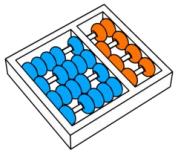


# Uma solução adaptativa baseada em aprendizado por reforço para contenção do tráfego de mensagens de controle em FANETs

Universidade Estadual de Campinas  
Instituto de Computação



# Uma solução adaptativa baseada em aprendizado por reforço para contenção do tráfego de mensagens de controle em FANETs

Aluno:

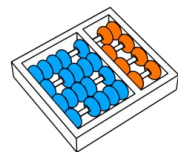
Elcio Pereira de Souza Junior

Orientador:

Leandro Aparecido Villas

Coorientadora:

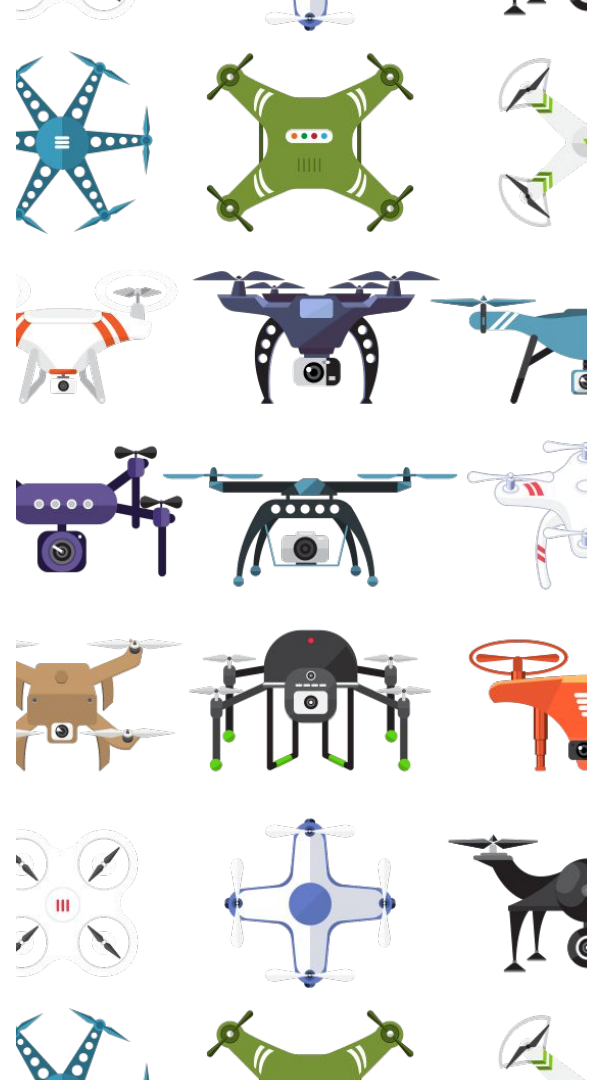
Esther Luna Colombini



# Agenda

— — —

- Motivação
- Objetivos
- Referencial Teórico
- Trabalhos Relacionados
- Modelo Proposto
- Simulação
- Resultados



# Motivação

— — —

- Utilização de VANTs
- Cenário multi VANT – FANETs
- Desafio → Comunicação
- Protocolos e mensagens de controle



# Motivação

— — —

- Utilização de VANTs
- Cenário multi VANT – FANETs
- **Desafio → Comunicação**
- **Protocolos e mensagens de controle**
  - Sobrecarga dos canais de comunicação
  - Aumento do consumo energético



# Objetivo

— — —

Construção de uma solução adaptativa baseada em aprendizado por reforço que permita a otimização do tráfego de mensagens de controle em FANETs.

- Propor um **modelo baseado em aprendizado por reforço para a otimização do envio de mensagens de controle**;
- Construir um **módulo de aprendizado** que seja **adaptável** às diferentes categorias de protocolos de roteamento existentes;
- Construir um **conhecimento global** do estado da rede **de modo a prolongar a efetividade do modelo proposto**.

# Referencial Teórico

## FANETs

— — —

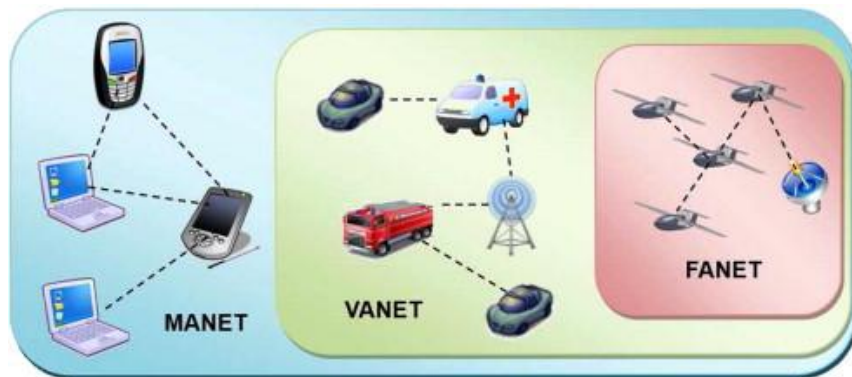


Figura 1: MANET, VANET e FANET.<sup>1</sup>

- **Grau de mobilidade**
- **Padrão de mobilidade**
- **Densidade dos nós**
- **Mudanças topológicas**
- **Modelo de rádio propagação**
- **Consumo energético**
- **Poder computacional**
- **Localização**

<sup>1</sup>Ilker Bekmezci, Ozgur Koray Sahingoz e Samil Temel. “Flying Ad-Hoc Networks (FANETs): A survey”. in Ad Hoc Networks 11.3 (2013), pp. 1254-1270. issn: 1570-8705. doi:<https://doi.org/10.1016/j.adhoc.2012>.

