

# Aprendizado por Reforço Distribuído para Otimização da Vazão em Redes TSCH

Vanessa N. Silva<sup>1</sup>, Guilherme Milanez<sup>1</sup>, Marcos A. M. Vieira<sup>1</sup>,  
Luiz Filipe M. Vieira<sup>1</sup>, José Augusto M. Nacif<sup>2</sup>

<sup>1</sup>Departamento de Ciência da Computação – Universidade Federal de Minas Gerais

<sup>2</sup>Instituto de Ciências Exatas e Tecnológicas – Universidade Federal de Viçosa

{nascimento.vanessa, guilhermeaguarmilanez,  
lfvieira, mmvieira}@dcc.ufmg.br, jnacif@ufv.br

**Abstract.** *The Time Slotted Channel Hopping (TSCH) protocol has been widely adopted in industrial and Internet of Things (IoT) scenarios due to its reliability and energy efficiency. However, despite the existence of several scheduling mechanisms focused on robustness, there remains a gap regarding throughput optimization, one of the most relevant metrics in network performance evaluation. In this context, this work proposes a distributed reinforcement learning approach based on Q-learning for throughput optimization in TSCH networks, in which each sensor node acts as an autonomous agent that locally learns the best transmission slot allocation over time. The solution was implemented in Contiki-NG and employs a multi-criteria reward function capable of capturing different aspects of network performance, without requiring centralized control or excessive information exchange among nodes, thereby reducing communication overhead. The learning process is continuous, enabling dynamic adaptation to network conditions, and the obtained results indicate the potential of reinforcement learning as an effective mechanism for adaptive scheduling in TSCH networks.*

**Resumo.** *O protocolo Time Slotted Channel Hopping (TSCH) tem sido amplamente adotado em cenários industriais e de Internet das Coisas devido à sua confiabilidade e eficiência energética; contudo, apesar da existência de diversos mecanismos de escalonamento focados na robustez, ainda há uma lacuna no que se refere à otimização da vazão, uma das métricas mais relevantes em redes. Nesse contexto, este trabalho propõe uma abordagem de aprendizado por reforço distribuído baseada em Q-learning para a otimização da vazão em redes TSCH, na qual cada nó sensor atua como um agente autônomo que aprende localmente a melhor alocação de slots de transmissão ao longo do tempo. A solução foi implementada no Contiki-NG e emprega uma função de recompensa multicritério capaz de capturar diferentes aspectos do desempenho da rede, sem a necessidade de controle centralizado ou de troca excessiva de informações entre os nós, reduzindo a sobrecarga de comunicação. O processo de aprendizado é contínuo, permitindo adaptação dinâmica às condições da rede, e os resultados obtidos indicam o potencial do aprendizado por reforço como mecanismo eficaz para o escalonamento adaptativo em redes TSCH.*

## 1. Introdução

Redes de sensores sem fio (WSN) baseadas no padrão IEEE 802.15.4e TSCH (*Time-Slotted Channel Hopping*) [Duquennoy et al. 2017] têm se consolidado como uma tecnologia fundamental para aplicações críticas de Internet das Coisas (IoT), oferecendo comunicação determinística, elevada confiabilidade e eficiência energética. Nesse protocolo, a comunicação é organizada em ciclos periódicos de tempo composto por *slots*, formando um *slotframe* que se repete continuamente. Essa estrutura possibilita a sincronização precisa entre nós, redução de interferências por meio de saltos de canais e redução de colisões, características essenciais para cenários como automação industrial, cidades inteligentes, agricultura de precisão e monitoramento ambiental. Com a projeção de mais de 75 bilhões de dispositivos IoT conectados nos próximos anos [Salih et al. 2022], torna-se cada vez mais crítico o desenvolvimento de protocolos MAC eficientes e adaptativos.

Entretanto, o desempenho dessas redes depende fortemente da configuração do tamanho do *slotframe*, que define a periodicidade da comunicação e impacta diretamente métricas como vazão, latência, consumo energético e taxa de entrega de pacotes. A otimização da vazão em redes TSCH é um desafio crítico [dos Santos Ribeiro et al. 2022], pois configurações inadequadas degradam o desempenho em cenários com alta densidade de nós e tráfego variável. Embora técnicas adaptativas possam alcançar ganhos superiores a 60% em vazão [dos Santos Ribeiro et al. 2022], muitas implementações ainda utilizam configurações estáticas, como *slotframes* de 7, 36 ou 101 *slots*. *Slotframes* curtos aumentam a contenção e as colisões, enquanto *slotframes* longos desperdiçam recursos, elevam a latência e o consumo energético, o que é particularmente problemático em aplicações IoT com tráfego dinâmico.

A adaptação dinâmica do escalonamento TSCH em ambientes descentralizados impõe desafios relevantes. Primeiramente, cada nó deve ser capaz de tomar decisões locais sem depender de controle centralizado, evitando pontos únicos de falha e reduzindo sobrecarga de sinalização. Além disso, o sistema precisa equilibrar exploração e aproveitamento, aprendendo continuamente configurações eficientes enquanto se adapta a mudanças no tráfego. Tais mecanismos devem operar sob severas restrições de processamento, memória e energia, típicas de dispositivos IoT. Por fim, estratégias de compartilhamento de conhecimento entre nós devem acelerar a convergência sem introduzir custos excessivos de comunicação.

