



UNIVERSIDAD NACIONAL DE CHIMBORAZO
FACULTAD DE INGENIERÍA
CARRERA DE TELECOMUNICACIONES

Análisis de un protocolo de control de acceso al medio basado en *Q-learning* para incrementar el *throughput* en una red inalámbrica de sensores asistidos por superficies reconfigurables inteligentes

Trabajo de Titulación para optar al título de
INGENIERO EN TELECOMUNICACIONES

AUTOR:

Michael Luis Allauca Ramirez

TUTOR:

PhD. Luis Patricio Tello Oquendo

Riobamba, Ecuador. 2026

DECLARATORIA DE AUTORÍA

Yo, Michael Luis Allauca Ramirez, con cédula de ciudadanía 060474858-2, autor del trabajo de investigación titulado: **Análisis de un protocolo de control de acceso al medio basado en Q-Learning para incrementar el throughput en una red inalámbrica de sensores asistidos por superficies reconfigurables inteligentes**, certifico que la producción, ideas, opiniones, criterios, contenidos y conclusiones expuestas son de mi exclusiva responsabilidad.

Asimismo, cedo a la Universidad Nacional de Chimborazo, en forma no exclusiva, los derechos para su uso, comunicación pública, distribución, divulgación y/o reproducción total o parcial, por medio físico o digital; en esta cesión se entiende que el cesionario no podrá obtener beneficios económicos. La posible reclamación de terceros respecto de los derechos de autor (a) de la obra referida, será de mi entera responsabilidad; librando a la Universidad Nacional de Chimborazo de posibles obligaciones.

En Riobamba, a los 9 días del mes de junio de 2026.



Michael Luis Allauca Ramirez

C.I: 060474858-2

DICTAMEN FAVORABLE DEL TUTOR

Quien suscribe, **PhD. Luis Patricio Tello Oquendo** catedrático adscrito a la Facultad de Ingeniería, por medio del presente documento certifico haber asesorado y revisado el desarrollo del trabajo de investigación titulado: **Análisis de un protocolo de control de acceso al medio basado en Q-Learning para incrementar el throughput en una red inalámbrica de sensores asistidos por superficies reconfigurables inteligentes**, bajo la autoría de Michael Luis Allauca Ramirez; por lo que se autoriza ejecutar los trámites legales para su sustentación.

Es todo cuanto informar en honor a la verdad; en Riobamba, a los 6 días del mes de abril de 2026

A handwritten signature in blue ink, reading "Luis P. Tello O", is written over a horizontal line.

PhD. Luis Patricio Tello Oquendo

TUTOR

CERTIFICADO DE LOS MIEMBROS DEL TRIBUNAL

Quienes suscribimos, catedráticos designados Miembros del Tribunal de Grado para la evaluación del trabajo de investigación **Análisis de un protocolo de control de acceso al medio basado en Q-Learning para incrementar el throughput en una red inalámbrica de sensores asistidos por superficies reconfigurables inteligentes** presentado por **Michael Luis Allauca Ramirez**, con cédula de identidad número **0604748582**, bajo la tutoría de **PhD. Luis Patricio Tello Oquendo**; certificamos que recomendamos la **APROBACIÓN** de este con fines de titulación. Previamente se ha evaluado el trabajo de investigación y escuchada la sustentación por parte de su autor; no teniendo más nada que observar.

De conformidad a la normativa aplicable firmamos, en Riobamba a los 9 días del mes de junio de 2026.

Deysi Vilma Inca Balseca, Mgs.
PRESIDENTE DEL TRIBUNAL DE GRADO



Dr. Klever Hernán Torres Rodríguez.
MIEMBRO DEL TRIBUNAL DE GRADO



Leonardo Fabian Rentería Bustamante, PhD.
MIEMBRO DEL TRIBUNAL DE GRADO



CERTIFICACIÓN DEL ANTI-PLAGIO

Que, **ALLAUCA RAMIREZ MICHAELL LUIS** con CC: **060474858-2**, estudiante de la Carrera de **TELECOMUNICACIONES**, Facultad de **INGENIERÍA**; ha trabajado bajo mi tutoría el trabajo de investigación titulado " **ANÁLISIS DE UN PROTOCOLO DE CONTROL DE ACCESO AL MEDIO BASADO EN Q-LEARNING PARA INCREMENTAR EL THROUGHPUT EN UNA RED INALÁMBRICA DE SENSORES ASISTIDOS POR SUPERFICIES RECONFIGURABLES INTELIGENTES**", cumple con el <1% de similitud y el 6% de Inteligencia Artificial, de acuerdo al reporte del sistema Anti plagio **COMPILATIO**, porcentaje aceptado de acuerdo a la reglamentación institucional, por consiguiente autorizo continuar con el proceso.

Riobamba, 26 de mayo de 2026



PhD. Luis Patricio Tello Oquendo
TUTOR

DEDICATORIA

Doy gracias a Dios por haberme acompañado en cada paso de este camino, por darme la fortaleza para superar los obstáculos, la sabiduría para aprender de cada experiencia.

A mis padres, Luis Allauca y Norma Ramírez, quienes han sido el pilar fundamental de mi vida. Gracias por su amor incondicional, por cada sacrificio realizado y por enseñarme que el esfuerzo y la perseverancia son la clave para alcanzar los sueños. Este logro también les pertenece a ustedes, porque han sido mi guía, mi apoyo y la fuerza que me impulsó a seguir adelante en los momentos más difíciles.

A mis abuelos, Miguel Ramírez y Ester Iguasña, por sus consejos, valores y el cariño que me han brindado durante esta etapa. Su ejemplo de vida ha sido una inspiración constante para seguir adelante y enfrentar cada desafío con humildad y determinación.

A mis hermanas, Joselyn, Nayeli, Belén y Gloria, por su apoyo, comprensión y por acompañarme en cada etapa de este camino. Gracias por creer siempre en mí, por sus palabras de aliento y por compartir conmigo cada alegría y cada reto que se presentó durante mi formación.

A mi querida sobrina, Paulet, por regalarme momentos de alegría con su inocencia, cariño y ocurrencias, convirtiéndose en una luz que iluminó muchos de mis días y recordándome la importancia de disfrutar cada instante de la vida.

A todas las personas que me acompañaron durante este recorrido, gracias por su apoyo, confianza y cariño. Este logro también lleva una parte de cada uno de ustedes.

Michaell Luis Allauca Ramirez

AGRADECIMIENTO

Primeramente, agradezco a Dios por darme la vida, la salud, la fortaleza y la sabiduría necesarias para superar cada desafío y permitirme culminar con éxito esta importante etapa de mi vida. Gracias por acompañarme en cada momento y guiar mis pasos hacia el cumplimiento de esta meta.