# Referencial Teórico

## Comunicação

— — —

### Características:

- Topologia dinâmica
- Links intermitentes
- Segmentação da rede
- Baixos intervalos de conexão
- Baixa disponibilidade energética

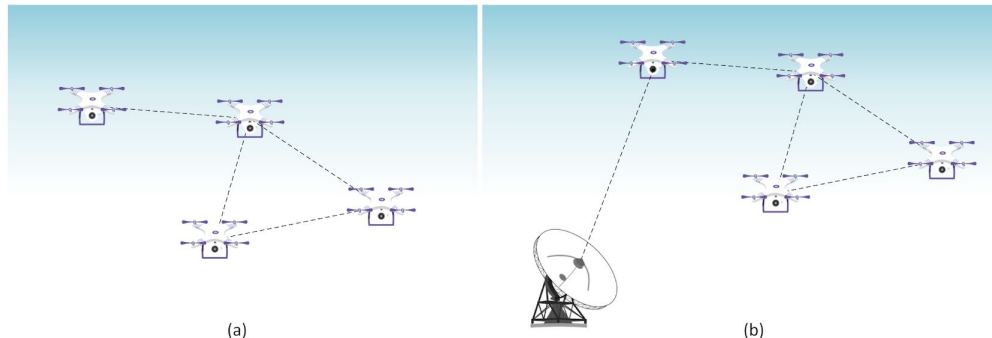


Figura 2: Comunicação U2U (a) e híbrida (b).

### Desafios:

- Mobilidade
- Recursos do sistema
- **Broadcast Storming**



# Referencial Teórico

## Protocolos e Mensagens de Controle

— — —

- Protocolos estáticos
  - Tabelas fixas de roteamento
  - Não adaptável a cenários dinâmicos
- Protocolos proativos
  - Tabelas de roteamento atualizáveis
  - Maior custo computacional (cálculo de rotas ótimas) e consumo energético
- Protocolos reativos
  - Rotas sob demanda
  - Aumento do delay

# Referencial Teórico

## Protocolos e Mensagens de Controle

— — —

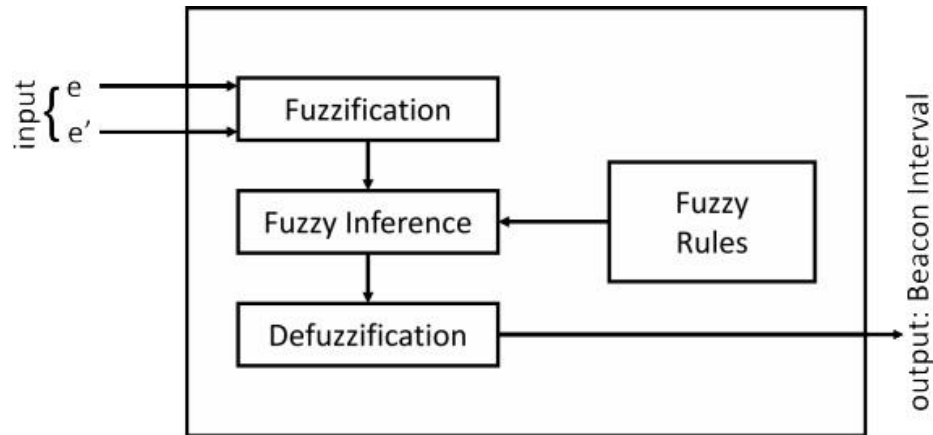
- Protocolos híbridos
  - Combinação modelos proativos e reativos
  - Problemas em maior comunicação entre zonas
- Protocolos geográficos
  - Não é necessário o armazenamento das tabelas de roteamento
  - Envio constante de mensagens de controle
- Protocolos hierárquicos
  - Múltiplos níveis
  - Maior complexidade na gestão das tabelas de roteamento

# Trabalhos Relacionados

## ABPP<sup>3</sup>

— — —

- Mobilidade como principal fator de ajuste
- alta velocidade = baixo intervalo
- baixa velocidade = alto intervalo



$e$ : erro de previsão de posição

$e'$ : taxa de mudança de erro

$b$ : intervalo de envio de mensagens

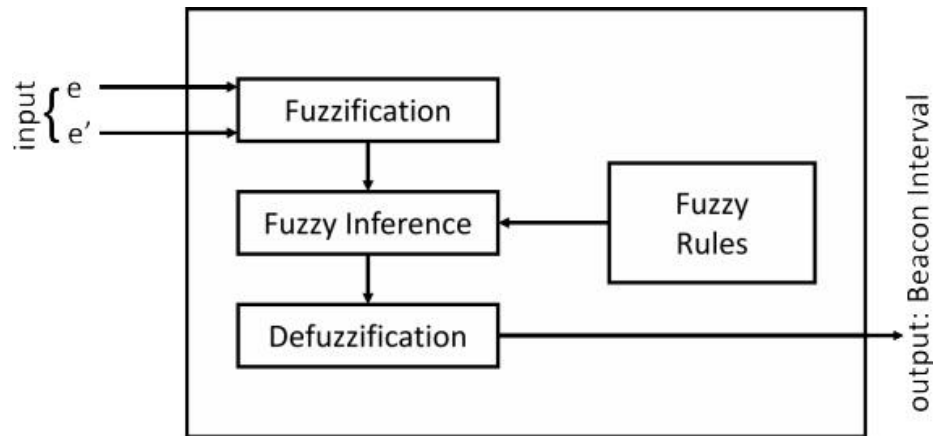
<sup>4</sup>X. Li and J. Huang, "ABPP: An Adaptive Beacon Scheme for Geographic Routing in FANET," 2017 18th International Conference on Parallel and Distributed Computing, Applications and Technologies (PDCAT), Taipei, Taiwan, 2017, pp. 293–299, doi: 10.1109/PDCAT.2017.00055.

