long.csv` (que contém os locais dos pontos de acesso e suas coordenadas geográficas) para inserir as coordenadas dos locais no arquivo principal. Na Figura 4.11 é apresentado um exemplo do *dataframe* com adição das colunas de latitude e longitude.

Figura 4.11 – Dataframe com Latitude e Longitude

	nome	acesspoint	dia	momento	Latitude	Longitude
0	...	PAVILHAO2	2022-06-22	16:15	-21.22554	-44.97743
1	...	PAVILHAO2	2022-06-22	16:30	-21.22554	-44.97743
2	...	PAVILHAO2	2022-06-22	16:45	-21.22554	-44.97743
3	...	DCF	2022-06-03	10:00	-21.22604	-44.97349
4	...	DCF	2022-06-03	10:15	-21.22604	-44.97349

Fonte: do autor (2022)

7) **Anonymous**: Realiza a leitura do arquivo gerado pelo script `Insert_lat_long` e do `usuarios.csv` (esse arquivo contém o login do usuário e a sua categoria), anonimiza o dataframe e insere a categoria do usuário. Na Figura 4.12 é demonstrado um exemplo do *dataframe* anonimizado.

Figura 4.12 – Dataframe anonimizado

	nome	acesspoint	dia	momento	Latitude	Longitude	categoria
0	discente_graduação_308	PAVILHAO2	2022-06-22	16:15	-21.22554	-44.97743	discente_graduação
1	discente_graduação_308	PAVILHAO2	2022-06-22	16:30	-21.22554	-44.97743	discente_graduação
2	discente_graduação_308	PAVILHAO2	2022-06-22	16:45	-21.22554	-44.97743	discente_graduação
3	discente_pos_graduacao_587	DCF	2022-06-03	10:00	-21.22604	-44.97349	discente_pos_graduacao
4	discente_pos_graduacao_587	DCF	2022-06-03	10:15	-21.22604	-44.97349	discente_pos_graduacao

Fonte: do autor (2022)

8) **Heatmap**: Exibe o mapa interativo, utilizando a biblioteca Folium, que exibe as conexões em forma de mapa de calor por local em intervalos de 15 em 15 minutos. Para esse mapa foram removidas as conexões compreendidas entre as 23:01 e as 06:59 horas.

9) **Mapa_categoria**: Exibe o mapa interativo, utilizando a biblioteca Plotly, que mostra as conexões no mapa por categoria dos usuários. Para esse mapa foram removidas as conexões compreendidas entre as 23:01 e as 06:59 horas.

Após a execução dos algoritmos descritos na seção 4.5, foi possível gerar um *dataset* anonimizado e com agrupamento dos usuários em categorias e intervalos de tempo. O *dataset* possui 10.550.396 registros de conexões no intervalo de junho a setembro de 2022. A Figura 4.13 apresenta um recorte do

dataset final. O login do usuário foi trocado pela sua categoria adicionado de uma numeração sequencial. Pode-se visualizar em intervalos de 15 minutos em que local determinado usuário estava conectado, em qualquer dia, dentro dos meses de junho a setembro de 2022. O intervalo de 15 minutos foi definido para termos quatro vezes a posição do usuário a cada hora. Com isso é possível analisar as mudanças de local entre intervalos de aula, por exemplo. As conexões compreendidas entre as 23:01 e as 06:59 horas foram removidas por se tratar de um intervalo de tempo em que a Universidade não possui aulas.

Figura 4.13 – Recorte dataframe

	nome	acesspoint	dia	momento	Latitude	Longitude	categoria
0	discente_graduação_308	PAVILHAO2	2022-06-22	13:00	-21.22554	-44.97743	discente_graduação
1	discente_graduação_308	PAVILHAO2	2022-06-22	13:30	-21.22554	-44.97743	discente_graduação
2	discente_graduação_308	PAVILHAO2	2022-06-22	13:45	-21.22554	-44.97743	discente_graduação
3	discente_graduação_308	PAVILHAO2	2022-06-22	14:15	-21.22554	-44.97743	discente_graduação
4	discente_graduação_308	PAVILHAO2	2022-06-22	15:00	-21.22554	-44.97743	discente_graduação
5	discente_graduação_308	PAVILHAO2	2022-06-22	16:00	-21.22554	-44.97743	discente_graduação
6	discente_graduação_308	PAVILHAO2	2022-06-22	16:15	-21.22554	-44.97743	discente_graduação
7	discente_graduação_308	PAVILHAO2	2022-06-22	16:30	-21.22554	-44.97743	discente_graduação
8	discente_graduação_308	PAVILHAO2	2022-06-22	16:45	-21.22554	-44.97743	discente_graduação
9	discente_pos_graduacao_587	DCF	2022-06-03	10:00	-21.22604	-44.97349	discente_pos_graduacao
10	discente_pos_graduacao_587	DCF	2022-06-03	10:15	-21.22604	-44.97349	discente_pos_graduacao
11	discente_pos_graduacao_587	DCF	2022-06-03	10:30	-21.22604	-44.97349	discente_pos_graduacao
12	discente_pos_graduacao_587	DCF	2022-06-03	10:45	-21.22604	-44.97349	discente_pos_graduacao
13	discente_pos_graduacao_587	DCF	2022-06-03	11:00	-21.22604	-44.97349	discente_pos_graduacao
14	discente_pos_graduacao_587	DCF	2022-06-03	11:15	-21.22604	-44.97349	discente_pos_graduacao
15	discente_pos_graduacao_587	DCF	2022-06-03	11:30	-21.22604	-44.97349	discente_pos_graduacao

Fonte: do autor (2022)

4.6 Dimensionamento do número de APs

Em redes sem fio, o recurso de rádio é limitado e escasso, e a qualidade do canal de cada usuário pode variar ao longo do tempo. Dada a condição do canal de cada usuário e o recurso total do sistema disponível, a quantidade de largura de banda atribuída a cada usuário pode ser guiada por métricas de desempenho como taxa de transferência ou de acordo com o tipo de tráfego (KUO; LIAO, 2008).

Segundo Hills (2001) o projeto de uma rede sem fio de grande escala deve considerar a seleção dos locais de instalação e a atribuição de frequências dos APs, de forma a garantir uma completa cobertura de sinal e capacidade adequada para transportar a carga de tráfego prevista.

Para dimensionar a quantidade ideal de APs em cada local, foi utilizado como base os dados do arquivo de *log*, com o número de conexões em intervalos de tempo. A base para esse dimensionamento foi definida pelo número de usuários conectados simultaneamente. Em seu trabalho Ghosh et al. (2011) relata que medições detalhadas nos níveis de pacote e fluxo são difíceis de coletar e raramente são coletadas continuamente, na prática, muitas decisões sobre dimensionamento, são baseadas na capacidade dos equipamentos de rede em termos do número de usuários conectados simultaneamente.

Neste trabalho foi definido como premissa de QoS a largura de banda mínima reservada para cada usuário. Não será avaliado atribuição de frequências, sobreposição de cobertura entre APs operando no mesmo canal de rádio e controle de potência de transmissão. Todas essas questões são resolvidas de forma automática pela controladora dos equipamentos da rede sem fio.