A mis padres, por su amor incondicional, esfuerzo, sacrificio y apoyo constante. Gracias por creer en mí, por sus consejos y por brindarme siempre la motivación necesaria para seguir adelante y no rendirme ante las dificultades.

A mis hermanas, por su cariño, comprensión y apoyo durante todo este proceso. Gracias por acompañarme en cada etapa de este camino y por celebrar conmigo cada logro alcanzado.

A mis amigos, quienes durante esta etapa universitaria me brindaron su amistad, apoyo y compañía. Gracias por las experiencias compartidas, las risas, los aprendizajes y los momentos que hicieron de este recorrido una etapa inolvidable llena de nuevas vivencias y alegrías. De manera especial, agradezco a Katy por su amor, compañía, comprensión y apoyo incondicional. Gracias por estar presente en los momentos más importantes de este camino, por tus palabras de aliento y por motivarme a seguir adelante cuando más lo necesitaba.

A mi tutor de tesis, Doctor Luis Tello, por su guía, paciencia, dedicación y valiosos conocimientos brindados durante el desarrollo de esta investigación. Su orientación fue fundamental para la culminación de este trabajo.

A todas las personas, familiares y seres queridos que, de una u otra manera, me brindaron su apoyo, confianza y motivación durante este proceso. Gracias por acompañarme y ser parte de este importante logro en mi vida.

Michaell Luis Allauca Ramirez

ÍNDICE GENERAL

DECLARATORIA DE AUTORÍA	
DICTAMEN FAVORABLE DEL TUTOR	
CERTIFICADO DE LOS MIEMBROS DEL TRIBUNAL	
CERTIFICACIÓN DEL ANTI-PLAGIO	
DEDICATORIA	
AGRADECIMIENTO	
ÍNDICE GENERAL	
RESUMEN	
ABSTRACT	
CAPÍTULO I.....	17
INTRODUCCIÓN.....	17
1.1 Antecedentes.....	17
1.2 Planteamiento del problema	18
1.3 Justificación.....	19
1.4 Objetivos.....	19
1.4.1 Objetivo general	19
1.4.2 Objetivos específicos.....	19
CAPITULO II.....	21
MARCO TEÓRICO	21
2.1. Estado del Arte	21
2.2. Fundamentación teórica.....	22
2.2.1. Redes Inalámbricas de Sensores (WSN)	22
2.2.1.1. Definición	22
2.2.1.2. Arquitectura y topologías de red	22
2.2.1.3. Componentes de un nodo sensor	24
2.2.1.4. Aplicaciones	24
2.2.1.5. Caso	25
2.2.1.6. Desafíos en las redes de sensores	25
2.2.2. Protocolos de control de acceso al medio.....	26
2.2.2.1. Concepto.....	26
2.2.2.2. Clasificación de protocolos MAC	26
2.2.2.3. Limitaciones de los protocolos tradicionales de control de acceso al medio	29
2.2.2.4. Importancia de un protocolo eficiente en redes de sensores inalámbricos.....	30

2.2.2.5.	Problemática del acceso al medio.....	30
2.2.3.	Aprendizaje por refuerzo	31
2.2.3.1.	Fundamentos.....	31
2.2.4.	Q-Learning.....	31
2.2.4.1.	Concepto de Q-Learning	31
2.2.4.2.	Formulación matemática del Q-Learning.....	32
2.2.4.3.	Convergencia del Q-Learning	33
2.2.4.4.	Aplicaciones de Q-Learning en redes de comunicación	33
2.2.4.5.	Algoritmos de Q-Learning en redes inalámbricas.....	34
2.2.4.6.	Ventajas de Q-Learning para la optimización de recursos en redes de sensores	35
2.2.5.	Superficies Reconfigurables Inteligentes (RIS)	35
2.2.5.1.	Definición y funcionamiento de RIS	35
2.2.5.2.	Ventajas de RIS en redes inalámbricas.....	36
2.2.5.3.	Aplicaciones de RIS en redes 5G y 6G	36
2.2.5.4.	Integración de RIS en redes de sensores inalámbricos.....	36
2.2.5.5.	Modelo del canal asistido por RIS.....	38
2.2.6.	Protocolos MAC basados en aprendizaje por refuerzo	38
2.2.6.1.	Justificación del uso de RL en protocolos MAC.....	38
2.2.6.2.	Modelado del problema MAC como MDP	39
2.2.7.	Optimización del throughput en redes inalámbricas	39
2.2.7.1.	Definición de throughput.....	39
2.2.7.2.	Factores que afectan el throughput en redes de sensores	40
2.2.7.3.	Métodos para mejorar el throughput en redes inalámbricas.....	41
2.2.7.4.	Uso de técnicas de optimización, como Q-Learning, para aumentar el throughput	41
2.2.8.	Python como entorno de simulación	42
2.2.8.1.	Definición	42
2.2.8.2.	Características.....	42
CAPITULO III		43
METODOLÓGICO		43
3.1.	Tipo de investigación	43
3.1.1.	Investigación cuantitativa.....	43
3.1.2.	Investigación exploratoria	43
3.1.3.	Investigación comparativa.....	43

3.2. Diseño de investigación.....	43
3.3. Métodos, Técnicas e Instrumentos de Investigación.....	43
3.3.1. Método analítico.....	43
3.3.2. Técnicas de recolección de datos.....	44
3.3.3. Instrumentos de Recolección de Datos.....	44
3.4. Población.....	44
3.5. Operacionalización de variables.....	45
3.6. Procedimiento y Análisis.....	46
3.6.1. Estructura y funcionamiento del algoritmo de simulación.....	48
CAPÍTULO IV.....	54
RESULTADOS Y DISCUSIÓN.....	54
4.1. Resultado del diagnóstico.....	54
4.1.1. Análisis de throughput bajo carga de tráfico baja (0.5).....	54
4.1.2. Eficiencia Energética bajo carga de tráfico baja (0.5).....	55
4.1.3. Convergencia del Algoritmo Q-Learning bajo carga baja.....	56
4.1.4. Throughput bajo carga de tráfico media (50).....	56
4.1.5. Eficiencia energética bajo carga de tráfico media (50).....	57
4.1.6. Convergencia del algoritmo bajo carga media.....	58
4.1.7. Throughput con carga de tráfico alta (100).....	59
4.1.8. Eficiencia Energética bajo carga de tráfico alta (100).....	60
4.1.9. Convergencia del Algoritmo con alta densidad de sensores.....	60
4.2. Resultados estadísticos.....	61
4.2.1. Estadísticos Descriptivos por Protocolo.....	61
4.2.2. Análisis por Número de Sensores.....	62
4.2.3. Verificación del Supuesto de Normalidad.....	63
4.2.4. Evidencia Gráfica de No Normalidad.....	64
4.2.5. Identificación del Marco Estadístico: Datos No Paramétricos.....	66
4.2.6. Pruebas de Hipótesis No Paramétricas.....	66
4.2.7. Efecto del Número de Sensores: Prueba de Kruskal-Wallis.....	69
4.2.8. Tamaño del Efecto: Eta Cuadrado (η^2).....	72
4.2.9. Discusión Integrada de Resultados.....	72
CAPÍTULO V.....	74
5. CONCLUSIONES Y RECOMENDACIONES.....	74
5.1. Conclusiones.....	74