# Trabalhos Relacionados

## ABPP<sup>3</sup>

— — —

- + Melhora na sobrecarga do beacon e taxa de entrega de pacotes
- + Aplicável a diferentes protocolos geográficos
- Limitações em altas velocidades
- Modelos de predição degrada em ambientes dinâmicos



$e$ : erro de previsão de posição

$e'$ : taxa de mudança de erro

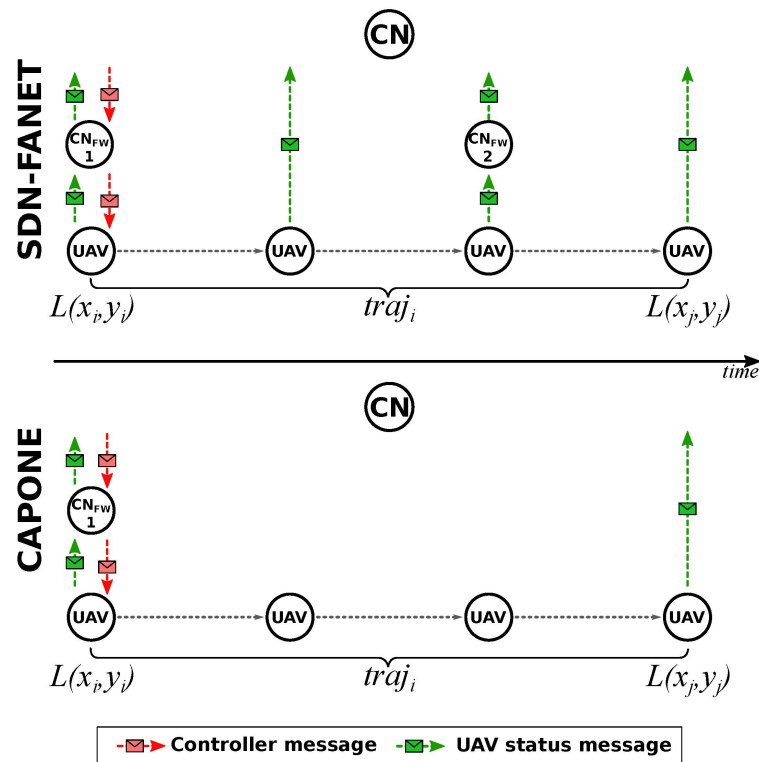
$b$ : intervalo de envio de mensagens

<sup>4</sup>X. Li and J. Huang, "ABPP: An Adaptive Beacon Scheme for Geographic Routing in FANET," 2017 18th International Conference on Parallel and Distributed Computing, Applications and Technologies (PDCAT), Taipei, Taiwan, 2017, pp. 293–299, doi: 10.1109/PDCAT.2017.00055.

# Trabalhos Relacionados

## CAPONE<sup>2</sup>

- Novo paradigma de rede conhecido como SDN-FANET
- Organização hierárquica da rede
- Algoritmo para clusterização e previsão de posicionamento
- Confirmação de recepção



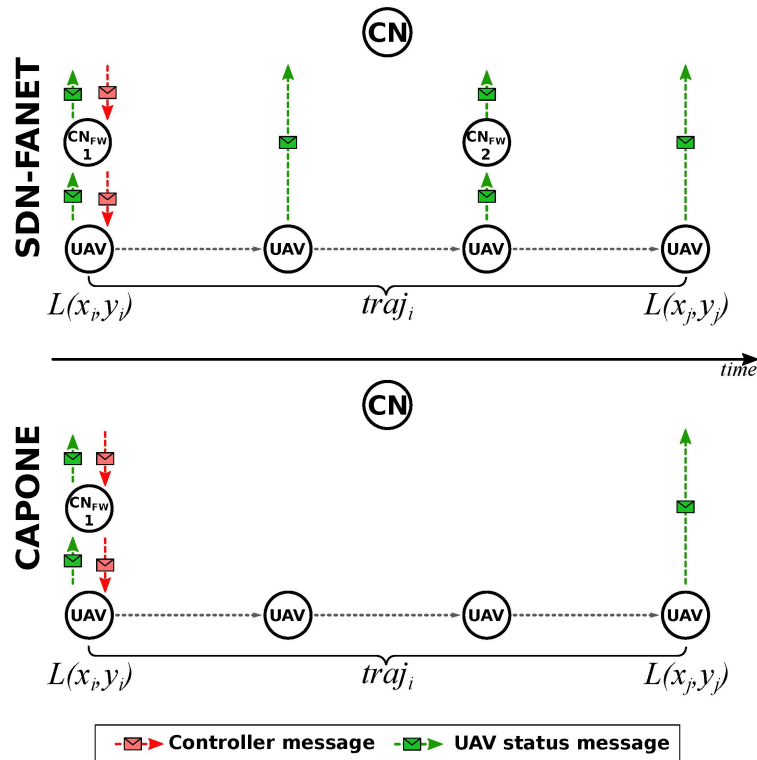
<sup>3</sup>Cumino P., Maciel K., Tavares T., Oliveira H., Rosário D., Cerqueira E., "Cluster-Based Control Plane Messages Management in Software-Defined Flying Ad-Hoc Network". in Sensors. 2020; 20(1):67. <https://doi.org/10.3390/s20010067>

# Trabalhos Relacionados

## CAPONE<sup>2</sup>

— — —

- + Aumento da taxa de entrega de pacotes e redução do consumo energético e overhead
- Cenários densos: suscetível a sobrecarga do controlador