Para a definição da banda mínima reservada para cada usuário foi utilizado um guia dos serviços onlines típicos com a velocidade mínima de download necessária para o desempenho adequado de cada aplicativo, conforme Tabela 4.1. As velocidades são baseadas na execução de uma atividade por vez. Esses números são diretrizes aproximadas divulgadas no site da FCC - *Federal Communications Commission* (FCC, 2023). A FCC regulamenta as comunicações interestaduais e internacionais por rádio, televisão, fio, satélite e cabo em todos os 50 estados, no Distrito de Columbia e nos territórios dos EUA. É uma agência independente do governo dos EUA supervisionada pelo Congresso e responsável pela implementação e aplicação das leis e regulamentações de comunicações no país.

Tabela 4.1 – Serviços e velocidade de download

Categoria	Velocidade mínima de download (Mbps)
Email	1
Streaming online de rádio	0,5
VoIP	0,5
Aluno (navegação)	6
Redes sociais	1
Streaming alta definição de vídeo	5
Videoconferência	6

Fonte: do autor (2023)

De acordo com a definição dos serviços selecionados, a reserva de largura de banda mínima para cada usuário é de 20 Mbps. Assim foi preciso definir um critério para saber qual a quantidade de usuários pode ser conectada por AP, baseado na reserva de largura de banda.

O Wi-Fi utiliza o método CSMA/CA para compartilhamento de recursos de rádio. Devido à sobrecarga, podem ocorrer colisões de tempos em tempos, o que reduz o uso de recursos de rádio (HU et al., 2012). A largura de banda do protocolo 802.11ac também é influenciada por outros fatores como, largura do canal utilizado, modulação, quantidade de antenas do equipamento, fluxos espaciais, tipo de

onda e relação sinal/ruído(SNR) (CISCO, 2018). Como essas informações não estão disponíveis no *dataset*, foi utilizado como base, a banda da porta LAN do AP (1 Gbps) para garantir o mínimo de banda definido. Para o percentual da banda reservada foi utilizado o valor de 100%, ou seja, os 20Mbps precisam estar garantidos para cada usuário.

A Fórmula 4.1 apresenta o cálculo da quantidade máxima de usuário em cada AP.

$$\left(\frac{Banda_lan}{Banda_user * Porcent_util} \right) = N_user_ap \quad (4.1)$$

Banda_lan: Banda da interface LAN do equipamento sem fio.

Banda_user: Banda mínima reservada para cada usuário.

Porcent_util: Percentual de utilização da banda reservada.

Entrando com os dados banda_lan (1000 Mbps), banda_user (20 Mbps) e percent_util (100%) na fórmula, chegou-se ao resultado de 50 usuários por AP, para a garantia mínima de reserva de largura de banda.

Para cálculo do número de APs necessários, foi utilizada a Fórmula 4.2:

$$\left(\frac{N_user_connect}{\sum(ap_local) * N_user_ap} \right) - 1 = N_crescimento \quad (4.2)$$

N_user_connect: Esse campo é o número máximo de usuários conectados simultaneamente nos locais dos *hotspots* durante o período de 4 meses.

AP_local: APs existente em cada local que representa um *hotspots*.

N_user_ap: Número máximo de usuários por AP de acordo com a Fórmula 4.1.

Caso o valor de N_crescimento seja menor ou igual a zero, significa que nesse local não será preciso o incremento no número de equipamentos.

O último passo é multiplicar o número de APs do local pelo N_crescimento para chegar a quantidade de equipamentos que precisam ser adicionados, conforme Fórmula 4.3.

$$N_aps_local * N_crescimento = Qtd_aps \quad (4.3)$$

As informações extraídas do arquivo de *log* que serviram como parâmetros de entrada para o dimensionamento de recursos e a obtenção dos resultados são descritos no próximo capítulo.

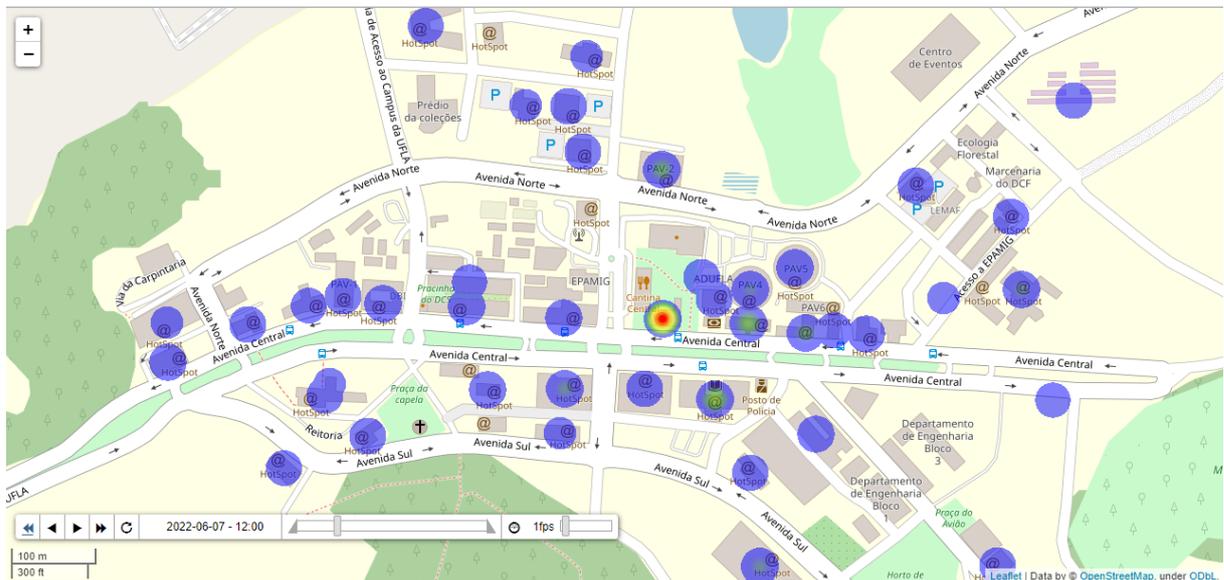
5 RESULTADOS

Esta capítulo mostra os resultados obtidos a partir das execuções dos algoritmos criados para tratamentos dos dados dos arquivos de *logs* de conexão, apresenta gráficos e mapas (mapa de calor e pontos de dispersão) para comparações de comportamento do usuário de acordo com local e intervalo de tempo e a tabela de dimensionamento de equipamentos.

5.1 Mapa de calor - (*Heatmap*)

Para avaliar os locais com o maior número de usuários e consequentemente auxiliar no dimensionamento de equipamentos, foi criado um mapa de calor (*heatmap*). Esse é um mapa interativo onde o usuário pode avançar no tempo através de uma linha do tempo e rapidamente ter uma noção da densidade e da distribuição espacial de seus usuários. Assim, é possível observar a mudança das zonas quentes e zonas frias das conexões dos usuários, a fim de analisar o padrão de distribuição espacial por local no decorrer do dia.