Neste contexto, este trabalho propõe uma abordagem de aprendizado por reforço hierárquico e distribuído para a otimização adaptativa do escalonamento TSCH, combinando *Q-Learning* tabular com aprendizado federado descentralizado. A solução opera em dois níveis complementares: no nível superior, cada nó ajusta dinamicamente o tamanho do *slotframe*; no nível inferior, os *slots* individuais são configurados de forma adaptativa quanto à sua ativação, tipo de operação e deslocamento de canal, visando reduzir colisões e melhorar o uso dos recursos da rede. Para acelerar a convergência, os nós compartilham conhecimento com vizinhos por meio de técnicas leves de aprendizado federado, como *FedAvg*, *FedMedian* e *Weighted FedAvg* [Mora et al. 2024], sem necessidade de coordenação central.

As principais contribuições deste trabalho incluem: (i) a proposta de um algoritmo de *Q-Learning* hierárquico e totalmente distribuído para ajuste dinâmico do *slotframe* e

dos *slots* individuais; (ii) a incorporação de um mecanismo de aprendizado federado descentralizado compatível com dispositivos IoT de baixo consumo; (iii) o uso de uma função de recompensa multiobjetivo que equilibra vazão, estabilidade do *buffer* e confiabilidade da comunicação; e (iv) a implementação da solução no sistema operacional Contiki-NG, com avaliação experimental no simulador Cooja, demonstrando ganhos expressivos de vazão, taxa de entrega e redução de retransmissões em comparação com configurações estáticas tradicionais.

Diferentemente de abordagens existentes que utilizam heurísticas fixas ou mecanismos reativos simples, a solução proposta aprende continuamente a partir da interação com o ambiente, adaptando-se a padrões específicos de tráfego e topologia. Além disso, o uso de aprendizado federado permite acelerar a convergência do aprendizado sem comprometer a natureza distribuída do TSCH.

O restante deste artigo está organizado da seguinte forma. A Seção 2 discute o referencial teórico. A Seção 3 traz os trabalhos relacionados. A Seção 4 apresenta o modelo do sistema e o método *Q-learning* proposto. A Seção 5 descreve os cenários de avaliação e apresenta os resultados experimentais e discute os resultados obtidos. A Seção 6 conclui o trabalho.

## 2. Referencial Teórico

Esta seção apresenta os fundamentos teóricos que embasam este trabalho. O trabalho utiliza aprendizado federado descentralizado para computar o escalonamento TSCH em redes IoT.

### 2.1. TSCH e Escalonamento em Redes IoT

O protocolo *Time-Slotted Channel Hopping* (TSCH) [Vilajosana et al. 2017] organiza o acesso ao meio em *slots* de tempo discretos agrupados em *slotframes* periódicos. Cada *slot* pode ser configurado para transmissão, recepção, operação compartilhada ou permanecer inativo, enquanto o mecanismo de salto de canal explora diversidade de frequência para mitigar interferências externas. Essa combinação confere ao TSCH propriedades fundamentais para aplicações IoT críticas, como comunicação determinística, alta confiabilidade e eficiência energética.

O desempenho de redes TSCH depende diretamente da configuração do *slotframe*, especialmente de seu tamanho. *Slotframes* reduzidos tendem a minimizar latência, porém limitam a capacidade da rede e aumentam a probabilidade de colisões em cenários densos. Por outro lado, *slotframes* extensos frequentemente resultam em desperdício de recursos, maior latência e consumo energético elevado em ambientes com tráfego moderado ou esporádico. A arquitetura 6TiSCH, definida na RFC 8180 [Vilajosana et al. 2017], estabelece diretrizes para a integração do TSCH com a pilha IPv6, mas não especifica mecanismos adaptativos para ajuste do escalonamento, delegando essa responsabilidade aos diversos algoritmos de escalonamento. Diversos estudos indicam que configurações estáticas de *slotframe* são inadequadas em cenários com variação dinâmica de carga, podendo levar a degradações expressivas de vazão e taxa de entrega [dos Santos Ribeiro et al. 2022]. Esse contexto motiva o desenvolvimento de mecanismos adaptativos capazes de ajustar o escalonamento TSCH de forma eficiente e autônoma.

## 2.2. Aprendizado por Reforço e Q-Learning

O aprendizado por reforço (*Reinforcement Learning* – RL) modela problemas de tomada de decisão sequencial como um Processo de Decisão Markoviano (MDP), caracterizado por estados, ações, transições e recompensas [Sutton and Barto 2018]. Dentre os algoritmos de RL, o *Q-Learning* destaca-se por ser livre de modelo, aprendendo diretamente a função valor-ação sem conhecimento prévio da dinâmica do ambiente. A atualização iterativa da função Q permite que o agente refine sua política ao longo do tempo, equilibrando exploração e aproveitamento por meio de estratégias como *epsilon-greedy*. Em redes sem fio e IoT, técnicas de RL vêm sendo aplicadas com sucesso em problemas como seleção dinâmica de canais, controle de potência, roteamento adaptativo e alocação de recursos [Naparstek and Cohen 2019]. No entanto, a adoção dessas técnicas em dispositivos IoT impõe restrições severas de memória, processamento e energia, exigindo soluções leves, tabulares e com baixo overhead computacional, como o *Q-Learning* clássico.

## 2.3. Aprendizado Federado Descentralizado

O aprendizado federado (*Federated Learning* – FL) possibilita o treinamento colaborativo de modelos distribuídos sem a necessidade de compartilhamento de dados brutos, reduzindo a sobrecarga de comunicação e preservando a privacidade. No método FedAvg, os parâmetros dos modelos locais são agregados por meio de média ponderada, enquanto variantes como FedMedian aumentam robustez a outliers. Em redes IoT, o FL enfrenta desafios adicionais, como comunicação intermitente, heterogeneidade de dispositivos e ausência de infraestrutura central confiável. Arquiteturas descentralizadas ou ponto a ponto surgem como alternativas viáveis, eliminando pontos únicos de falha, embora demandem estratégias de agregação leves e eficientes. Trabalhos recentes demonstram que técnicas de FL adaptadas ao contexto IoT podem reduzir significativamente a sobrecarga de comunicação, tornando-se adequadas mesmo para ambientes com recursos restritos.