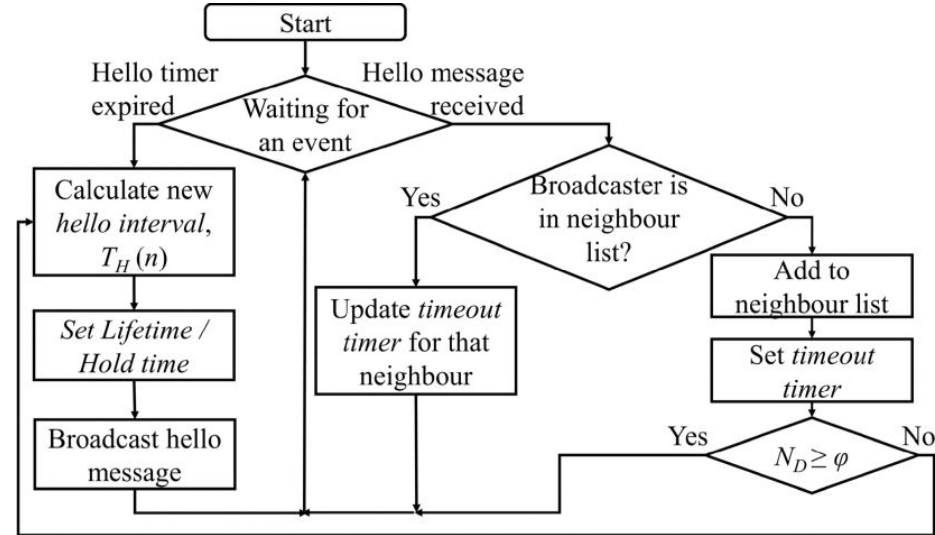
<sup>3</sup>Cumino P., Maciel K., Tavares T., Oliveira H., Rosário D., Cerqueira E., “Cluster-Based Control Plane Messages Management in Software-Defined Flying Ad-Hoc Network”. in Sensors. 2020; 20(1):67. <https://doi.org/10.3390/s20010067>

# Trabalhos Relacionados

## EE-Hello<sup>1</sup>

— — —

- Algoritmo de intervalo de saudação adaptável
- Mecanismo de feedback de mensagem de saudação instantânea



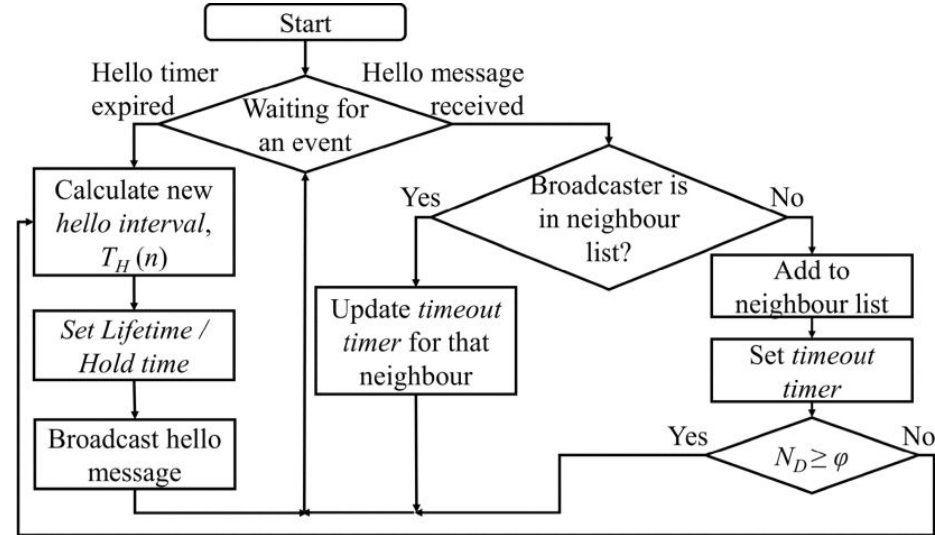
<sup>2</sup>I. Mahmud and Y. Cho, "Adaptive Hello Interval in FANET Routing Protocols for Green UAVs," in IEEE Access, vol. 7, pp. 63004–63015, 2019, doi: 10.1109/ACCESS.2019.2917075.

# Trabalhos Relacionados

## EE-Hello<sup>1</sup>

— — —

- + 25% de redução do consumo energético
- Restrição do espaço aéreo
- Inserção de dados offline



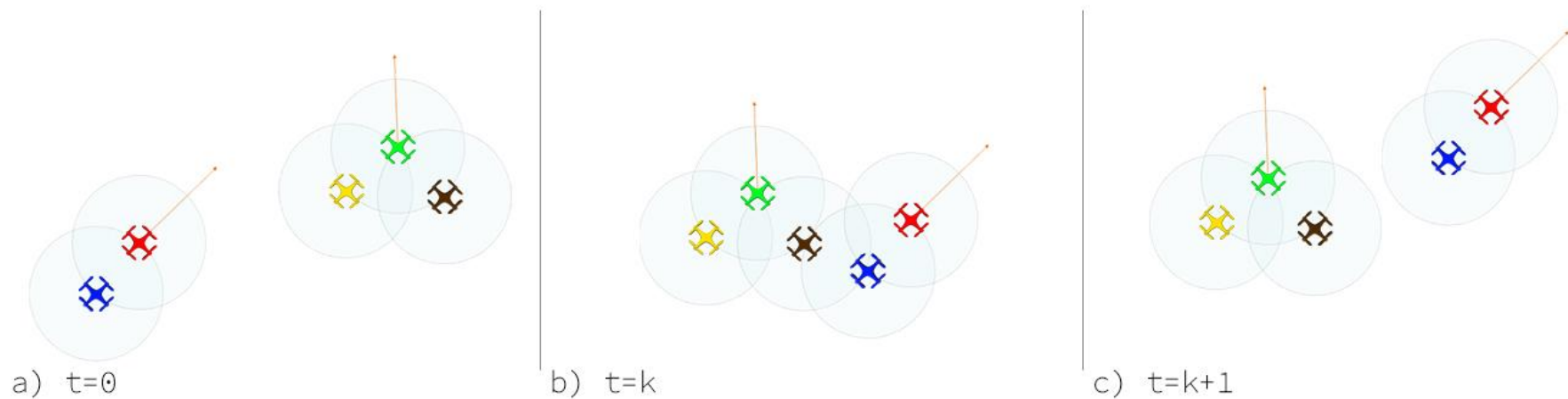
<sup>2</sup>I. Mahmud and Y. Cho, "Adaptive Hello Interval in FANET Routing Protocols for Green UAVs," in IEEE Access, vol. 7, pp. 63004–63015, 2019, doi: 10.1109/ACCESS.2019.2917075.