Figura 5.1 – Mapa de calor - 07/06/22 - 12:00



Fonte: do autor (2022)

Na Figura 5.1 estão listadas as conexões dos usuários do dia 07 de junho de 2022 (07/06/22) que ocorreu entre 12:00 e 12:15. Essa data foi escolhida aleatoriamente, para demonstrar a diferença de conexões por local dependendo do horário. Pelo mapa é possível visualizar que o local com maior concentração de pessoas é a área da Cantina, isso se deve ao fato do horário de almoço, quando os usuários se concentram nesta região. Nesses acessos, estão contabilizados todas as categorias, discentes

Figura 5.2 – Número de usuários conectados em 07/06/22 - 12:00

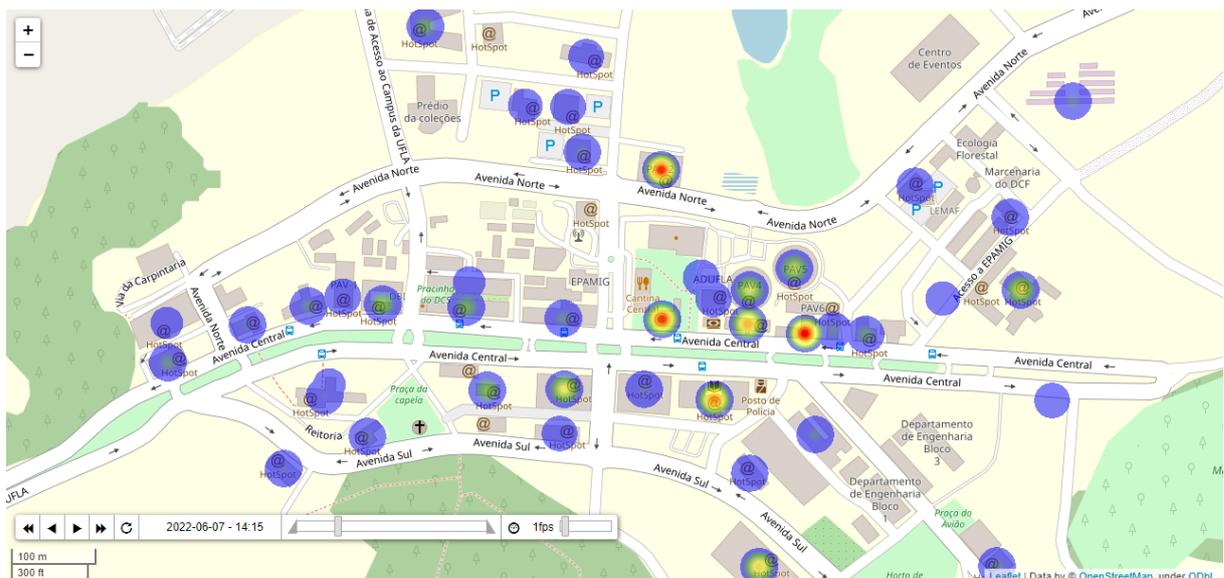
acesspoint	dia	momento	counts
CANTINA	2022-06-07	12:00	360
PORTARIA	2022-06-07	12:00	145
BIBLIOTECA	2022-06-07	12:00	118
PAVILHAO3	2022-06-07	12:00	105
PAVILHAO2	2022-06-07	12:00	83

Fonte: do autor (2022)

de graduação e pós, docentes, técnicos administrativos, terceirizados e setores. A Figura 5.2 apresenta, em números, a quantidade de conexões.

Na Figura 5.3 estão listadas as conexões de todos os usuários do dia 07 de junho de 2022 (07/06/22) que ocorreu entre 14:15 e 14:30. Pelo mapa é possível visualizar que a local com maior concentração de pessoas não é mais a área da Cantina. A concentração maior é no prédio da biologia, seguida pelos prédios das engenharias, biblioteca e pró-reitorias. A Figura 5.4 apresenta, em números, a quantidade de conexões.

Figura 5.3 – Mapa de calor - 07/06/22 - 14:15



Fonte: do autor (2022)

Para que seja possível analisar os locais de conexão dos usuários foi criado um endereço Web para consulta ao mapa interativo. O mapa pode ser consultado no endereço: <http://mapswifi.ufla.br>

Figura 5.4 – Número de usuários conectados em 07/06/22 - 14:15

acesspoint	dia	momento	counts
PAVILHAO6	2022-06-07	14:15	156
PAVILHAO2	2022-06-07	14:15	148
CANTINA	2022-06-07	14:15	145
PORTARIA	2022-06-07	14:15	139
BIBLIOTECA	2022-06-07	14:15	124

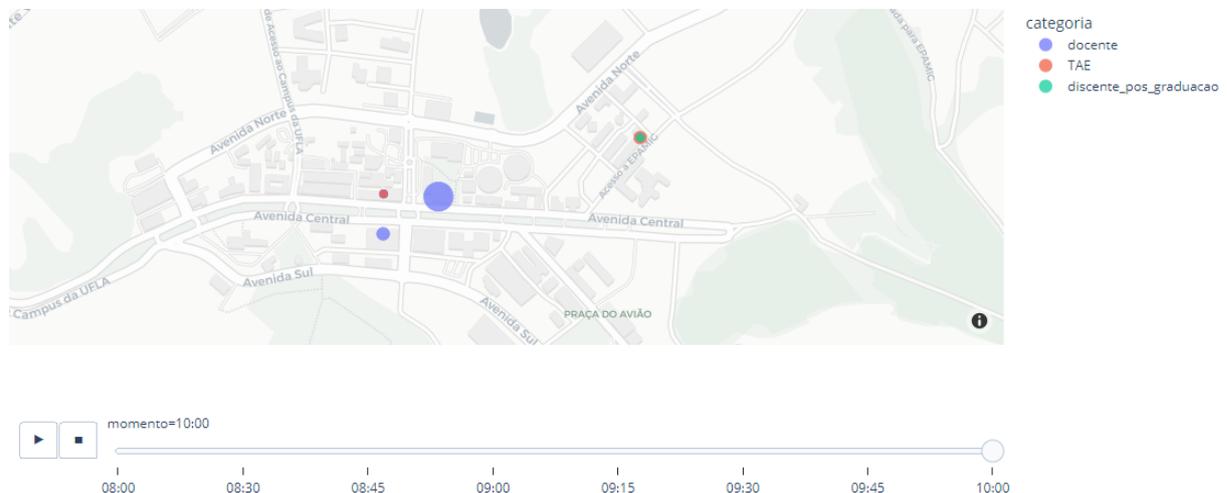
Fonte: do autor (2022)

5.2 Mapa com pontos de dispersão

O mapa de calor apresentado na seção anterior mostra de forma intuitiva e rápida os locais com maior concentração de conexões. Com um mapa com pontos de dispersão é possível visualizar a quantidade de conexões por intervalo de tempo separado em categorias. Assim é possível relacionar o número de conexões no mesmo período de tempo do mapa de calor com a quantidade de usuários de cada categoria.

Para avaliar o número de conexões por categoria foi criado um mapa separado por categorias. Esse mapa é interativo, onde o usuário pode avançar no tempo através de uma linha do tempo e rapidamente ter uma noção da densidade e da distribuição espacial de seus usuários por categoria. Assim é possível observar os locais em que cada categoria tem maior concentração por local no decorrer do dia, conforme Figura 5.5.