5.2. Recomendaciones	75
BIBLIOGRAFÍA	76
ANEXOS	82

ÍNDICE DE TABLAS

Tabla 1. Herramientas e instrumentos para la Recolección de Datos.	44
Tabla 2. Operacionalización de variables.....	45
Tabla 3. Parámetros del modelo de simulación.....	48
Tabla 4. Resultados de throughput y eficiencia energética bajo carga de tráfico baja.....	54
Tabla 5. Resultados de throughput y eficiencia energética bajo carga de tráfico media.....	57
Tabla 6. Resultados de throughput y eficiencia energética bajo carga de tráfico alta	59
Tabla 7. Estadísticos descriptivos del throughput y eficiencia energética por protocolo....	61
Tabla 8. Estadísticos descriptivos del throughput y eficiencia energética por protocolo....	62
Tabla 9. Medias de throughput y eficiencia energética por nivel de sensores	63
Tabla 10. Resultados de las pruebas de normalidad de Kolmogorov-Smirnov y Shapiro-Wilk para throughput y eficiencia energética por protocolo (n = 225).	64
Tabla 11. Resultados de la prueba U de Mann-Whitney para la comparación de throughput y eficiencia energética entre protocolos Q-Learning y BEB ($n_1 = n_2 = 225$, $\alpha = .05$).	67
Tabla 12. Resultados de la prueba de Kruskal-Wallis para el efecto del número de sensores y la carga de tráfico sobre el throughput y la eficiencia energética ($\alpha = .05$).	69

ÍNDICE DE FIGURAS

Figura 1. Topologías en WSN [14].	23
Figura 2. Componentes de WSN [15].	24
Figura 3. Fases del proyecto de investigación.	46
Figura 4. Parámetros generales.	48
Figura 5. Modelo de canal con RIS.	49
Figura 6. Modelo de energía.	49
Figura 7. Parámetros del Q-Learning.	50
Figura 8. Configuración del entrenamiento.	50
Figura 9. Decisión ϵ -greedy.	51
Figura 10. Actualización del Q-learning	51
Figura 11. Estrategia aprendida.	51
Figura 12. Mido energía y bits exitoso.	52
Figura 13. Calculo throughput y eficiencia.	52
Figura 14. Protocolo Binary Exponential Backoff (BEB).	52
Figura 15. Visualización de resultados de la simulación	53
Figura 16. Throughput promedio vs número de sensores (carga 0.5, 60 sensores).	54
Figura 17. Evolución del throughput durante el entrenamiento (carga 0.5).	55
Figura 18. Throughput promedio vs número de sensores (carga 0.5)	56
Figura 19. Throughput promedio vs número de sensores (carga 50, 60 sensores).	57
Figura 20. Eficiencia energética vs número de sensores (carga 50).	58
Figura 21. Evolución del throughput durante entrenamiento (carga 50).	58
Figura 22. Throughput promedio vs número de sensores (carga 100, 60 sensores).	59
Figura 23. Eficiencia energética vs número de sensores (carga 100).	60
Figura 24. Evolución del throughput durante entrenamiento (100 sensores).	61
Figura 25. Histogramas de throughput por protocolo (Q-Learning y BEB).	64
Figura 26. Histogramas de eficiencia energética por protocolo (Q-Learning y BEB).	65
Figura 27. Gráficos Q-Q normales de throughput por protocolo (Q-Learning y BEB).	65
Figura 28. Gráficos Q-Q normales de eficiencia energética por protocolo (Q-Learning y BEB).	66
Figura 29. Diagrama de caja de throughput por protocolo.	68
Figura 30. Diagrama de caja de eficiencia energética por protocolo	68

Figura 31. Diagrama de caja de throughput por número de sensores.	70
Figura 32. Diagrama de caja de eficiencia energética por número de sensores.	70
Figura 33. Diagrama de caja de throughput por carga de tráfico.	71
Figura 34. Diagrama de caja de eficiencia energética por carga de tráfico.	71

RESUMEN

Las redes inalámbricas de sensores (WSN) enfrentan serias limitaciones de throughput y eficiencia energética cuando emplean protocolos tradicionales de control de acceso al medio (MAC), como el Back-off Exponencial Binario (BEB), especialmente en escenarios de alta densidad y tráfico elevado. Paralelamente, las Superficies Reconfigurables Inteligentes (RIS) han emergido como tecnología clave para mejorar la cobertura y calidad de señal en redes de próxima generación. Sin embargo, la sinergia entre RIS y protocolos MAC inteligentes permanece poco explorada. El presente trabajo analiza un protocolo MAC basado en Q-Learning (QL) aplicado a una WSN asistida por RIS, con el objetivo de incrementar el throughput y mejorar la eficiencia energética. La metodología empleó un diseño de simulación computacional, evaluando 15 escenarios que combinaron variaciones en la densidad de sensores (40 a 60 nodos) y tres niveles de carga de tráfico (0.5, 50 y 100 paquetes por unidad de tiempo). Los resultados permitieron validar el comportamiento del sistema, demostrando que el protocolo MAC basado en Q-Learning supera significativamente al esquema tradicional BEB en throughput y eficiencia energética, mientras que el incremento en la densidad de sensores constituye el factor determinante del rendimiento, en contraste con la carga de tráfico, cuya variación no genera cambios sustanciales. En conjunto, el modelo desarrollado constituye una base sólida para futuras investigaciones orientadas a la integración de algoritmos de aprendizaje profundo y arquitecturas multiagente en redes inalámbricas de sensores asistidas por RIS, con miras a su posible despliegue en escenarios reales de redes de próxima generación 6G e IoT masivo.

Palabras clave: Q-Learning, control de acceso al medio (MAC), redes inalámbricas de sensores (WSN), superficies reconfigurables inteligentes (RIS), throughput, eficiencia energética, aprendizaje por refuerzo, Back-off Exponencial Binario (BEB).