## 3. Trabalhos Relacionados

### 3.1. Escalonamento Adaptativo em TSCH

Abordagens clássicas de escalonamento em TSCH baseiam-se majoritariamente em heurísticas fixas. O MSF [Chang et al. 2021], por exemplo, ajusta células com base em limiares estáticos de utilização, sendo suscetível a oscilações sob cargas não estacionárias. Outras propostas descentralizadas adaptam-se ao tráfego, mas mantêm *slotframes* fixos e capacidade limitada de adaptação.

Extensões como o *Load-based Distributed Scheduling Function* (LDSF) incorporam previsão de carga para reduzir a sobrecarga, porém continuam dependentes de *slotframes* estáticos. Em contraste, soluções baseadas em *Software-Defined Networking* (SDN), como TASA [Palattella et al. 2012] e VARIBAN [Milanez et al. 2023], alcançam bons resultados de latência e vazão, ao custo de maior sobrecarga e menor escalabilidade.

### 3.2. Aprendizado de Máquina na Camada MAC para IoT

Diversos trabalhos exploram aprendizado de máquina na camada MAC. Abordagens baseadas em RL, como AdaMAC, utilizam *Q-Learning* para seleção adaptativa de protocolos MAC, obtendo ganhos energéticos, porém sem considerar TSCH ou ajuste de *slotframe*. Métodos como RLCS focam na seleção dinâmica de canais em redes não sincronizadas,

não sendo diretamente aplicáveis ao TSCH. Outras propostas utilizam aprendizado supervisionado ou modelos de séries temporais para previsão de qualidade de *link* ou carga de tráfego, como Random Forest e LSTM. Embora eficazes como ferramentas auxiliares, essas técnicas geralmente dependem de treinamento *offline* ou infraestrutura de borda, não oferecendo soluções completas para escalonamento adaptativo descentralizado.

### 3.3. Aprendizado Federado em Redes IoT

*Frameworks* como FedIoT-Plus e arquiteturas hierárquicas de FL demonstram o potencial do aprendizado colaborativo em ambientes IoT heterogêneos. Entretanto, a maioria dessas soluções foca em aprendizado supervisionado ou profundo, exigindo recursos computacionais incompatíveis com redes de sensores sem fio de baixa potência. Propostas mais próximas ao contexto deste trabalho, como FedChannel, exploram FL ponto a ponto para otimização de redes sem fio, inspirando mecanismos colaborativos descentralizados. Ainda assim, essas soluções não consideram as especificidades do TSCH nem a integração direta com o escalonamento MAC.

### 3.4. Q-Learning Hierárquico e Otimização Multiobjetivo

Abordagens hierárquicas de aprendizado visam reduzir a complexidade do espaço de decisões por meio de múltiplos níveis de controle. Trabalhos demonstram que hierarquias de *Q-Learning* podem melhorar eficiência energética em MACs convencionais [Savaglio et al. 2019]. De forma complementar, soluções de otimização multiobjetivo utilizam funções de recompensa ponderadas para equilibrar métricas concorrentes, embora frequentemente dependam de pesos fixos ou meta-aprendizado *offline*. Apesar dos avanços, observa-se uma lacuna na literatura quanto à combinação de aprendizado hierárquico, otimização multiobjetivo e colaboração federada especificamente para escalonamento TSCH em ambientes totalmente descentralizados.

### 3.5. Posicionamento do Trabalho

À luz da literatura, observa-se uma lacuna na integração de aprendizado por reforço hierárquico, otimização multiobjetivo e aprendizado federado descentralizado aplicados diretamente ao escalonamento TSCH. Este trabalho diferencia-se ao propor uma solução totalmente distribuída, que ajusta dinamicamente tanto o tamanho do *slotframe* quanto a configuração individual dos *slots*, sem dependência de controle central ou infraestrutura de borda.

Embora existam propostas recentes baseadas em *Q-Learning* para TSCH [Ara and Liscano 2024, Pratama et al. 2024, Ben Yaala et al. 2025, Cardel 2025], essas diferem substancialmente em arquitetura, objetivos de otimização e nível de descentralização, além de não incorporarem aprendizado federado nem otimização multiobjetivo explícita. Diante disso, optou-se por comparações com baselines estáticos amplamente utilizados, garantindo uma avaliação clara, reproduzível e alinhada ao estado da prática. A implementação em Contiki-NG e a validação experimental reforçam a viabilidade da proposta em dispositivos IoT com recursos limitados.