# Modelo Proposto

## Eventos de Interesse

— — —



**Figura 3:** Evento de interesse, segmentação e conexão total dos nós.

# Modelo Proposto

## Aprendizado por Reforço

Equação  
de Bellman

$$Q(s, a)_{\text{New Q-Value}} = Q(s, a)_{\text{Current Q-Value}} + \underbrace{\alpha}_{\text{Learning rate}} \left[ \underbrace{R(s, a)}_{\text{Reward}} + \underbrace{\gamma \max_{a'} Q'(s', a')}_{\text{Maximum predicted reward, given new state and all possible actions}} - Q(s, a)_{\text{Current Q-Value}} \right]$$

Valor futuro ótimo  
estimado dado o par  
estado - ação

$$Q^* : State \times Action \rightarrow \mathbb{R}$$

- No entanto, não sabemos tudo sobre o mundo, por isso não temos acesso a  $Q^*$
- Uma vez que as redes neurais são aproximadores de função, podemos então criar uma e treiná-la para se assemelhar  $Q^*$

# Modelo Proposto

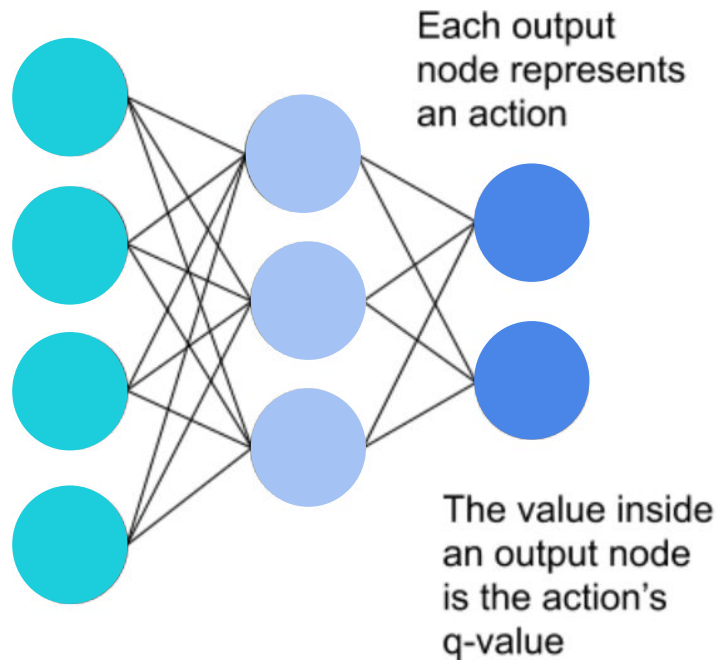
## Deep Q Network

— — —

$$Q^* : State \times Action \rightarrow \mathbb{R}$$

- Algoritmo livre de modelo (que se baseia apenas em experiências)
- A rede neural mapeia os estados de entrada para pares (ação, valor Q)

Input States



# Modelo Proposto

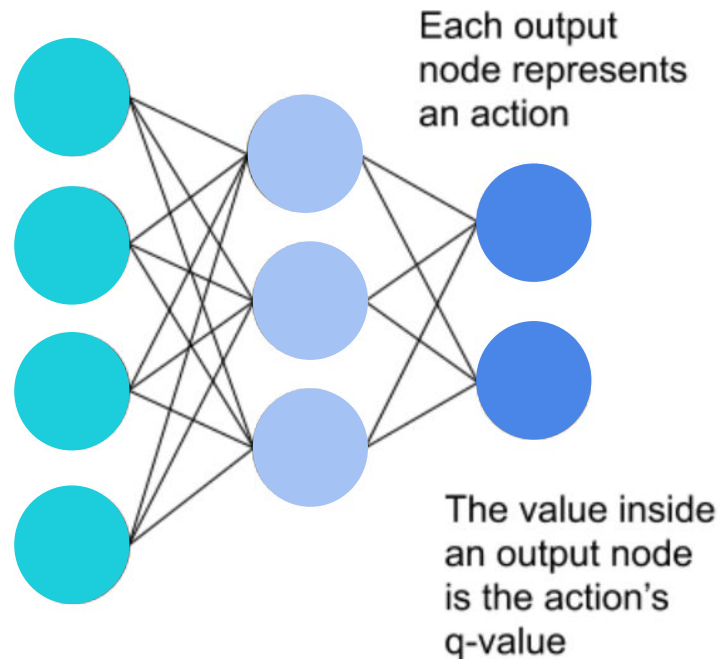
## Deep Q Network

— — —

$$Q^* : State \times Action \rightarrow \mathbb{R}$$

- Epsilon-Greedy Exploration, o agente escolhe uma ação aleatória com probabilidade épsilon e explora a ação mais conhecida com probabilidade 1-épsilon
- Espaço de ações discreto