ABSTRACT

Wireless Sensor Networks (WSN) face significant throughput and energy-efficiency limitations when using traditional Medium Access Control (MAC) protocols, such as Binary Exponential Back-off (BEB), especially in high-density, high-traffic scenarios. In parallel, Reconfigurable Intelligent Surfaces (RIS) have emerged as a key technology for improving coverage and signal quality in next-generation networks. However, the synergy between RIS and intelligent MAC protocols remains largely unexplored. This work analyzes a Q-Learning (QL)-based MAC protocol for a RIS-assisted WSN, aiming to increase throughput and improve energy efficiency. The methodology employed a computational simulation design, evaluating 15 scenarios combining variations in sensor density (40 to 60 nodes) and three traffic load levels (0.5, 50, and 100 packets per time unit). The results validated the system's behavior, demonstrating that the Q-Learning-based MAC protocol significantly outperforms the traditional BEB scheme in both throughput and energy efficiency. The increase in sensor density is the determining performance factor, whereas traffic load variation does not produce substantial changes. Overall, the developed model provides a solid foundation for future research aimed at integrating deep learning algorithms and multi-agent architectures into RIS-assisted wireless sensor networks, with a view to potential deployment in real-world next-generation 6G and massive IoT network scenarios.

Keywords: Q-Learning; Medium Access Control (MAC); Wireless Sensor Networks (WSN); Reconfigurable Intelligent Surfaces (RIS); Throughput; Energy Efficiency; Reinforcement Learning; Binary Exponential Backoff (BEB).



Reviewed by:
Mgs. Hugo Romero
ENGLISH PROFESSOR
C.C. 0603156258

CAPÍTULO I

INTRODUCCIÓN

En los últimos años, las superficies reconfigurables inteligentes (RIS) han emergido como una tecnología innovadora en el ámbito de las redes inalámbricas, cabe mencionar que las RIS consisten en materiales electromagnéticos que incluyen numerosos elementos pasivos, capaces de modificar dinámicamente las propiedades de propagación de las señales inalámbricas, como su dirección, amplitud y fase. Esta capacidad de reconfiguración permite optimizar la cobertura, aumentar la tasa de datos y mejorar significativamente la eficiencia energética de las redes, convirtiendo a las RIS en una solución prometedora para los retos actuales de las comunicaciones inalámbricas [1], [2].

Sin embargo, la incorporación de RIS en redes inalámbricas plantea desafíos relacionados con la coordinación del acceso de múltiples usuarios al medio compartido. La interferencia generada por accesos simultáneos puede reducir la eficiencia energética y el throughput de la red. Para abordar esta problemática, se han propuesto estrategias basadas en inteligencia artificial, como los protocolos de control de acceso al medio (MAC) asistidos por aprendizaje por refuerzo, los cuales permiten la asignación dinámica y eficiente de recursos en sistemas complejos [3], [4].

Este trabajo se centra en analizar un protocolo MAC basado en Q-Learning diseñado para redes inalámbricas de sensores asistidas por RIS. El objetivo principal de este enfoque es aumentar el throughput de la red y reducir las pérdidas de paquetes mediante un mecanismo de retroceso inteligente que minimiza la interferencia en el acceso al medio sin requerir señalización adicional [5]. Además, se evalúa el desempeño del protocolo propuesto en términos de throughput y eficiencia energética, comparándolo con esquemas tradicionales como el Back-off Exponencial Binario (BEB) [6]. A través de un modelo de simulación, este estudio pretende demostrar que la combinación de RIS y Q-Learning mejora el rendimiento de la red y establece un camino hacia el desarrollo de protocolos más avanzados, integrando técnicas como el aprendizaje profundo para una gestión aún más eficiente de los recursos.

1.1 Antecedentes

Investigaciones proponen un algoritmo basado en Deep Q-Network (DQN) para optimizar la multi conectividad en redes celulares heterogéneas, mostrando que el enfoque DQN mejora significativamente el rendimiento de la red, alcanzando un rendimiento cercano al óptimo. Las conclusiones subrayan la capacidad del DQN para optimizar la configuración de multi conectividad y su potencial para futuras implementaciones en redes 5G.

Además de que el enfoque DQN puede ofrecer mejoras significativas en la tasa de datos y eficiencia energética, especialmente en escenarios de redes 5G y 6G, al optimizar la configuración de las superficies inteligentes reconfigurables (RIS).

1.2 Planteamiento del problema

El aumento de la demanda mundial de servicios inalámbricos de alta velocidad, especialmente en entornos urbanos y rurales altamente densos, ha ejercido una presión significativa sobre la infraestructura de telecomunicaciones. Las redes de sensores inalámbricos, que se utilizan extensamente para aplicaciones de IoT, monitoreo y ciudades inteligentes, sufren desafíos críticos de capacidad de throughput y eficiencia energética. La congestión canal, la interferencia y la variabilidad en la carga de tráfico son problemas comunes que afectan el rendimiento de estas redes. Aunque las soluciones tradicionales de MAC basadas en la tentativa de BEB son eficaces en entornos de tráfico bajo, su eficiencia disminuye significativamente con el volumen de datos y el número de dispositivos, por lo tanto, se requiere un enfoque innovador para gestionar el acceso al canal y mejorar el rendimiento de las redes de sensores inalámbricos [7].

Con el desarrollo de redes 5G y posteriores, emergente tecnología como las Superficies Inteligentes Reconfigurables (RIS) se considera una de las soluciones clave para mejorar la cobertura y alojamiento de redes. Por lo tanto, a diferencia de los reflectores de metamateriales, las RIS establecen la capacidad de modificar de manera direccional y activa las propiedades de las ondas electromagnéticas del entorno. Se podrían lograr mayores niveles de señal a través de la reducción del nivel de interferencia, y, por lo tanto, mayores velocidades de propagación, para utilizar completamente las capacidades de las RIS, se requiere una mayor eficiencia en el protocolo de acceso al medio. La solución más prometedora es el uso de algoritmos de aprendizaje de refuerzo, y en particular el Q-learning que ofrece una optimización del canal de acceso dinámica y adaptativo.

En América Latina, los desafíos con respecto al acceso y la calidad de los servicios de telecomunicaciones son aún más severos debido a la infraestructura de red poco desarrollada en muchas áreas y la disparidad en la cobertura de red. Aunque hay progreso para instalar redes 5G y ampliar las redes inalámbricas, hay obstáculos significativos debido a la escasez total de espectro, la propagación del ruido en entornos superpoblados y las dificultades que implica el terreno y las condiciones geográficas en áreas rurales. Se traduce en una capacidad limitada para las redes de sensores inalámbricos que son críticas para ciertas aplicaciones, como el monitoreo ambiental, la agricultura de precisión y la gestión de recursos, además del fin de los protocolos de control de acceso al medio sin optimizar, es ineficaz en la gestión de la creciente demanda de tráfico [8].