Figura 5.5 – Mapa de categoria



Fonte: do autor (2022)

Como se trata de um mapa interativo, é possível ocultar/exibir categorias do mapa e quando o mouse é posicionado em cima de cada bolha é exibido a categoria, momento da conexão, número de conexões, latitude e longitude. Na Figura 5.6 é possível visualizar exemplo dessas informações.

Figura 5.6 – Informações da categoria docente no local DEX



Fonte: do autor (2022)

5.3 Análise gráfica do número conexões

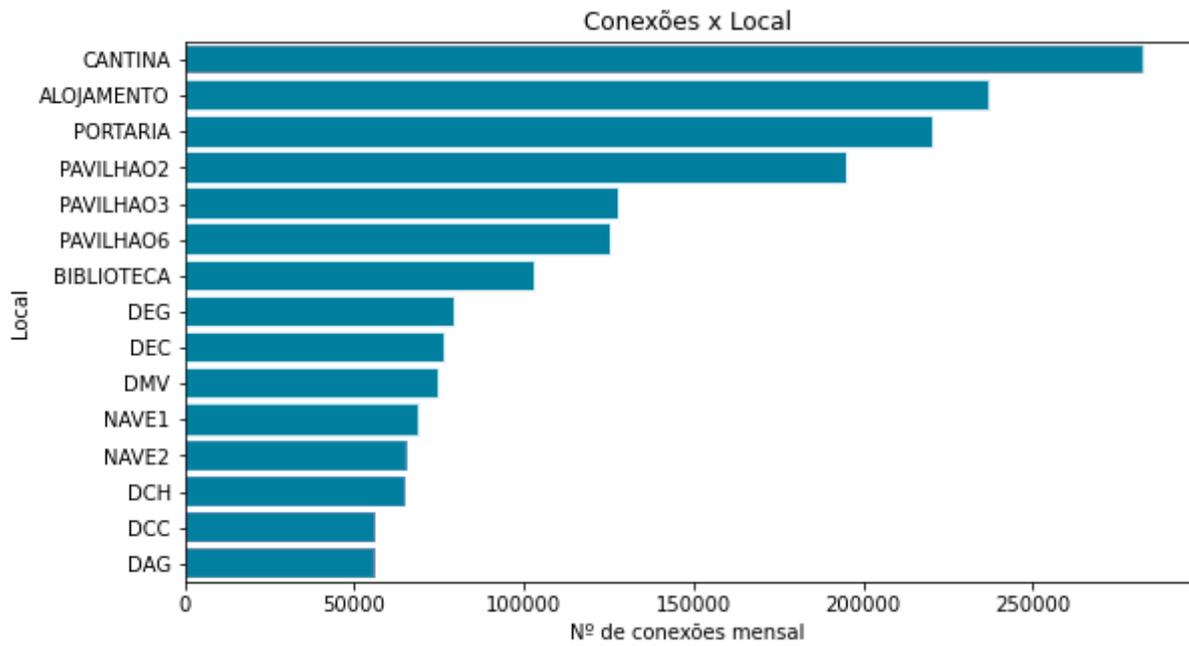
Para facilitar a compreensão das informações geradas pelo *dataset* serão gerados diversos gráficos a fim de ajudar a identificar padrões, verificar resultados e comparar medidas de forma ágil.

5.3.1 Maior número de conexões

Na Figura 5.7 é possível visualizar o gráfico que exhibe os quinze locais com maior número de conexões durante o período do mês de junho (01/06/22 a 30/06/22).

Analisando os 5 locais com maior número de acessos, a Cantina foi o local com maior número de conexões. Isso se deve ao fato de ser um local de encontro dos usuários nos intervalos entre aulas e para lanche. Os usuários frequentam esse local diversas vezes por dia, gerando novas conexões. O alojamento estudantil foi o segundo local em número de conexões, vale ressaltar que esse local é composto por 3 prédios, 30 APs e aproximadamente 480 moradores. A portaria tem um grande número de acessos, pois quando os usuários entram ou saem da Universidade há grandes chances do dispositivo se conectar na rede automaticamente. Outro fator, é o ponto de carona que fica localizado em frente a portaria e o embarque/desembarque de alunos no transporte interno, que é feito através de um ônibus articulado. O quarto e quinto local com maior número de conexão são os pavilhões de aula.

Figura 5.7 – Número de conexões x local

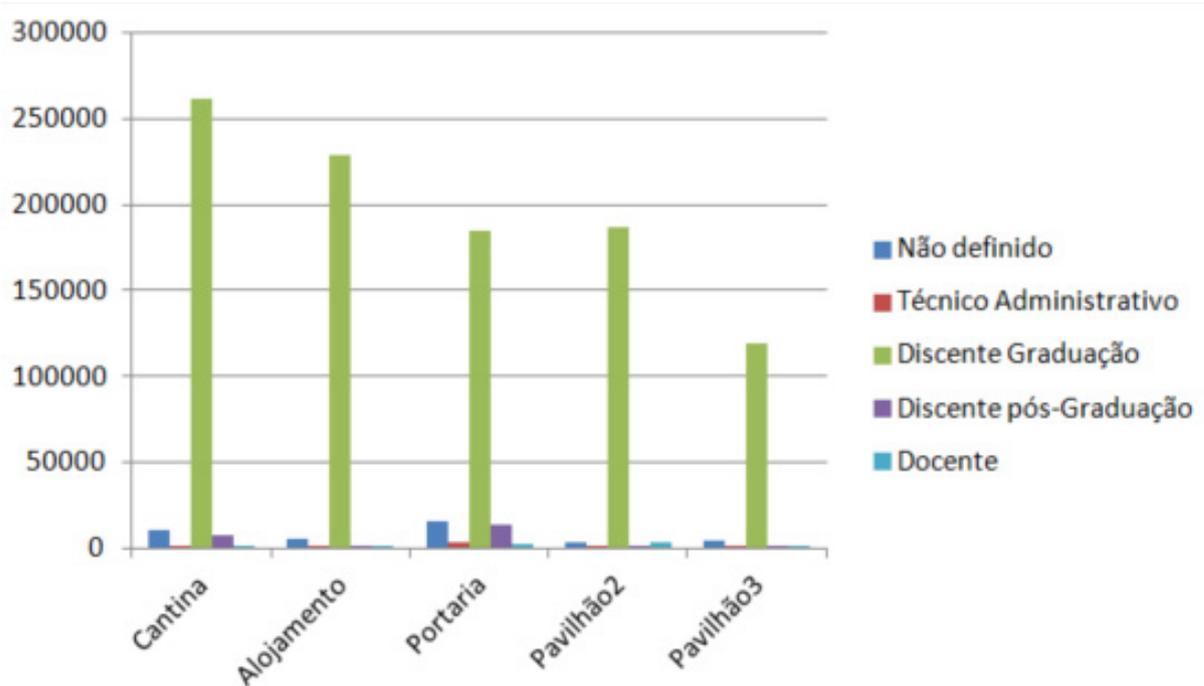


Fonte: do autor (2022)

5.3.2 Número de acessos por categorias

Nesse gráfico serão analisados os 5 locais com maior número de acessos por categorias no mês de junho/22.