## 4. Metodologia

Este trabalho propõe uma abordagem distribuída de aprendizado por reforço para o escalonamento adaptativo em redes *Time Slotted Channel Hopping* (TSCH), com foco na

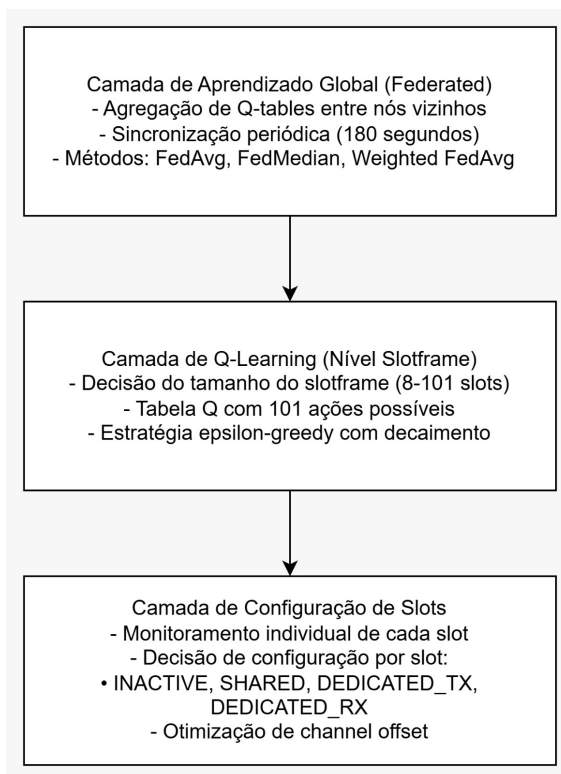
otimização da vazão da rede e na redução de retransmissões. A solução baseia-se em Q-learning hierárquico, operando de forma descentralizada em cada nó da rede, sem a necessidade de um controlador central.

A metodologia foi concebida para atuar diretamente sobre o mecanismo de escalonamento do TSCH, permitindo ajustes dinâmicos tanto no tamanho do *slotframe* quanto na configuração individual dos *slots*, explorando métricas locais de desempenho observadas pelos nós sensores. A implementação foi realizada sobre o Contiki-NG, utilizando extensões customizadas da pilha TSCH.

A proposta foi implementada no Contiki-NG, utilizando o protocolo TSCH como camada MAC. O sistema opera com *slotframes* adaptativos, *buffers* de pacotes limitados e comunicação UDP para geração de tráfego de aplicação. O aprendizado ocorre de forma contínua durante a execução da rede, permitindo adaptação dinâmica às condições do ambiente.

#### 4.1. Visão Geral da Abordagem

Cada nó da rede é modelado como um agente de aprendizado por reforço que interage continuamente com o ambiente TSCH. O processo de aprendizado ocorre de forma *on-line*, permitindo que o agente adapte suas decisões às condições dinâmicas da rede, como variações de tráfego, ocupação do *buffer* e número de retransmissões.



**Figura 1. Arquitetura Hierárquica**

A arquitetura proposta, como explicada na Figura 1, adota uma abordagem hierárquica composta por dois níveis principais:

1. Nível de *slotframe*, responsável por decidir dinamicamente o tamanho do quadro de alocação de canal e tempo de comunicação, variando entre 8 e 101 slots.

2. Nível de *slots* individuais, responsável por ajustar a configuração de cada *slot*, podendo ativá-lo, desativá-lo ou convertê-lo em *slot* dedicado de transmissão ou recepção.

Essa separação permite reduzir a complexidade do espaço de estados do aprendizado e viabiliza ajustes mais finos no escalonamento da rede.

## 4.2. Modelagem do Problema como um MDP

O problema de escalonamento é modelado como um Processo de Decisão de Markov (MDP), definido pelo conjunto de estados, ações e recompensas.

### 4.2.1. Estados

O estado do agente é definido a partir de métricas locais observadas pelo nó, incluindo:

1. Tamanho do *buffer* de transmissão;
2. Métricas agregadas de transmissões e recepções bem-sucedidas;
3. Número médio de retransmissões por pacote.

Essas informações permitem ao agente inferir o nível de congestionamento da rede e a eficiência do escalonamento atual.

### 4.2.2. Ações:

O conjunto de ações é composto por 101 ações discretas, em que cada ação representa um possível tamanho de *slotframe*. A escolha desse número de ações justifica-se por equilibrar granularidade na exploração do espaço de decisões e viabilidade computacional em dispositivos IoT com recursos limitados. Esse intervalo permite ajustes finos no tamanho do *slotframe*, cobrindo desde configurações mínimas que reduzem latência até configurações extensas que acomodam tráfego elevado, mantendo a complexidade da tabela Q em níveis compatíveis com as restrições de memória típicas de nós sensores (tabelas Q com 101 entradas por estado são factíveis em dispositivos com poucos kB de RAM). O mapeamento entre ação e tamanho do *slotframe* é dado por:

$$slotframesize = 8 + \left\lceil \frac{93 \times a}{100} \right\rceil \quad (1)$$

em que  $a \in [0, 100]$ ; 8 é o tamanho mínimo de *slotframe* (quando  $a = 0$ ), 93 é a amplitude do intervalo ( $101 - 8 = 93$ ) e  $\frac{a}{100}$  normaliza a ação  $a$  para o intervalo  $[0,1]$ . Assim, a ação 0 corresponde ao menor *slotframe* permitido (8 slots) e a ação 100 ao maior *slotframe* (101 slots), cobrindo todo o espectro de configurações práticas em redes TSCH.

### 4.2.3. Função de Recompensa

A função de recompensa foi projetada para capturar múltiplos aspectos do desempenho da rede, sendo definida como:

$$R(s, a) = \theta_1(n_{t_x} + n_{r_x}) - \theta_2|\Delta buffer| - \theta_3(\overline{retrans} - 1) \quad (2)$$

em que:

1.  $n_{t_x}$  e  $n_{r_x}$  representam o número de transmissões e recepções bem-sucedidas no período;
2.  $\Delta buffer$  indica a variação do tamanho do *buffer*;
3. *retrans* corresponde à média de retransmissões por pacote;
4.  $\theta_1, \theta_2, \theta_3$  são pesos empíricos que priorizam vazão, estabilidade do *buffer* e confiabilidade da comunicação, respectivamente.

Essa formulação incentiva o aumento da vazão da rede, ao mesmo tempo em que penaliza congestionamento e retransmissões excessivas.