Input States



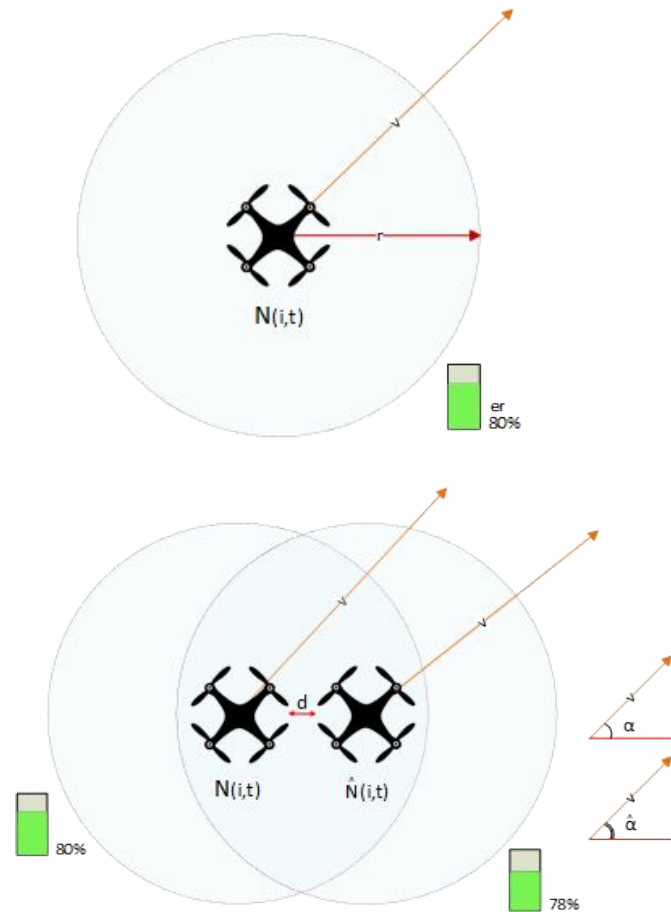
# Modelo Proposto

## Estado ou Entrada

— — —

## Entrada ou Estado:

- $Position_{[t-1,t]}$
- Range
- $Energy_{[t-1,t]}$
- $Vizinhança_{[t-1,t]}$
- $Time_{[t-1,t]}$
- $Tx_{[t-9,t]}$
- $Rx_{[t-9,t]}$
- $Actions_{[t-9,t]}$



# Modelo Proposto

## Reforço

— — —

$$reward = \left[ \underbrace{\left( |Rx_t - Rx_{t-1}|^{hello} - |Tx_t - Tx_{t-1}|^{hello} \right)}_{\text{Variação de mensagens recebidas e enviadas do tipo hello}} \times \underbrace{\left( \sum_{i=1}^t \frac{Rx_i^{hello} + Tx_i^{hello}}{Rx_i^{control} + Tx_i^{control}} \right)}_{\text{Razão entre mensagens do tipo hello e as demais}} - \underbrace{(energy_{t-1} - energy_t) * 10}_{\text{energia residual}} \right]$$

Variação de mensagens  
recebidas e enviadas do  
tipo hello

Razão entre mensagens do  
tipo hello e as demais

energia residual

# Modelo Proposto

## Intervalo de Comunicação

— — —

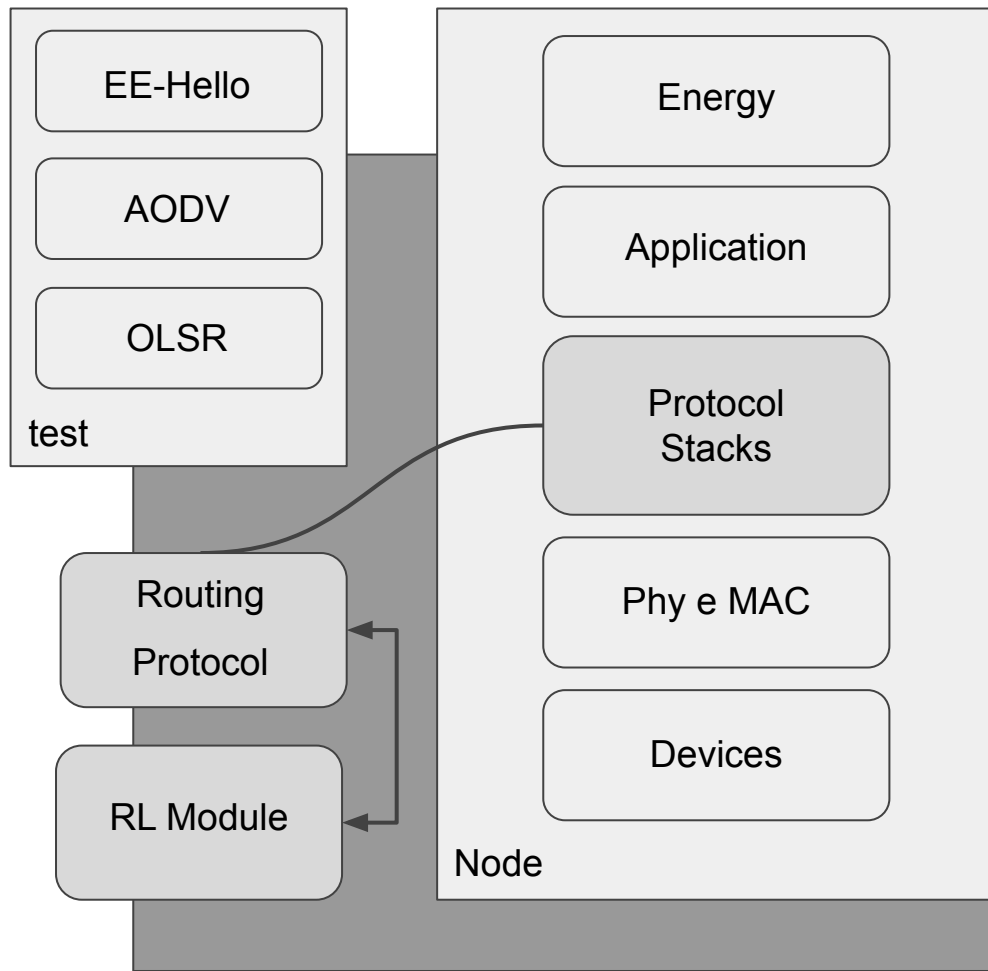
### Saída ou Ações:

- Intervalo de comunicação  
(intervalo específico ao tipo  
de mensagem);
- Mensagens do tipo Hello.