Figura 5.8 – Número de conexões x local x categoria - junho/22



Fonte: do autor (2022)

Tabela 5.1 – Número de conexões x categoria - junho/22.

Categoria	Cantina	Alojamento	Portaria	Pavilhão2	Pavilhão3
Não definido	10502	5976	15295	3467	4433
Técnico Administrativo	1275	305	3093	463	371
Discente de graduação	261506	229004	185252	186715	119543
Discente de pós-Graduação	7808	1193	13636	1124	1602
Docente	1219	197	2834	3301	1457

Fonte: do autor (2023)

A Figura 5.8 exibe o gráfico dos 5 locais com maior número de acessos no mês de junho/22 e Tabela 5.1 quantifica o número de conexões. O acesso dos discentes de graduação em todos os locais é bem superior ao das outras categorias. Isso se deve ao fato do número de discentes de graduação ser muito maior do que as outras categorias. A quantidade de conexões de técnicos administrativos e docentes na cantina, alojamento e portaria são bem próximas, já nos pavilhões de aula as conexões dos docentes superam as conexões de técnicos em mais de quatro vezes. O motivo é que os docentes ministram aulas nos pavilhões e têm um acesso mais frequente a esses locais.

5.4 Dimensionamento de recursos de rede

Concluídas as etapas de coleta e tratamentos dos dados, criação do *dataset* e apresentação da análise georreferenciada, foi feito o dimensionamento do número de equipamentos por local, baseado em uma reserva de largura de banda mínima. Para isso foi realizado uma consulta ao *dataset* e listado por local o máximo de conexões em um intervalo de tempo. A Tabela 5.2 apresenta o local e qual foi o número máximo de conexões recebidas dentro do intervalo de tempo de 15 minutos, durante o período de 4 meses.

Tabela 5.2 – Local e número máximo de conexões (continua)

Local	Número máximo de conexões
Almoxarifado	6
Alojamento internacional	29
Alojamento	197
Biblioteca	293
Cantina	768
Centro de convivência	78
Ciuni	86
DAE	62
DAG	143
DBI	171
DCA	85
DCC	129
DCF	50
DCH	157
Administrativo	68
DCS	9
DED	80
DEF	109
DEG	197
DEN	18
DEX	112
DFI	71
DFP	49

Tabela 5.2 – Local e número máximo de conexões (continua)

DGA	25
DGTI	38
DIRED	40
DIR	14
DLA	59
DMV	160
DNU	51
DQI	77
DRI	5
DME	96
DZO	41
DEC	144
Fisiologia	63
Inbatec	5
Inovacafé	98
Lemaf	110
LIFE	15
Nave1	168
Nave2	175
NEDI	54
Pavilhão 1	37
Pavilhão 2	353
Pavilhão 3	240
Pavilhão 6	230
Pavilhão 9	155
Portaria	206
PRAEC	142
PROINFRA	86
Pró-Reitorias	56
Reitoria	65
Salão de convenções	110

Tabela 5.2 – Local e número máximo de conexões (conclusão)

Sindicato	92
Transporte	12
TVU	8
Usina	43

Fonte: do autor (2023)

Além de dimensionar o número de APs por local, é importante dimensionar a largura de banda da ligação entre o *switch* onde estão ligados os APs com o núcleo da rede. A ligação de locais com mais de 50 conexões simultâneas deve ser de no mínimo 10Gbps. Com essa velocidade e considerando a reserva de largura de banda de 20Mbps por usuário da rede sem fio, seria possível suportar o fluxo de 500 conexões com utilização máxima. Nesse cálculo está sendo desconsiderado a utilização de banda de dispositivos conectados via cabo nesses *switchs*.

Para definir a quantidade de APs por local foram utilizadas as fórmulas 4.2 e 4.3. O resultado é apresentado na Tabela 5.3.

Tabela 5.3 – Locais e número de equipamentos a serem adicionados.

Local	N Conexões	Qtd APs local	Qtd APs à adicionar
Biblioteca	293	5	1
Cantina	768	6	9
Ciuni	86	1	1
Nave1	168	2	2
Nave2	175	3	1
Nedi	54	1	1
Portaria	206	1	3
Sindicato	92	1	1

Fonte: do autor (2023)

Diante disso podemos analisar que os locais citados na Tabela 5.3 irão necessitar de acréscimo de equipamentos para manter a reserva de banda mínima. O local com maior criticidade é a cantina, onde temos o maior número de conexões de todo o campus. O prédio do sindicato dos técnicos é próximo da cantina e sua utilização se deve a conexões de usuários que frequentam essa região. O *hotspot* do Nedi obteve 4 conexões acima do definido na Fórmula 4.1, foi realizado a análise desse local nos outros intervalos de tempo e somente em 5 vezes foi ultrapassada a quantidade de 50 conexões. Assim, nesse local não será necessário adição de novo equipamento para manter a reserva de largura de banda mínima. O restante dos locais, que não estão relacionados na Tabela 5.3 não necessitam adição de novos APs pelo

requisito de largura de banda, mas pode ser necessário adição de equipamentos devido a cobertura de sinal, que não foi alvo desse estudo.

6 CONCLUSÃO

Neste trabalho foi proposta a geração de uma base de dados (*dataset*) fundamentada nos arquivos de *logs* de conexão da rede Wi-Fi de um campus universitário. Esse *dataset* contém informações de um período de 4 meses, com intervalos de tempo divididos em 15 minutos, locais de conexão, tipos de usuários, quantidade de conexões, entre outros dados. Com essas informações foi definida uma metodologia baseada em largura de banda mínima por usuário para cálculo do dimensionamento de recursos. O resultado indicam quais os locais que necessitam adição de equipamentos. Também foram criados mapas georreferenciados interativos, onde foi possível localizar os locais com maior concentração de usuários no decorrer do tempo.

Uma limitação encontrada se deve ao desafio do chamado efeito Ping Pong, que corresponde ao registro de uma sucessão de conexões/desconexões entre APs enquanto o usuário está parado. Para contornar esse problema foi realizado o agrupamento dos APs de um mesmo prédio/setor. Esse agrupamento pode gerar uma limitação na interpretação dos dados, já que um AP do prédio/setor pode estar sendo subutilizado, porém devido ao agrupamento essa informação não será percebida.

Apesar dessa questão, o resultado obtido mostra uma visão geral dos acessos dos usuários em cada local da Universidade. Outra limitação relacionada aos dados de entrada é o intervalo de tempo de 15 minutos das conexões, que foi definido para esse conjunto de dados. A escolha desse intervalo de conexões pode afetar a precisão do local onde o usuário está conectado, já que neste intervalo o usuário pode estar em deslocamento e conectar-se em diversos pontos de acesso. Para contornar esse problema, foi definido que o local que o usuário permaneceu mais tempo conectado, dentro do intervalo de 15 minutos, fosse contabilizado.

Considerando a diminuição do intervalo e adaptação da análise de tempo, o conjunto de dados seria mais preciso, ou seja, aumentaria a precisão da descrição da localização da conexão, porém a base de dados aumentaria consideravelmente e o tempo para processamento dos algoritmos também. A visualização interativa dos mapas também ficaria prejudicada devido a quantidade de informações no período de tempo que precisariam ser exibidas.