Además, los sensores inalámbricos a menudo están formados por dispositivos de bajo consumo, lo que significa que la gestión de la energía es de vital importancia para garantizar la mayor vida útil posible de la red. Lamentablemente, tal como se mencionó anteriormente, el control de acceso al medio puede ser ineficiente y la tentativa constante de retransmitir los datos que no se procesan junto con el uso inexacto del espectro puede abrumar el consumo energético, por lo tanto, tal como se mencionó anteriormente en relación con el uso de Q-learning, donde la RIS admite una distribución de recurso más eficiente y una toma de decisiones más oportuna debido al cambio del estado del canal, estas soluciones pueden ayudar a mejorar la eficiencia energética. Además, implementar tales soluciones en Latinoamérica puede garantizar la sostenibilidad de las redes y el uso optimizado de los

recursos, ya que la región consume muchos servicios de telecomunicaciones sin una infraestructura establecida.

1.3 Justificación

La implementación de un protocolo de control de acceso al medio basado en Q-learning para redes inalámbricas de sensores asistidas por RIS es de gran relevancia debido a la creciente demanda de redes más rápidas y eficientes. Las redes de sensores inalámbricos son esenciales para aplicaciones en áreas como monitoreo ambiental, ciudades inteligentes y agricultura de precisión. Sin embargo, estas redes enfrentan importantes desafíos en términos de throughput y eficiencia energética, particularmente en escenarios con alta densidad de tráfico y interferencia. El protocolo propuesto ofrece una solución adaptativa que optimiza el acceso al medio, mejorando significativamente el rendimiento de la red al aprovechar el aprendizaje automático para gestionar de manera eficiente los recursos de la red y la configuración de las RIS, que a su vez mejora la calidad de la señal.

Por tanto, un protocolo basado en Q-learning permitirá a las redes de sensores adaptarse en tiempo real a las condiciones cambiantes del entorno, mejorando la eficiencia espectral y reduciendo la congestión del canal. Las RIS proporcionan una forma innovadora de manipular las ondas electromagnéticas, mejorando las señales en propagación y disminuyendo la interferencia, lo que es especialmente beneficioso en las redes de sensores de bajo coste y bajo consumo. Al combinar Q-learning con RIS, el sistema es capaz de tomar decisiones informadas sobre cómo utilizar las comunicaciones y los recursos de virtualización, otorgando un rendimiento óptimo con la menor cantidad de energía, lo que es esencial para la sostenibilidad de las redes.

Este enfoque tiene un alto potencial para transformar las redes inalámbricas de sensores al proporcionar una solución escalable y eficiente para el control del acceso al medio. La comparación con el protocolo tradicional de Back-off Exponencial Binario (BEB) permitirá demostrar las ventajas del Q-learning en términos de throughput y eficiencia energética. Este avance no solo contribuirá al desarrollo de redes más eficientes y sostenibles en el contexto de 5G y 6G, sino que también abrirá nuevas posibilidades para la integración de tecnologías emergentes, como el Internet de las Cosas (IoT) y las ciudades inteligentes, mejorando la calidad de vida de los usuarios y la eficiencia de las infraestructuras.

1.4 Objetivos

1.4.1 Objetivo general

Analizar un protocolo de control de acceso al medio (MAC) basado en Q-learning (QL) para incrementar el throughput en una red inalámbrica de sensores asistidos por superficies reconfigurables inteligentes (RIS).

1.4.2 Objetivos específicos

- Desarrollar un modelo de simulación del protocolo de control de acceso al medio basado en Q-learning en una red inalámbrica de sensores asistidos por RIS.

- Aplicar el modelo del protocolo de control de acceso al medio basado en Q-learning para determinar el throughput de la red en escenarios con diferentes cargas de tráfico.
- Comparar los resultados del protocolo basado en Q-learning con un protocolo tradicional de control de acceso al medio (Back-off Exponencial Binario), para evaluar las mejoras alcanzadas en términos de throughput y eficiencia energética en la red.

CAPITULO II

MARCO TEÓRICO

2.1. Estado del Arte

Según una investigación realizada por Hernández et al. titulada “A Deep Q-Network-Based Algorithm for Multi-Connectivity Optimization in Heterogeneous Cellular-Network” propone un algoritmo basado en Deep Q-Network (DQN) para optimizar la conectividad en redes celulares heterogéneas, buscando mejorar la asignación de tráfico entre diversas celdas. El objetivo es distribuir eficientemente el tráfico entre diferentes tecnologías de acceso de radio (RAT) para maximizar el rendimiento de la red. La metodología emplea un modelo de aprendizaje por refuerzo, donde el agente DQN aprende a partir de experiencias para seleccionar las configuraciones óptimas de conectividad. Los resultados muestran que el enfoque DQN alcanza un rendimiento cercano al óptimo, superando estrategias basadas en la relación señal-ruido (SINR) en términos de eficiencia de throughput. Las conclusiones destacan el potencial de DQN en redes 5G/6G y su capacidad para mejorar la eficiencia sin requerir una búsqueda exhaustiva de configuraciones [9].

Alnoman et al. en su investigación titulada “Emerging AI and 6G-Based User Localization Technologies for Emergencies and Disasters” aborda las tecnologías de localización emergentes que asisten en la localización de personas durante emergencias, utilizando redes inalámbricas de sensores, UAVs y superficies reconfigurables inteligentes (RIS). El objetivo es explorar cómo las tecnologías de localización basadas en 6G, como RIS, UAVs y aprendizaje automático, pueden mejorar la precisión de la localización en escenarios de desastre. La metodología incluye el análisis de diferentes enfoques tecnológicos y su aplicabilidad a situaciones de emergencia. Los resultados muestran que el uso de RIS y UAVs puede mejorar significativamente la precisión y eficiencia de la localización. Las conclusiones destacan el potencial de estas tecnologías para transformar la respuesta ante desastres, subrayando la necesidad de integrarlas en redes 6G para optimizar la comunicación y la localización en tiempo real [10].

Por su parte, Ali et al. en su investigación titulada “A Critical Analysis of Cooperative Caching in Ad Hoc Wireless Communication Technologies: Current Challenges and Future Directions” aborda los desafíos técnicos y las oportunidades del uso de caching cooperativo en redes inalámbricas ad hoc como MANET, WMN, WSN y VANET. El objetivo es analizar cómo el caching cooperativo mejora la eficiencia en la distribución de contenido, reduciendo la congestión de red, el consumo de energía y la latencia en la transmisión de datos. La metodología incluye una revisión de diversas técnicas de caching, destacando sus aplicaciones y limitaciones en redes ad hoc. Los resultados indican que el caching cooperativo mejora significativamente el rendimiento, aumentando la tasa de aciertos y reduciendo el retraso en la recuperación de datos. Las conclusiones subrayan la importancia de estas técnicas en el diseño de redes inalámbricas más eficientes y energéticamente sostenibles, especialmente en el contexto de IoT y 6G [11].