# Modelo Proposto

## Visão Geral

- O agente é treinado em ambiente simulado;
- O treinamento é individual;
- Estima-se que o comportamento individual de múltiplos agentes forneça uma otimização global da rede.





# Simulação

## Cenário

— — —

- **Wifi:** taxa 2 Mb/s em modo ad hoc
- **Aplicação:** 4 pacotes de 64 bytes por segundo
- Os dados do aplicativo são iniciados em um tempo aleatório entre 30 e 31 segundos e continuam até o final da simulação

**NS3:** Simulador Open-Source, baseado em eventos discretos, destinado a pesquisa e uso educacional de sistemas de internet

<b>Área:</b>	1000 x 1000 x 600 m
<b>Número de VANTs:</b>	10
<b>Velocidade:</b>	[5,20] m/s
<b>Transmissão:</b>	150 m
<b>Mobilidade:</b>	Gauss Markov Mobility Model
<b>Número de VANTs:</b>	10
<b>Source/Destination:</b>	3
<b>Protocolo Phy/Mac:</b>	Wifi 802.11b

# Simulação

## Treinamento

— — —

- **Timesteps:** intervalo de tempo específico;
- **Episódio:** um conjunto finito de timesteps;
- Cada timestep está definido em 1s e cada episódio possui de 300 a 500 timesteps (tempo de simulação)

**Stable Baseline:** Conjunto de implementações de aprendizado por reforço destinado a pesquisadores e entusiastas da área.

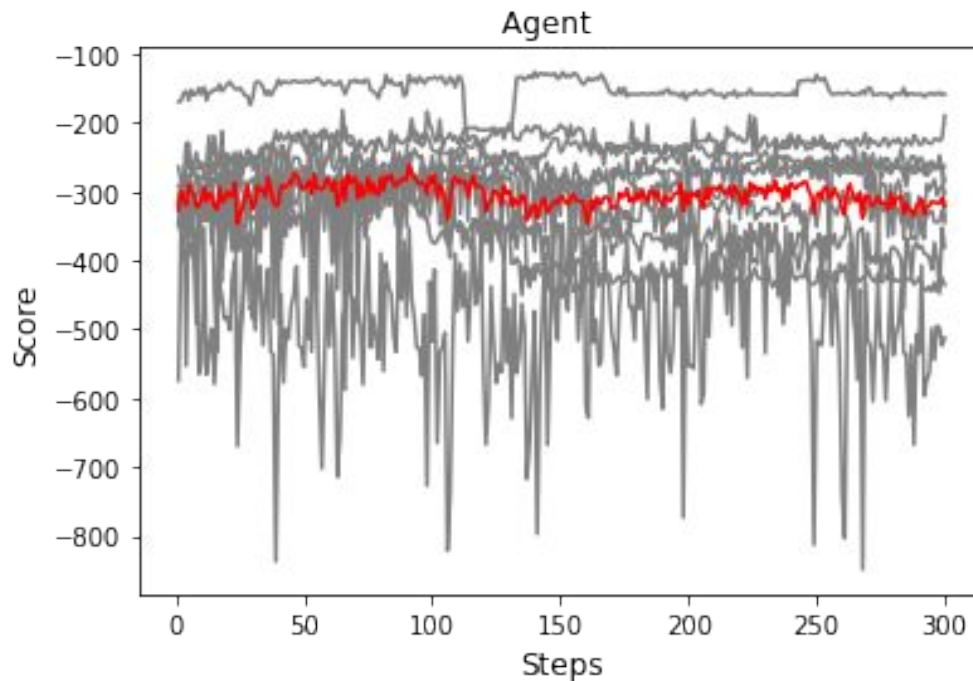
<b>Política:</b>	MlpPolicy
<b>Taxa de aprendizado:</b>	0,001
<b>Batch</b>	64
<b>Frequência de treinamento:</b>	4
<b>Epsilon:</b>	[0.05, 1]
<b>Fração de exploração:</b>	0.1