Então o trabalho foi baseado no conceito de conexões realizadas pelos usuários nos diferentes locais dentro do intervalo de 15 minutos. Para o modelo proposto a base de dados gerada foi suficiente para as informações de conexão.

Com a análise dos gráficos foi possível ter informações dos locais que precisam de atenção especial do setor de infraestrutura de TI, para evitar congestionamento e melhorar o desempenho da rede. Existem diversos locais em que os equipamentos de interconexão (*switchs*), onde estão ligados os

APs, possuem largura de banda de ligação (*backbone*) com o núcleo da rede de 1Gbps. Isso pode ser um gargalo para a conexão dos usuários da rede Wi-Fi.

Este trabalho se diferencia dos trabalhos anteriores da área pela criação e disponibilização da base de dados em <https://mapswifi.ufla.br>, categorizada e com locais de conexão e por propor o dimensionamento de equipamentos baseado em largura de banda para o usuário baseado no número de conexões por local.

Como trabalhos futuros, pretende-se realizar outras entradas de dados considerando um intervalo de tempo menor e comparando com os resultados obtidos nesse trabalho, analisar os arquivos de *logs* online, para que seja gerado um painel com informações em tempo real das conexões dos usuários e tentar determinar quando um local poderá entrar em colapso e também propor uma metodologia para economia de energia em dispositivos ociosos.

REFERÊNCIAS

- AGGARWAL, R.; KOKSAL, C. E.; SCHNITER, P. On the design of large scale wireless systems. **IEEE Journal on Selected Areas in Communications**, Institute of Electrical and Electronics Engineers (IEEE), v. 31, n. 2, p. 215–225, feb 2013.
- ARUBA. **O que é 802.11ax (WI-FI 6)?** Aruba, 2020. Acessado em 05/10/2022. Disponível em: <<https://www.arubanetworks.com/br/resource/o-que-e-802-11ax-wi-fi-6-e-por-que-voce-precisa-dele>>.
- ARYAFAR, E. et al. Rat selection games in hetnets. **Proceedings IEEE INFOCOM**, Institute of Electrical and Electronics Engineers (IEEE), p. 998–1006, 2013.
- BAHRI, A.; CHAMBERLAND, S. On the wireless local area network design problem with performance guarantees. **Computer Networks**, Elsevier BV, v. 48, n. 6, p. 856–866, aug 2005.
- BASAGNI, S. et al. Heuristics for lifetime maximization in wireless sensor networks with multiple mobile sinks. **International Conference on Communications**, Institute of Electrical and Electronics Engineers (IEEE), jun 2009.
- BHANAGE, D. A.; PAWAR, A. V.; KOTECHA, K. IT infrastructure anomaly detection and failure handling: A systematic literature review focusing on datasets, log preprocessing, machine & deep learning approaches and automated tool. **IEEE Access**, Institute of Electrical and Electronics Engineers (IEEE), v. 9, p. 156392–156421, 2021.
- BONILHA, C. **Wireless LAN (WLAN)**. Teleco, 2003. Acessado em 05/10/2022. Disponível em: <www.telecom.com.br/wlan>.
- BRIETZIG, N. G. **Análise de dados como ferramenta de tomada de decisão para micro e pequenas empresas**. Dissertação (Mestrado) — Universidade do Sul de Santa Catarina - Unisul, 2022.
- ÇEKEN, C.; YARKAN, S.; ARSLAN, H. Interference aware vertical handoff decision algorithm for quality of service support in wireless heterogeneous networks. **Computer Networks**, Elsevier BV, v. 54, n. 5, p. 726–740, apr 2010.
- CISCO. **The Benefits of Centralization in Wireless LANs via the Cisco Unified Wireless Network**. Cisco, 2006. Acessado em 05/10/2022. Disponível em: <https://www.cisco.com/web/AP/wireless/pdf/Benefits_of_centralizedWlan.pdf>.
- CISCO. **802.11ac: The Fifth Generation of Wi-Fi**. Cisco, 2018. Acessado em 16/10/2022. Disponível em: <<https://www.cisco.com/c/dam/en/us/products/collateral/wireless/aironet-3600-series/white-paper-c11-713103.pdf>>.
- CISCO. **IEEE 802.11ax: The Sixth Generation of Wi-Fi**. Cisco, 2020. Acessado em 22/10/2022. Disponível em: <<https://www.cisco.com/c/en/us/products/collateral/wireless/white-paper-c11-740788.html>>.
- CISCO. **Wi-Fi 6E: The Next Great Chapter in Wi-Fi White Paper**. Cisco, 2021. Acessado em 22/10/2022. Disponível em: <<https://www.cisco.com/c/en/us/solutions/collateral/enterprise-networks/802-11ax-solution/nb-06-wi-fi-6e-wp-cte-en.html?oid=wprwls024270>>.
- DANALET, A.; FAROOQ, B.; BIERLAIRE, M. A bayesian approach to detect pedestrian destination-sequences from WiFi signatures. **Transportation Research Part C: Emerging Technologies**, Elsevier BV, v. 44, p. 146–170, jul 2014.

- DOMINGOS, N. **Análise de Dados: Como resolver problemas em 4 etapas**. Medium, 2019. Acessado em 02/11/2022. Disponível em: <<https://medium.com/@natanaeldgsantos/analise-de-dados-como-resolver-problemas-em-4-etapas-1914a6b9b54>>.
- DU, M. et al. Deeplog: Anomaly detection and diagnosis from system logs through deep learning. **Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security**, Association for Computing Machinery, p. 1285–1298, oct 2017.
- DUAN, X. et al. Oilog: An online incremental log keyword extraction approach based on mdp-lstm neural network. **Information Systems, Volume 95**, 2021.
- FCC. **Broadband Speed Guide**. Federal Communications Commission, 2023. Acessado em 07/01/2023. Disponível em: <<https://www.fcc.gov/consumers/guides/broadband-speed-guide>>.
- FILHO, J. G. Q. Dispatch: Uma abordagem sdwn para o gerenciamento do processo de handoff nas redes wi-fi. **Universidade Federal da Paraíba**, 2016.
- FOLIUM. **Python data, leaflet.js maps**. Folium, 2022. Acessado em 08/11/2022. Disponível em: <<https://python-visualization.github.io/folium/>>.
- FREERADIUS. **Documentation - A FreeRADIUS resource portal**. Freeradius, 2022. Acessado em 12/09/2022. Disponível em: <<https://freeradius.org/documentation/>>.
- FSCOMMUNITY. **Wireless LAN Controller Explained and FAQs**. FSCommunity, 2023. Acessado em 10/08/2023. Disponível em: <<https://community.fs.com/blog/wireless-lan-controller-explained.html>>.
- GANJI, F.; BUDZISZ, L.; WOLISZ, A. Assessment of the power saving potential in dense enterprise WLANs. **IEEE 24th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)**, Institute of Electrical and Electronics Engineers (IEEE), sep 2013.
- GERASIMENKO, M. et al. Characterizing performance of load-aware network selection in multi-radio (WiFi/LTE) heterogeneous networks. **IEEE Globecom Workshops (GC Wkshps)**, Institute of Electrical and Electronics Engineers (IEEE), dec 2013.
- GHOSH, A. et al. Modeling and characterization of large-scale wi-fi traffic in public hot-spots. **2011 Proceedings IEEE INFOCOM**, Institute of Electrical and Electronics Engineers (IEEE), apr 2011.
- GIBNEY, A. M.; KLEPAL, M.; PESCH, D. A wireless local area network modeling tool for scalable indoor access point placement optimization. In: **Proceedings of the 2010 Spring Simulation Multiconference**. [S.l.]: Society for Computer Simulation International, 2010.
- GIL, A. C. **Métodos e técnicas de pesquisa social**. [S.l.]: Editora Atlas - 6 ed. - São Paulo, 2008.
- GONDRAN, A. et al. Wireless LAN planning: a didactical model to optimise the cost and effective payback. **International Journal of Mobile Network Design and Innovation**, Inderscience Publishers, v. 2, n. 1, p. 13, 2007.
- HASSANI, M. et al. Studying and detecting log-related issues. **Empirical Software Engineering**, Springer Science and Business Media LLC, v. 23, n. 6, p. 3248–3280, mar 2018.
- HILLS, A. Large-scale wireless LAN design. **IEEE Communications Magazine**, Institute of Electrical and Electronics Engineers (IEEE), v. 39, n. 11, p. 98–107, 2001.
- HU, L. et al. How much can wi-fi offload? a large-scale dense-urban indoor deployment study. In: **2012 IEEE 75th Vehicular Technology Conference (VTC Spring)**. [S.l.]: IEEE, 2012.