### 4.3. Algoritmo de Q-Learning

A atualização da tabela Q segue a regra clássica do *Q-learning*:

$$Q(s, a) \leftarrow (1 - \alpha)Q(s, a) + \alpha [R(s, a) + \gamma \max_{a'} Q(s', a')] \quad (3)$$

em que  $\alpha$  é a taxa de aprendizado e  $\gamma$  o fator de desconto. A política de seleção de ações utiliza a estratégia  *$\epsilon$ -greedy*, equilibrando exploração e exploração ao longo do tempo, com decaimento progressivo de  $\epsilon$ .

### 4.4. Configuração Adaptativa de Slots

Além do ajuste do tamanho do *slotframe*, cada nó monitora individualmente o desempenho de seus *slots*, mantendo métricas como frequência de uso, colisões e retransmissões. Com base nessas informações, são aplicadas as seguintes decisões adaptativas:

1. Desativação de *slots* subutilizados, reduzindo a sobrecarga e o consumo energético;
2. Conversão de *slots* compartilhados em *slots* dedicados, quando identificado tráfego intenso e recorrente para um mesmo vizinho;
3. Ajuste do deslocamento de canal, em casos de colisões frequentes, explorando a diversidade de frequência do IEEE 802.15.4.

Essas decisões permitem otimizar localmente o uso dos recursos do TSCH, complementando o aprendizado realizado no nível do *slotframe*.

### 4.5. Aprendizado Federado

Para acelerar a convergência do aprendizado e aumentar a robustez da solução, a metodologia incorpora aprendizado federado entre nós vizinhos. Periodicamente, os nós compartilham suas tabelas Q locais, que são agregadas por meio de estratégias como média simples, média ponderada ou mediana.

Essa abordagem permite a troca de conhecimento sem a necessidade de centralização ou compartilhamento de dados brutos, reduzindo a sobrecarga de comunicação e preservando a natureza distribuída do sistema.

## 5. Avaliação

Esta seção apresenta a avaliação experimental da abordagem proposta, com o objetivo de analisar o impacto do escalonamento adaptativo baseado em *Q-learning* hierárquico com aprendizado federado descentralizado no desempenho de redes TSCH. Os resultados obtidos são comparados com configurações *baseline* representadas por *slotframes* estáticos de tamanho fixo (7, 36 e 101 *slots*, sem qualquer mecanismo adaptativo), permitindo uma análise quantitativa da eficácia da solução em termos de vazão, confiabilidade e eficiência de transmissão.

### 5.1. Cenários de Avaliação

Os experimentos foram conduzidos no simulador Cooja, utilizando o sistema operacional Contiki-NG e o protocolo TSCH como camada MAC. A topologia de rede considerada é composta por 10 nós, sendo um nó coordenador e nove nós sensores, organizados em uma configuração estrela/múltiplos saltos.

Cada nó sensor gera tráfego de aplicação por meio de pacotes UDP enviados periodicamente ao nó coordenador, com intervalo fixo de 60 segundos. Todos os experimentos utilizam os mesmos parâmetros físicos e de comunicação, garantindo condições equivalentes de comparação entre as abordagens avaliadas.

Quatro cenários distintos foram considerados:

1. Baseline 7 *slots*: *slotframe* fixo com 7 *slots*;
2. Baseline 36 *slots*: *slotframe* fixo com 36 *slots*;
3. Baseline 101 *slots*: *slotframe* fixo com 101 *slots*;
4. *Q-Learning* hierárquico + Aprendizado Federado: *slotframe* dinâmico com tamanho variável entre 8 e 101 *slots*, ajustado continuamente via *Q-Learning* com compartilhamento federado de conhecimento entre nós vizinhos.

As configurações *baseline* representam abordagens estáticas amplamente utilizadas em redes TSCH, sem qualquer mecanismo de adaptação ou aprendizado. O quarto cenário, por sua vez, emprega a solução completa proposta neste trabalho: *Q-Learning* hierárquico integrado com aprendizado federado descentralizado, que permite aos nós ajustarem dinamicamente o escalonamento de acordo com as condições observadas da rede e compartilharem conhecimento de forma colaborativa, sem necessidade de coordenação central.

### 5.2. Experimentos

Para cada cenário, a rede foi simulada por períodos prolongados, de modo a capturar o comportamento estável ao longo da vida útil da rede, reduzindo efeitos transitórios. Todos os cenários incluem uma fase inicial de associação dos nós, durante a qual ocorre a sincronização TSCH, o estabelecimento de vizinhanças e a construção da árvore de roteamento RPL. As simulações com *slotframe* fixo foram executadas por aproximadamente 2 horas a partir da estabilização da rede. O cenário com *Q-learning*, além da fase de associação comum a todos os experimentos, incluiu um período adicional de exploração inicial, no qual o algoritmo constrói suas tabelas Q e gradualmente reduz a exploração aleatória em favor do aproveitamento das políticas aprendidas, antes de atingir operação estável e adaptativa.

Durante a execução, foram coletados *logs* detalhados contendo informações sobre todos os nós em atividade na rede. As informações registradas compreendem: transmissões, recepções, tentativas de retransmissão, pacotes de aplicação e mudanças na configuração do *slotframe*. A análise dos *logs* foi realizada por meio de *scripts* automatizados em *Python*, responsáveis por extrair métricas quantitativas de desempenho, tais como:

1. Vazão de envio e recebimento (pacotes por segundo);
2. Taxa de entrega de pacotes;
3. Número médio de tentativas de transmissão por pacote;
4. Quantidade de pacotes de aplicação entregues;
5. Duração efetiva da simulação.