En otro aspecto, la investigación realizada por Ibarra et al. titulada "Machine Learning Strategies for Reconfigurable Intelligent Surface-Assisted Communication Systems—A

Review" aborda el uso de estrategias de aprendizaje automático (ML) para mejorar el rendimiento de sistemas de comunicación inalámbrica asistidos por superficies inteligentes reconfigurables (RIS). El objetivo es revisar y comparar diversas metodologías ML aplicadas a sistemas RIS para optimizar el rendimiento en redes 5G y 6G, especialmente en términos de eficiencia energética, maximización de tasa de datos y mejora de la calidad de señal. La metodología incluye la evaluación de enfoques como el aprendizaje supervisado, no supervisado, y el aprendizaje por refuerzo (RL). Los resultados indican que las técnicas de ML, en particular el RL, mejoran significativamente la eficiencia del sistema RIS. Las conclusiones subrayan el potencial de combinar ML con RIS para diseñar redes inalámbricas más eficientes, aunque también señalan la necesidad de superar desafíos como la estimación precisa de canales [12].

Mengnan et al. en su investigación titulada "Reconfigurable Intelligent Surfaces for Wireless Communications: Overview of Hardware Designs, Channel Models, and Estimation Techniques" aborda la tecnología emergente de superficies inteligentes reconfigurables (RIS) para mejorar las comunicaciones inalámbricas. El objetivo es ofrecer una visión general de los avances recientes en el diseño de hardware RIS, modelado de canales y técnicas de estimación de canales para su integración en redes de comunicaciones futuras. La metodología se centra en una revisión detallada de las arquitecturas RIS, sus modos de operación y los enfoques de estimación de canales. Los resultados indican que RIS tiene un potencial significativo para mejorar la cobertura, la capacidad y la eficiencia energética de las redes inalámbricas. Las conclusiones destacan la importancia de la estimación precisa del canal y el control inteligente del entorno electromagnético, sugiriendo futuras investigaciones para integrar RIS en redes 5G/6G. [8]

2.2. Fundamentación teórica

2.2.1. Redes Inalámbricas de Sensores (WSN)

2.2.1.1. Definición

Según una investigación realizada por Romero et al. "los WSN son formadas por dispositivos equipados con sensores que interactúan para realzar una tarea común, también son conocidos como nodos sensores los cuales captan información de su entorno, en donde es procesada y enviada a un nodo sumidero, el cual al mismo tiempo transmite los datos a una pasarela conectada a internet" [13, p. 2].

De la misma manera, otra investigación realizada por Tatayo y Llusi mencionan que "los WSN son sistemas prometedores para monitorear e interactuar con el mundo físico, presentando una diversidad de aplicaciones en las áreas de transporte, seguridad, demótica, etc." [13, p. 3]. Estas redes se emplean en entornos donde el cableado es limitado o las tecnologías inalámbricas son tradicionales como bluetooth, estos sensores tienen como objetivo adquirir variables del entorno mediante sensores, procesar datos y enviar la información de manera inalámbrica, caracterizándose por su bajo consumo de energía, lo que las hace útiles en lugares donde la energía eléctrica es limitada.

2.2.1.2. Arquitectura y topologías de red

Entre las arquitecturas de la red inalámbrica de sensores se encuentran tres principales:

1. Ad-hoc o IBSS, los cuales son equipos integrados que proveen información directamente entre si sin necesidad de infraestructura centralizada, esos se utilizan en lugares pequeños donde no se requiere mayor cobertura. Para la aplicación, los nodos sensores pueden comunicarse directamente entre vecinos sin depender de un punto de acceso central [13].
2. Infraestructura o BSS, en donde los dispositivos no se comunican directamente entre sí, necesitan un punto de acceso que sirve de enlace para interconectar los clientes. El punto de acceso es el encargado de transmitir paquetes de datos de datos y encaminarlos a su destino. Además, para la aplicación, corresponde a la topología estrella, donde los nodos sensores envían datos a un nodo coordinador o Gateway central.
3. Mesh (malla), los cuales son redes múltiples BSS conectados mediante un sistema de distribución, obteniendo amplia cobertura, si falla una conexión, la red puede seguir operando sin problemas gracias a rutas alternativas, esta es la topología más común en WSN, permitiendo redundancia y autoconfiguración (Ver figura 1).

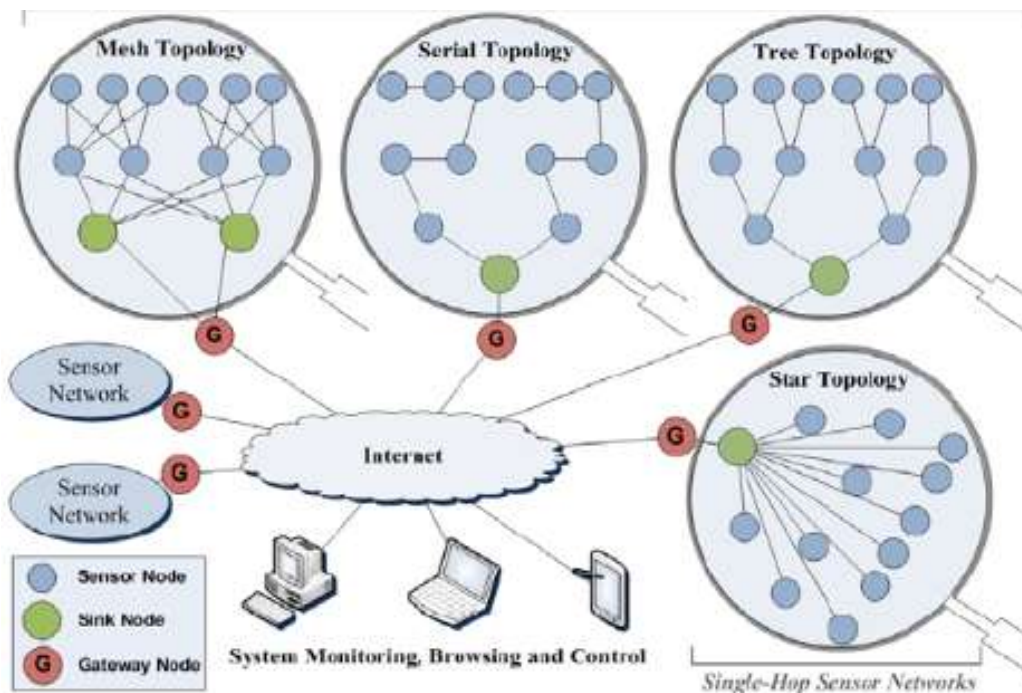


Figura 1. Topologías en WSN [14].