- HUAWEI. **Huawei Wireless Access Controller**. Huawei, 2022. Acessado em 29/09/2022. Disponível em: <<https://e.huawei.com/en/material/networking/wlan/d7ceca307a0f42fbb8ee5e417bae8d40>>.
- IEEE. Standard for information technology– local and metropolitan area networks– specific requirements– part 11: Wireless lan medium access control (mac)and physical layer (phy) specifications amendment 5: Enhancements for higher throughput. **IEEE Std 802.11n-2009 (Amendment to IEEE Std 802.11-2007 as amended by IEEE Std 802.11k-2008, IEEE Std 802.11r-2008, IEEE Std 802.11y-2008, and IEEE Std 802.11w-2009)**, 2009.
- JAIN, S. et al. Extracting the textual and temporal structure of supercomputing logs. **2009 International Conference on High Performance Computing (HiPC)**, Institute of Electrical and Electronics Engineers (IEEE), dec 2009.
- JARDOSH, A. P. et al. Green WLANs: On-demand WLAN infrastructures. **Mobile Networks and Applications**, Springer Science and Business Media LLC, v. 14, n. 6, p. 798–814, dec 2008.
- KAMENETSKY, M.; UNBEHAUN, M. Coverage planning for outdoor wireless LAN systems. **International Zurich Seminar on Broadband Communications Access - Transmission - Networking (Cat. No.02TH8599)**, Institute of Electrical and Electronics Engineers (IEEE), 2022.
- KBAR, G.; MANSOOR, W. Distributed resources management in wireless LANs. In: **Selected Readings on Telecommunications and Networking**. [S.l.]: IGI Global, 2006. p. 293–304.
- KOBAYASHI, H. et al. Towards sustainable heterogeneous wireless networks: A decision strategy for AP selection with dynamic graphs. **Computer Networks**, Elsevier BV, v. 132, p. 99–107, feb 2018.
- KONONENKO, I.; KUKAR, M. Machine learning and data mining: introduction to principles and algorithms. **Horwood Publishing**, 2007.
- KUO, W.-H.; LIAO, W. Utility-based radio resource allocation for QoS traffic in wireless networks. **IEEE Transactions on Wireless Communications**, Institute of Electrical and Electronics Engineers (IEEE), v. 7, n. 7, p. 2714–2722, jul 2008.
- LABCISCO. **Padrão IEEE 802.11ax de WiFi de Próxima Geração**. LabCisco, 2023. Acessado em 10/08/2023. Disponível em: <<http://labcisco.blogspot.com/2017/10/padrao-ieee-80211ax-de-wifi-de-proxima.html>>.
- LIMA, L. J. B. et al. Qualidade de serviços em redes do padrão 802.11. **V CTDA - Universidade Federal Rural da Amazônia - UFRA**, 2002.
- MATEUS, G. R.; LOUREIRO, A. A.; RODRIGUES, R. C. **Annals of Operations Research**, Springer Science and Business Media LLC, v. 106, n. 1/4, p. 331–345, 2001.
- MCKINNEY, W. **Python para análise de dados: Tratamento de dados com Pandas, NumPy e IPython**. [S.l.]: Novatec Editora, 2019.
- MELO, J. C. de. Estudos da utilização de mecanismos de qos em redes com enlaces de banda estreita. **Universidade Federal do Maranhão, Curso de pós-graduação em Engenharia de Eletricidade**, 2005.
- MENESES, F.; MOREIRA, A. Large scale movement analysis from WiFi based location data. In: **2012 International Conference on Indoor Positioning and Indoor Navigation (IPIN)**. [S.l.]: Institute of Electrical and Electronics Engineers (IEEE), 2012.

- MIZUTANI, M. Incremental mining of system log format. In: **2013 IEEE International Conference on Services Computing**. [S.l.]: Institute of Electrical and Electronics Engineers (IEEE), 2013.
- NETGEAR. **The Benefits of Wireless LAN Controller in Small and Mid-sized Business Networks**. Netgear, 2007. Acessado em 03/09/2022. Disponível em: <<https://usermanual.wiki/Document/NetGearWFS709TPProSafeSmartWirelessControllerWhitePaper.1310704068.pdf>>.
- NETWORKS, A. **Controllers, Cloud, and Cooperative Control - The 5 Generations Of WLAN Architecture**. Aerohive Networks, 2015. Acessado em 09/01/2023. Disponível em: <https://www.aerohive.com/wp-content/uploads/Aerohive_Whitepaper_Controllers_Cloud_Cooperative_Control.pdf>.
- NUMPY. **NumPy documentation**. Numpy, 2022. Acessado em 02/10/2022. Disponível em: <<https://numpy.org/>>.
- OPENSTREETMAP. **Documentação do OpenStreetMap**. OSM, 2021. Acessado em 21/10/2022. Disponível em: <<http://www.openstreetmap.org/traces/>>.
- PANDAS. **Documentation Pandas**. Pandas, 2022. Acessado em 28/10/2022. Disponível em: <<https://pandas.pydata.org/>>.
- PENG, M. et al. Network planning based on interference alignment in density WLANs. **Institute of Electrical and Electronics Engineers (IEEE) Access**, IEEE, v. 7, p. 70525–70534, 2019.
- PLOTLY. **Plotly Open Source Graphing Library for Python**. Plotly, 2022. Acessado em 05/11/2022. Disponível em: <<https://plotly.com/python/>>.
- RADIUS. **Documentation FreeRadius**. Radius, 2022. Acessado em 05/10/2022. Disponível em: <<https://freeradius.org/documentation/>>.
- RADIUS, N. **How does RADIUS Accounting work?** Network Radius, 2021. Acessado em 02/12/2022. Disponível em: <<https://networkradius.com/articles/2019/06/05/how-does-radius-accounting-work.html>>.
- RAJKUMAR, R. R. et al. Cyber-physical systems. **Proceedings of the 47th Design Automation Conference on - DAC '10**, ACM Press, 2010.
- ROCHOL, J. **Sistemas de comunicação sem fio - conceitos e aplicações**. [S.l.]: Porto Alegre: Bookman, 2018.
- RODRIGUEZ-LOZANO, D. et al. Context-aware prediction of access points demand in wi-fi networks. **Computer Networks**, Elsevier BV, v. 117, p. 52–61, apr 2017.
- RUCKUS. **Ruckus Zone Director: Enterprise-class smart wireless LAN controller**. Ruckus Networks, 2021. Acessado em 11/09/2022. Disponível em: <<https://www.commscope.com/globalassets/digizuite/61801-ds-zonedirector-1200.pdf>>.
- SANCHES, C. A. **Projetando redes WLAN - Conceitos e Práticas**. [S.l.]: São Paulo: Editora Érica, 2011.
- SCHIMUNECK, M. A. K. et al. Análise da qualidade de experiência em redes sem fio. **XX Seminário de Iniciação Científica, Universidade de Santa Cruz do Sul**, 2014.
- SEABORN. **Seaborn: statistical data visualization**. Seaborn, 2022. Acessado em 30/10/2022. Disponível em: <<https://seaborn.pydata.org/>>.