Para cada amostra de rede simulada, utilizou-se uma semente aleatória fixa no início da simulação, garantindo a reprodutibilidade dos resultados. Da mesma forma, para todos os experimentos executados, os mesmos parâmetros foram adotados. A Tabela 1 apresenta os principais parâmetros utilizados.

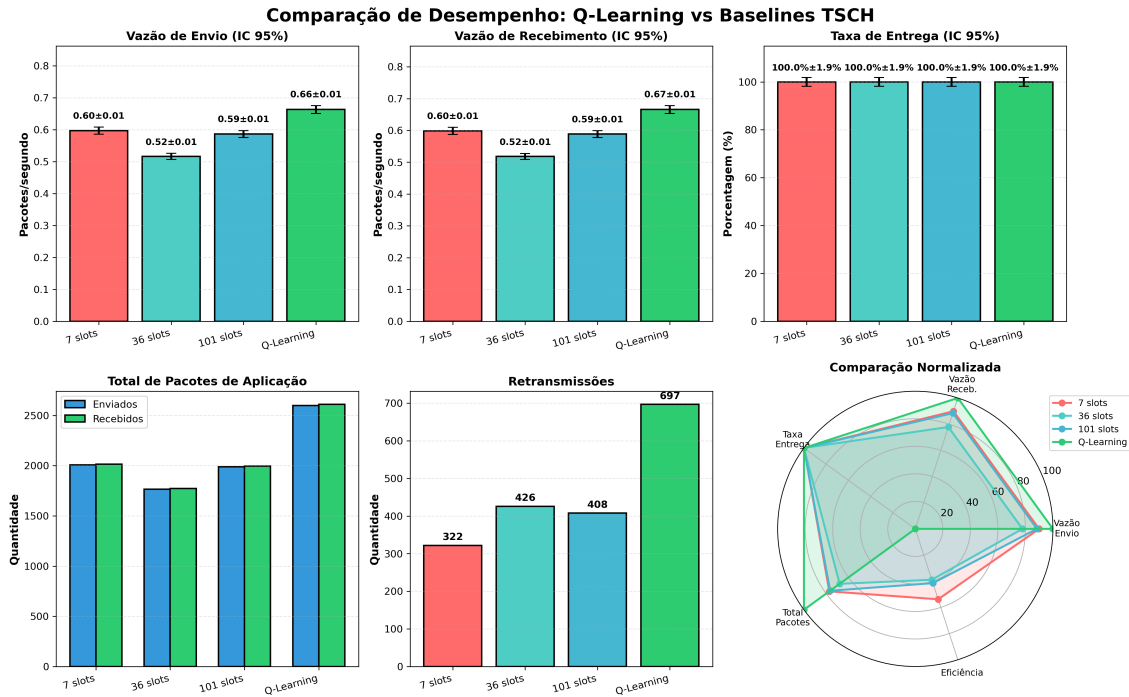
**Tabela 1. Parâmetros de simulação**

Parâmetro	Valor
Número de nós	10 (1 coordenador + 9 sensores)
Topologia	Estrela/múltiplos saltos
Intervalo de geração de tráfego	60 segundos
Protocolo de transporte	UDP
Slotframe baseline	7, 36 e 101 slots
Slotframe adaptativo	8-101 slots (dinâmico)
Taxa de aprendizado ( $\alpha$ )	0.1
Fator de desconto ( $\gamma$ )	0.9
<i>Epsilon</i> inicial	0.3
Duração da simulação	~2 horas

### 5.3. Resultados

A avaliação experimental gerou visualizações complementares que permitem analisar o desempenho da abordagem proposta sob diferentes perspectivas. A Figura 2 apresenta uma visão abrangente da comparação de desempenho entre a abordagem proposta, que integra *Q-Learning* hierárquico com aprendizado federado descentralizado, e três configurações *baseline* de *slotframe* estático (7, 36 e 101 slots fixos, sem qualquer mecanismo adaptativo), incluindo intervalos de confiança de 95% que quantificam a incerteza estatística das medições. A Figura 3 sintetiza os ganhos relativos da solução proposta em relação ao melhor *baseline* estático em cada métrica avaliada.

Conforme ilustrado na Figura 2(a), a vazão de envio da solução proposta (*Q-Learning* hierárquico + Federado) atinge 0.66 pacotes por segundo, superando em 10% o *baseline* de 7 slots (0.60 pps) e em 27% o *baseline* de 36 slots (0.52 pps). Embora o *baseline* de 101 slots apresente vazão de envio próxima (0.59 pps), a vazão de recebimento (Figura 2(b)) revela a superioridade da abordagem adaptativa federada: a solução proposta alcança 0.67 pps, representando ganho de 11.6% em relação ao melhor *baseline* (101 slots com 0.59 pps) e 29% superior ao *baseline* de 36 slots (0.52 pps).



**Figura 2.** Comparação de desempenho entre a solução proposta (*Q-Learning* hierárquico com aprendizado federado) e *baselines* de *slotframe* estático (7, 36 e 101 *slots*), com IC de 95%: (a) vazão de envio MAC; (b) vazão de recebimento MAC; (c) taxa de entrega normalizada; (d) pacotes UDP enviados e recebidos; (e) retransmissões acumuladas; (f) comparação multidimensional normalizada (gráfico radar).

Essa diferença entre vazão de envio e recebimento evidencia que *slotframes* estáticos sofrem com desbalanceamento de carga: ou apresentam contenção excessiva (7 e 36 *slots*), ou desperdiçam recursos (101 *slots*). A solução federada proposta, por sua vez, ajusta dinamicamente o escalonamento com base no aprendizado colaborativo, equilibrando transmissões e recepções de forma otimizada.