La figura 1 muestra un sistema de supervisión y control empleando distintos tipos de topologías WSN, cuando la topología es un árbol, se la identifica también como redes jerárquicas. Un clúster es administrado por un nodo seleccionado conocido como cabeza de clúster, el clúster head es responsable de coordinar las actividades de los miembros y comunicarse con otros CHs la BS. Las técnicas de clúster ayudan a limitar el consumo de energía a través de diferentes medios, entre estos se encuentran que: reducen el rango de comunicación en el interior el clúster que requiere menos potencia de transmisión, otro es que limitan el número de transmisiones gracias a la fusión realizada por el CH. además de reducir operaciones intensivas en energía como coordinación y agregación a la cabeza del clúster, también permite apagar algunos nodos dentro del clúster mientras que el CH asume

las responsabilidades de reenvío, de la misma manera también equilibran el consumo de energía entre los nodos a través de la rotación del CH.

2.2.1.3. Componentes de un nodo sensor

Entre los componentes se encuentran la puerta de enlace, router y nodo final, en donde la puerta de enlace o Gateway es el punto de unión de la WSN con la red LAN o con internet, además su función principal es recoger los datos de la red y transmitirlos al exterior y monitorear la red. La segunda es el router o enrutador el cual se encarga de retransmitir los datos de los nodos finales hacia la puerta de enlace, esto permite tener una cobertura extensa mediante comunicación multisalto. El tercero es el nodo final o end device, el cual está compuesto por sensores encargados de capturar los datos del ambiente, estos son los dispositivos que realizan las mediciones (Ver figura 2) [15].

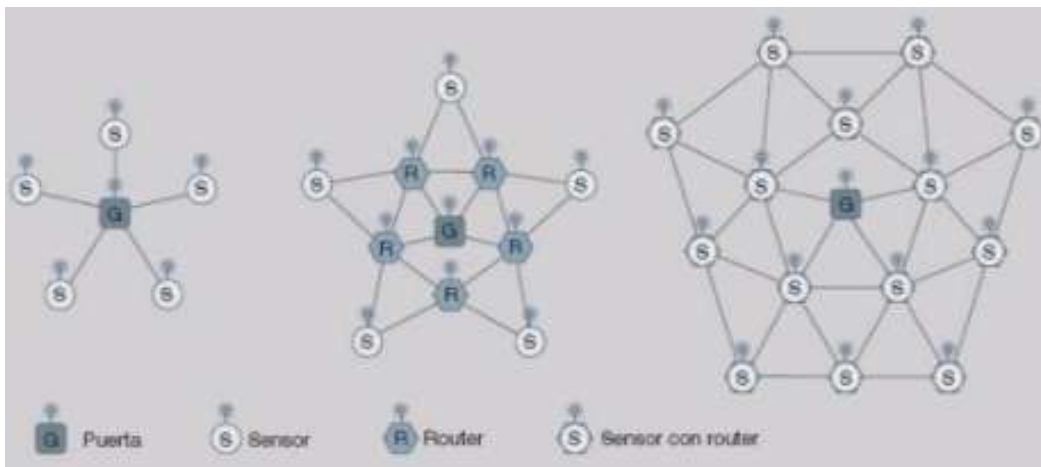


Figura 2. Componentes de WSN [15].

2.2.1.4. Aplicaciones

Las aplicaciones de las redes inalámbricas de sensores son variadas y cada vez más conocidas entre la sociedad, entre estas aplicaciones se encuentran, ciudades inteligentes, atención médica inteligente, redes hidráulicas, redes eléctricas, agricultura y sistemas de transporte inteligente, entre otros más [16].

Además de ello, estas redes inalámbricas presentan aplicaciones en la monitorización de ambientes, seguimiento de objetos, monitoreo de pacientes hospitalarios, y procesos industriales e infraestructuras; en entornos militares se utilizan también este tipo de redes para fines de seguridad y vigilancia. Por otro lado, se destaca su implementación en aplicaciones clave de ciudades inteligentes, agricultura inteligente y sistemas de transporte inteligentes, donde las WSN son esenciales para el monitoreo y control de diferentes variables del entorno. Así, se hace posible la interconexión y la realidad actual de dispositivos, objetos que colaboran logrando un fin, para el cumplimiento de los objetivos comunes, lo que exige la característica de conectividad en tiempo real esencial para el paradigma de Internet de las Cosas (IoT) [17]

2.2.1.5. Caso

Una investigación realizada en el entorno industrial menciona que esta red se utiliza en áreas laborales para medir variables como la temperatura, humedad e iluminación, con el fin de brindar información accesible y permanente, en este caso se observa por medio de la plataforma web, permitiendo el acceso y análisis en tiempo real. De esta forma, con la WSN se pueden establecer medidas de control y prevención que garanticen condiciones seguras y saludables para reducir la exposición del hombre a condiciones dispares y perjudiciales. Por ende, se puede considerar que, al mejorar las condiciones laborales, se logra una mayor productividad [18].

2.2.1.6. Desafíos en las redes de sensores

Una investigación realizada por Powell et al. menciona que existen desafíos, y entre estos principales retos que enfrenta esta tecnología se encuentran [19]:

- **Energía:** El consumo de energía es un problema significativo como la mayoría de las WSN se alimentan de baterías, que limitan la vida útil. El paradigma de las Redes Definidas por Software ayuda a disminuir el problema creando un controlador central al que se le transfieren las tareas del controlador, lo que reduce este consumidor de energía de los nodos sensores.
- **Gestión de red:** Difícil administración de la red y reconfiguración de los nodos debido a que son sistemas tradicionales bastante rígidos y no son fáciles de administrar ya que todos los nodos deben ser reevaluados y reconfigurados cada vez que se cambia cualquier decisión de enrutamiento.
- **Escalabilidad:** A medida que crece un WSN se vuelve ineficiente, el enfoque permite una mejor gestión de la escalabilidad de la red, lo que mantiene su eficiencia a medida que la red se expande.