- SHARDA, R.; DELEN, D.; TURBAN, E. **Business Intelligence e Análise de Dados para Gestão do Negócio**. [S.l.]: Bookman - 4 ed., 2019.
- SMITH JAKE WOODHAMS, R. M. J. **Controller-Based Wireless LAN Fundamentals**. 1st. ed. [S.l.]: Cisco Press, 2011. 320 p.
- STARTDUST. **QoS protocols & architectures - White paper**. Startdust, 1999. Acessado em 10/10/2022. Disponível em: <<https://sites.cs.ucsb.edu/~almeroth/classes/F01.201B/papers/qos.pdf>>.
- SYSTEMS, C. **Annual Internet Report**. Cisco, 2018. Acessado em 16/10/2022. Disponível em: <<https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.pdf>>.
- TANG, D.; BAKER, M. Analysis of a local-area wireless network. **Proceedings of the 6th annual international conference on Mobile computing and networking - MobiCom**, ACM Press, 2000.
- VALDMAN, J. **Log file analysis**. [S.l.]: University of West Bohemia in Pilsen, Czech Republic, 2001.
- WANG, Y.-T. et al. The implementation of wi-fi log analysis system with elk stack. In: **Frontier Computing**. Singapore: Springer Singapore, 2019. p. 46–255.
- ZHONG, Z. et al. Issues and challenges in dense WiFi networks. In: **2015 International Wireless Communications and Mobile Computing Conference (IWCMC)**. [S.l.]: IEEE, 2015.
- ZVANOVEC, S.; PECHAC, P.; KLEPAL, M. Wireless lan networks design: Site survey or propagation modeling? **RADIOENGINEERING, VOL. 12, NO. 4**, 2003.

APÊNDICE A

Algorithm 1 Log2CSV

```

1: function LOG2DF(log_file)
2:   READ (log_file)
3:   data = Readlines
4:   while line IN data do
5:     if line.strip() EQUAL "" then
6:       log_dict[counter] = new_element
7:       Increment counter
8:     if IS_DATE(line) then
9:       CONVERT_DATE(log_dict[counter])
10:  return DataFrame(log_dict)
11: while file IN listdir(current) do
12:  if file.startwith(detail) then
13:    df = log2df(file)
14:    SAVE df to CSV (file +"CSV")

```

Algorithm 2 Convert_mac_local

```

1: OPEN (aps_local_final.csv)
2: CREATE df_mac FROM aps_local_final.csv
3: reader = df_mac.readlines
4: while rown IN reader do
5:   dict_from_csv = df_mac.lines
6:   dict_from_csv = CONVERT(lowercase(dict_from_csv))
7: while file IN listdir(current) do
8:  if file.startwith(detail) then
9:    df = READ(file)
10:   df = REPLACE dict_from_csv.keys(), value = dict_from_csv.values()
11:   SAVE df to CSV (file +"Local"+"CSV")

```

Algorithm 3 Create_BD

```

1: engine = connection_bd (postgres:userbd@patchbd)
2: READ table.engine()
3: table = conexões
4: while file IN listdir(current) do
5:   if file.startwith(detail) then
6:     df = READ(file)
7:     SAVE df to SQL(table)

```

Algorithm 4 View_BD

```

1: CREATE View
2: READ table.engine()
3: table = conexões
4: for row in table do:
5:   CONVERT 'date' IN 'day', 'hour'
6:   Momento = SPLIT Column 'hour'
7:   table.intervalo = COUNT(SELECT conexão_agrupada.Local, conexão_agrupada.Usuario IN
"momento").unique

```

Algorithm 5 Create_CSV

```

1: engine = connection_bd (postgres:userbd@patchbd)
2: READ table.engine()
3: query_sql = select table ("intervalo")
4: READ query_sql
5: SAVE query_sql to CSV("contagem.csv")

```

Algorithm 6 Insert_lat_long

```

1: OPEN (contagem.csv, local_lat_long.csv)
2: CREATE dataframe_main FROM contagem.csv
3: CREATE dataframe_local FROM local_lat_long.csv
4: INSERT COLUMNS "latitude", "longitude" IN dataframe_main
5: reader = Readlines
6: for row IN reader do
7:   if dataframe_main["acesspoint"] EQUAL dataframe_local["local"] then
8:     INSERT latitude, longitude
9:   else
10:    Trata_erro
11: SAVE df to CSV ("contagem_lat_long.csv")

```

Algorithm 7 Anonymous

```

1: OPEN (contagem_lat_long.csv, usuarios.csv)
2: CREATE dataframe_main FROM contagem_lat_long.csv
3: CREATE dataframe_user FROM usuarios.csv
4: INSERT COLUMN 'categoria' IN dataframe_main
5: reader = Readlines
6: for row IN reader do
7:   if dataframe_main["usuario"] EQUAL dataframe_user["login"] then
8:     INSERT categoria
9:     dataframe_main['usuario'] = dataframe_user['categoria']+Counter
10:  else
11:    Trata_erro
12: SAVE df to CSV (base_final_anonimizada.csv)

```

Algorithm 8 HeatMap

```

1: OPEN (contagem_lat_long.csv)
2: CREATE dataframe_main FROM contagem_lat_long.csv
3: REMOVE period[23:01 at 06:59] FROM dataframe_main["momento"]
4: reader = Readlines
5: for row IN reader do
6:   Fullcoord = Count[dataframe_main["latitude"], dataframe_main["longitude"]] IN dataframe_main["momento"]
7: Generate MAP(Fullcoord)

```

Algorithm 9 Mapa categoria

```

1: OPEN (base_final_anonimizada.csv)
2: CREATE dataframe_main FROM base_final_anonimizada.csv
3: REMOVE period[23:01 at 06:59] FROM dataframe_main['momento']
4: df_count = dataframe_main.groupby('categoria').count.newcolumn('counts')
5: Generate MAP(df_count['counts'])

```