# Resultados

## Avaliação

— — —

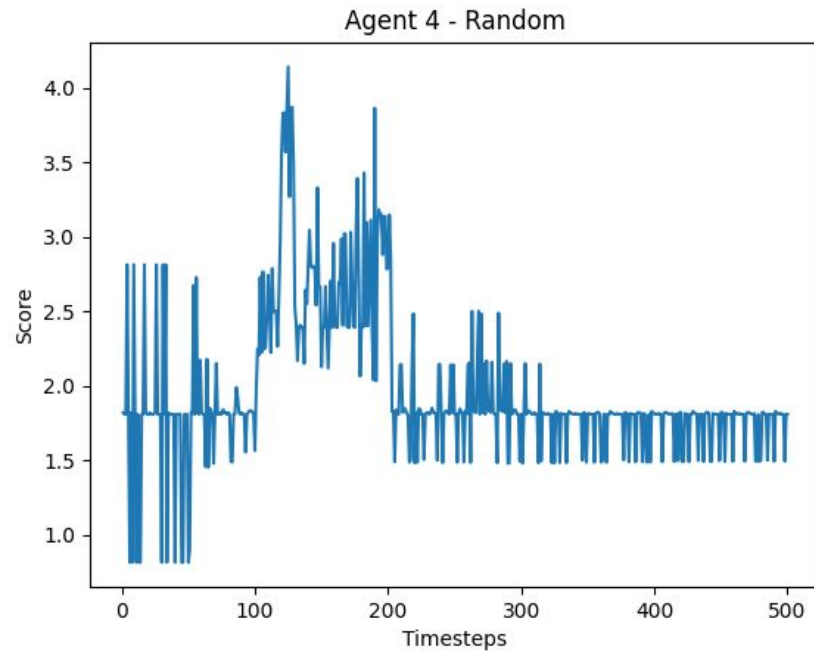
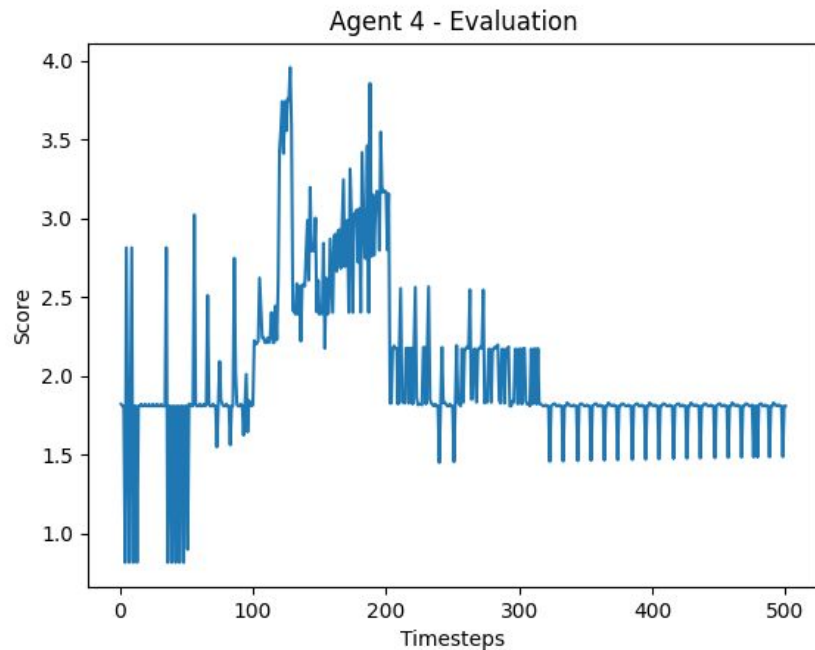
- Função de recompensa empregando o total das mensagens de controle;
- Desempenho médio dos agentes em um ambiente durante 300 timesteps.



# Resultados

## Avaliação

— — —



# Resultados

## Avaliação

— — —

- Mudança para função de recompensa mais específica;
- Revisão do par estado e ação;
- Alteração do modelo de treinamento.

	Random	Evaluate
Total sent packets	165277	165034
Total Received Packets	77593	80474
Total Lost Packets	87684	84560
Packet Loss ratio	53%	51%
Packet delivery ratio	46%	48%
Average Throughput	20.5689Kbps	19.8483Kbps
End to End Delay	1.5858e+12ns	1.36862e+12ns
End to End Jitter delay	1.12497e+11ns	1.1059e+11ns

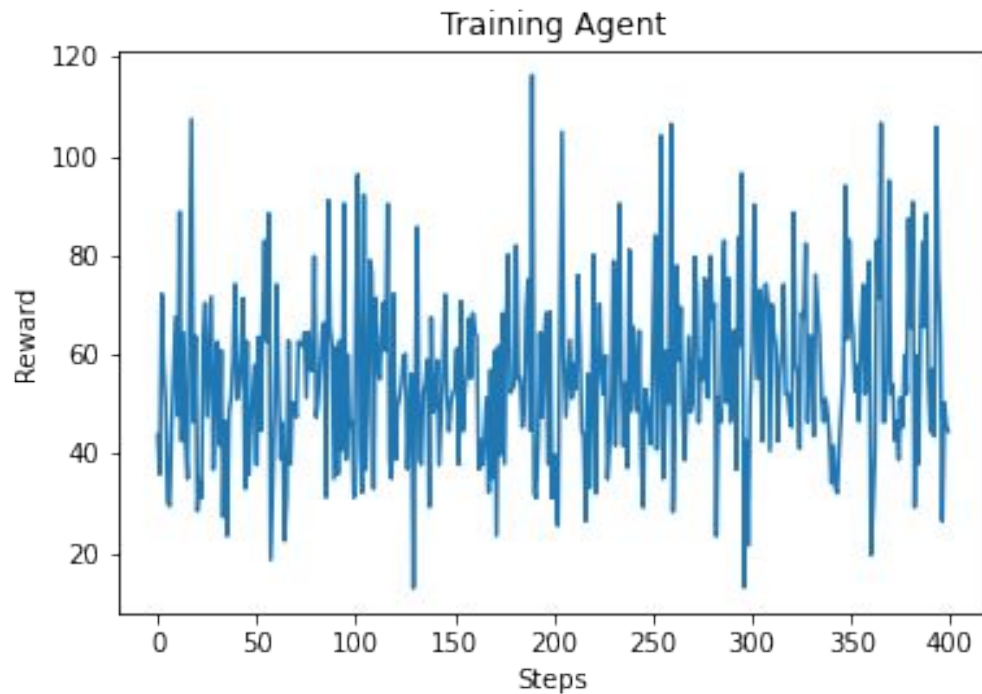
# Resultados

## Avaliação

- Convergência dos agentes;

### Próximos passos:

- Correção de bugs no processo de treinamento;
- Inserção da métrica de energia.



# Agradecimentos

— — —



Prof. Leandro A. Villas  
Prof<sup>a</sup>. Esther L. Colombini



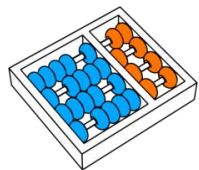
Computer Networks  
Laboratory



Conselho Nacional de  
Desenvolvimento Científico  
e Tecnológico



Laboratório de Robótica  
e Sistemas Cognitivos



Instituto de Computação e a  
Unicamp

# Obrigado!