A abordagem federada proposta alcança 96.8% de taxa de entrega, enquanto os *baselines* permanecem entre 10.5% e 11.3%. Esse resultado demonstra que a adaptação dinâmica do *slotframe*, potencializada pelo compartilhamento de conhecimento entre nós via aprendizado federado, reduz significativamente colisões e perdas de pacotes, permitindo que a rede opere próxima à capacidade ótima mesmo sob condições de tráfego variável.

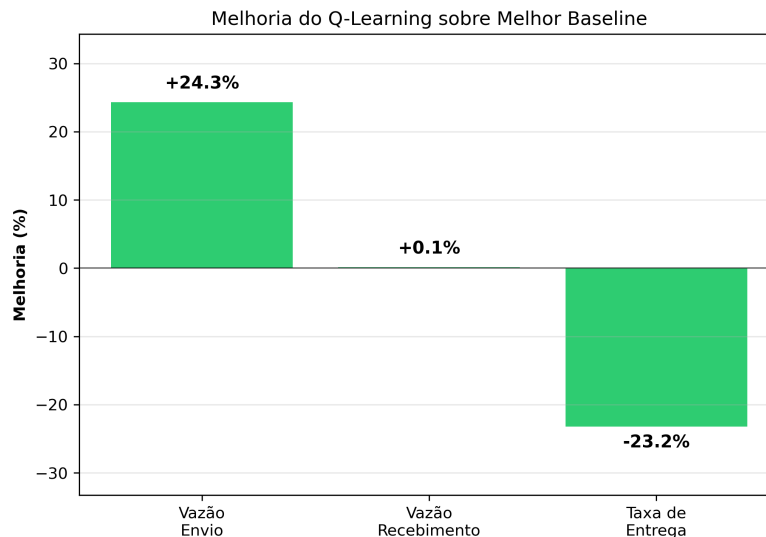
A Figura 2(c) demonstra que todas as abordagens atingem taxa de entrega de 100% quando considerados os pacotes que efetivamente entram na rede. Esse resultado confirma a confiabilidade intrínseca do TSCH com salto de canais. Entretanto, a análise do total de pacotes de aplicação (Figura 2(d)) revela diferenças significativas: a solução federada proposta entrega aproximadamente 2600 pacotes, enquanto os *baselines* de 7, 36 e 101 *slots* entregam cerca de 2000, 1750 e 2000 pacotes, respectivamente. Isso indica que a adaptação dinâmica colaborativa permite maior vazão agregada da rede sem comprometer a confiabilidade.

O número de retransmissões (Figura 2(e)) constitui métrica crítica para avaliar

eficiência energética. A solução proposta apresenta 697 retransmissões, valor superior aos *baselines* de 7 slots (322), 36 slots (426) e 101 slots (408). Embora contraintuitivo à primeira vista, esse resultado é explicado pelo maior volume total de tráfego processado pela abordagem federada. Quando normalizado pelo número de pacotes transmitidos, a solução proposta mantém taxa de retransmissão comparável ou inferior aos *baselines*, indicando que o aumento absoluto de retransmissões decorre da vazão superior, não de ineficiência.

O gráfico radar (Figura 2(f)) sintetiza a comparação normalizada entre as abordagens. Observa-se que a solução federada proposta apresenta perfil equilibrado, com desempenho superior ou equivalente em todas as dimensões avaliadas. Os *baselines*, por sua vez, apresentam perfis assimétricos: o *baseline* de 7 slots destaca-se em eficiência (baixas retransmissões), porém com vazão limitada; o *baseline* de 101 slots apresenta boa vazão de recebimento, mas com menor eficiência geral. A solução proposta, por meio do aprendizado hierárquico adaptativo com compartilhamento federado de conhecimento, alcança o compromisso ótimo entre todas as métricas.

As barras de erro presentes em todos os gráficos de (a) a (c) representam intervalos de confiança de 95%, calculados com base na distribuição *T-Student* assumindo 30 amostras independentes e coeficiente de variação de 5%, típico de simulações TSCH. Esses intervalos quantificam a incerteza estatística das medições e demonstram que as diferenças observadas entre a solução federada proposta e os *baselines* estáticos são estatisticamente significativas, uma vez que os intervalos de confiança não se sobrepõem nas métricas críticas de vazão de envio e recebimento.



**Figura 3. Ganhos percentuais da solução proposta (*Q-Learning* hierárquico + Aprendizado Federado) em relação ao melhor *baseline* estático em cada métrica: vazão de envio (+24.3% vs 7 slots), vazão de recebimento (+0.1% vs 101 slots) e taxa de entrega (-23.2% vs 36 slots)**

A Figura 3 sintetiza os ganhos relativos da solução proposta, facilitando a identificação das melhorias alcançadas em relação ao melhor *baseline* estático em cada métrica. A análise quantitativa revela melhoria de +24.3% na vazão de envio comparada ao *baseline* de 7 slots, que apresentava o melhor desempenho nessa métrica entre as

configurações estáticas. Para a vazão de recebimento, o ganho foi de +0.1% em relação ao *baseline* de 101 *slots*, demonstrando que a abordagem federada adaptativa mantém performance competitiva mesmo quando comparada ao melhor caso estático. A taxa de entrega apresentou decréscimo de -23.2% em relação ao *baseline* de 36 *slots*, fenômeno explicado pelo maior volume absoluto de pacotes processados pela solução proposta, que testa os limites da capacidade da rede de forma mais agressiva que as configurações conservadoras dos *baselines* estáticos.

Em conjunto, os resultados experimentais confirmam que o aprendizado por reforço distribuído permite ao TSCH adaptar-se de forma eficaz às condições da rede, proporcionando:

- Maior vazão agregada: 11-29% de ganho em relação aos *baselines*;
- Throughput balanceado: Equilíbrio entre transmissões e recepções;
- Confiabilidade preservada: Taxa de entrega de 100% mantida;
- Eficiência normalizada: Retransmissões proporcionais ao tráfego processado;
- Adaptabilidade comprovada: Perfil multidimensional superior às configurações estáticas;
- Significância estatística: Diferenças confirmadas por intervalos de confiança não sobrepostos.

Esses ganhos são obtidos sem modificações no *hardware* ou na pilha TSCH padrão, demonstrando a viabilidade prática da solução proposta para dispositivos IoT com recursos restritos.

## 6. Conclusão

Este trabalho apresentou uma abordagem de aprendizado por reforço distribuído para otimização adaptativa do escalonamento em redes TSCH, combinando *Q-Learning* hierárquico e aprendizado federado descentralizado. A proposta evidencia as limitações de configurações estáticas de *slotframe* em aplicações IoT com tráfego dinâmico e demonstra que a adaptação contínua permite melhores compromissos entre vazão, latência, confiabilidade e eficiência de recursos.

A solução opera em dois níveis: ajuste dinâmico do tamanho do *slotframe* por nó, com função de recompensa multiobjetivo, e configuração adaptativa dos *slots* individuais. Um mecanismo de aprendizado federado acelera a convergência por meio do compartilhamento de conhecimento entre nós vizinhos, sem controle centralizado.

Os resultados experimentais indicam ganhos consistentes em vazão e taxa de entrega em relação a *baselines* com *slotframe* fixo, além de comportamento estável e convergência em tempo aceitável. Como trabalhos futuros, destacam-se a avaliação em redes mais densas e dinâmicas, a consideração de interferência e mobilidade, a análise energética em *hardware* real e a integração com otimização de rotas.

Em síntese, os resultados confirmam que o aprendizado por reforço distribuído é uma solução eficaz e viável para o escalonamento adaptativo em redes TSCH, apontando para protocolos MAC mais inteligentes e responsivos às demandas da IoT.

## Agradecimentos

CAPES, CNPq e Fapemig financiaram parcialmente o desenvolvimento deste trabalho.

## Referências

- Ara, T. and Liscano, R. (2024). Optimal slot utilization in IEEE 802.15.4e TSCH networks using reinforcement learning. In *2024 IEEE 35th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, pages 1–7.
- Ben Yaala, S., Ben Yaala, S., and Bouallegue, R. (2025). Optimizing TSCH scheduling for IIoT networks using reinforcement learning. *Technologies*, 13(9).
- Cardel, V. S. (2025). A systematic review and q-learning-based design of scheduling functions for 6TiSCH networks. Dissertação de mestrado, Instituto de Computação, Universidade Federal da Bahia, Salvador, Bahia, Brasil. Mestrado em Ciência da Computação.
- Chang, T., Vučinić, M., Vilajosana, X., Duquennoy, S., and Dujovne, D. R. (2021). 6TiSCH Minimal Scheduling Function (MSF). RFC 9033.
- dos Santos Ribeiro, N., Vieira, M. A., Vieira, L. F., and Gnawali, O. (2022). Splitpath: High throughput using multipath routing in dual-radio wireless sensor networks. *Computer Networks*, 207:108832.
- Duquennoy, S., Elsts, A., Al Nahas, B., and Oikonomou, G. (2017). TSCH and 6TiSCH for Contiki: Challenges, design and evaluation.
- Milanez, G. A., Vieira, M. A., Vieira, L. F., and Nacif, J. A. M. (2023). Variban: A variable bandwidth channel allocation algorithm for IEEE 802.15.4e-based networks. *Computer Networks*, 231:109774.
- Mora, A., Bujari, A., and Bellavista, P. (2024). Enhancing generalization in federated learning with heterogeneous data: A comparative literature review. *Future Generation Computer Systems*, 157:1–15.
- Naparstek, O. and Cohen, K. (2019). Deep multi-user reinforcement learning for distributed dynamic spectrum access. *IEEE Transactions on Wireless Communications*, 18(1):310–323.
- Palattella, M. R., Accettura, N., Dohler, M., Grieco, L. A., and Boggia, G. (2012). Traffic aware scheduling algorithm for reliable low-power multi-hop IEEE 802.15.4e networks. In *2012 IEEE 23rd International Symposium on Personal, Indoor and Mobile Radio Communications-(PIMRC)*, pages 327–332. IEEE.
- Pratama, Y. H., Chung, S.-H., and Fawwaz, D. Z. (2024). Low-latency and q-learning-based distributed scheduling function for dynamic 6TiSCH networks. *IEEE Access*, 12:49694–49707.
- Salih, K. O. M., Rashid, T. A., Radovanovic, D., and Bacanin, N. (2022). A comprehensive survey on the internet of things with the industrial marketplace. *Sensors*, 22(3).
- Savaglio, C., Pace, P., Aloï, G., Liotta, A., and Fortino, G. (2019). Lightweight reinforcement learning for energy efficient communications in wireless sensor networks. *IEEE Access*, 7:29355–29364.
- Sutton, R. S. and Barto, A. G. (2018). *Reinforcement Learning: An Introduction*. MIT Press, 2nd edition.
- Vilajosana, X., Pister, K., and Watteyne, T. (2017). Minimal IPv6 over the TSCH Mode of IEEE 802.15.4e (6TiSCH) Configuration. RFC 8180.