Por otra parte, Vera et al. mencionan otras complicaciones que tienen estas redes de sensores, entre estas se observan [20]:

- **Desafíos en la comunicación y transmisión de datos:** Existe una dificultad en asegurar una transmisión de datos confiables, en particular en entornos industriales donde las condiciones de operaciones son extremas, con una red vulnerable a ataques mal intencionados.
- **Limitaciones en la capacidad de procesamiento de los sensores:** Entre estas limitaciones se encuentran la potencia de procesamiento, memoria y capacidad de comunicación, dificultando de esta manera la eficiencia.
- **Seguridad y protección de datos:** Debido a que estas redes son propensas a ataques cibernéticos como la interceptación de datos y alteraciones.
- **Costos de implementación y mantenimiento:** Este es uno de los principales desafíos por el que pasan las industrias, debido a que estos sistemas a gran escala tienen un costo elevado, y por la complicación del sistema, es necesario las capacitaciones al personal, teniendo de esta manera un costo adicional para la implementación del sistema de red inalámbrico por sensores.

2.2.2. Protocolos de control de acceso al medio

2.2.2.1. Concepto

El control de acceso al medio, conocido como MAC por su abreviatura en inglés, son conocidos como los métodos usados para manejar y controlar la comunicación dentro de una red, especialmente la que está utilizando un solo canal de comunicación, como en el caso de redes inalámbricas. Involucrando numerosos aparatos que solo pueden comunicarse sin interrupción, MAC maneja la autorregulación de transmisión y la frecuencia, la secuenciación de la secuencia de acceso, y la resolución de problemas a fin de evitar choques [21].

El MAC es un conjunto de mecanismos y protocolos que regulan cómo los dispositivos en una red comparten un canal de comunicación común, con el objetivo de evitar colisiones, maximizar la eficiencia del ancho de banda y garantizar un acceso justo y ordenado al medio de transmisión [22].

2.2.2.2. Clasificación de protocolos MAC

Basado en contención (CSMA/CA, ALOHA)

En estos protocolos, los nodos compiten por acceder al canal cuando necesitan transmitir, en este no existe una coordinación preestablecida estricta, lo que relaja los requisitos de sincronización de tiempo, cabe señalar que su funcionamiento se basa en el acceso aleatorio, un nodo con datos para enviar intenta acceder al canal, si está ocupado, espera. Entre los protocolos destacados se encuentran ALOHA, la cual es la forma más simple, donde los nodos transmiten en cuanto tienen datos. Este se ha combinado con técnicas de muestreo de preámbulo para ahorrar energía en WSN. También se encuentra el CSMA el cual, el nodo escucha el canal antes de transmitir, se detecta actividades, difiere la transmisión. Además, el CSMA/CA el cual es utilizado en estándares como IEEE 802.15.4. este intenta evitar colisiones mediante el intercambio de mensajes de control antes de enviar los datos [23].

Estos protocolos son escalables y se adaptan bien a cambios en la topología, pero sufren de mayores gastos energéticos debido a colisiones, escuchas ociosas y sobre – escuchas.

Basado en planificación (TDMA, FDMA)

Estos protocolos organizan a los nodos siguiendo un esquema definido para que las comunicaciones ocurran de manera ordenada y libre de colisiones, en estas se encuentran las TDMA, FDMA, en donde las primeras el tiempo se divide en ranuras, a cada nodo se le asigna una ranura específica para transmitir o recibir, permitiéndole dormir el resto del tiempo, entre sus ventajas se encuentra que elimina colisiones y escuchas ociosas, y entre sus desventajas se encuentra que requiere una sincronización de tiempo estricta y es difícil adaptarse a cambio de topologías dinámicas. La segunda, aunque permite el acceso libre de colisiones asignando frecuencias o códigos distintos, suelen requerir hardware adicional o alta complejidad computacional, lo que aumenta el costo y el consumo de energía, contradiciendo los objetivos de bajo costo de la WSN [24].

Protocolo de retroceso exponencial binario

Definición y principio de funcionamiento

El protocolo de Retroceso Exponencial Binario (Binary Exponential Backoff, BEB) es un algoritmo de resolución de colisiones ampliamente utilizado en protocolos MAC basados en contención, particularmente en CSMA/CA según el estándar IEEE 802.11 e IEEE 802.15.4. Su principio fundamental consiste en que, tras detectar una colisión, cada nodo involucrado espera un tiempo aleatorio antes de reintentar la transmisión, y dicho tiempo de espera se incrementa exponencialmente con cada colisión consecutiva, de modo que se reduce progresivamente la probabilidad de que los nodos colisionen nuevamente al intentar retransmitir simultáneamente [23][24].

Algoritmo y formulación matemática del BEB

El funcionamiento del protocolo BEB se rige por los siguientes pasos:

Paso 1 - Detección de colisión: Cuando un nodo intenta transmitir y detecta que el canal está ocupado o que se ha producido una colisión, incrementa su contador de reintentos (n) en una unidad.

Paso 2 - Cálculo de la ventana de contención: La ventana de contención (Contention Window, CW) se calcula de forma exponencial. Tras la n -ésima colisión, el nodo selecciona un tiempo de backoff aleatorio dentro del rango $[0, CW - 1]$, donde el tamaño de la ventana de contención se define mediante la siguiente fórmula:

$$CW(n) = \min(2^n \times CW_{\min}, CW_{\max})$$

Donde:

$CW(n)$ es el tamaño de la ventana de contención tras la n -ésima colisión.

n es el número de colisiones consecutivas experimentadas ($n = 0, 1, 2, \dots, m$).

CW_{\min} es el tamaño mínimo de la ventana de contención (valor inicial). En IEEE 802.15.4, $CW_{\min} = 8$ ranuras.

CW_{\max} es el tamaño máximo de la ventana de contención, que actúa como límite superior para evitar tiempos de espera excesivos. En IEEE 802.15.4, $CW_{\max} = 256$ ranuras.

m es el número máximo de reintentos permitidos antes de descartar el paquete.

Paso 3 - Selección del tiempo de backoff: El nodo selecciona un valor aleatorio uniforme dentro del intervalo $[0, CW(n) - 1]$ y espera ese número de ranuras de tiempo antes de intentar retransmitir. El tiempo de backoff se expresa como:

$$T_{\text{backoff}} = \text{random}(0, CW(n) - 1) \times T_{\text{slot}}$$

Donde T_{slot} es la duración de una ranura de tiempo definida por el estándar de comunicación.

Paso 4 - Transmisión o descarte: Si la transmisión es exitosa, el contador de reintentos se reinicia a $n = 0$ y la ventana de contención retorna a CW_{\min} . Si se alcanza el número máximo de reintentos ($n = m$), el paquete se descarta y se notifica a la capa superior [23][24].

Ejemplo numérico del funcionamiento del BEB

Para ilustrar el funcionamiento del BEB, considérese un escenario con $CW_{min} = 8$ y $CW_{max} = 256$:

En el primer intento ($n = 0$): $CW(0) = \min(2^0 \times 8, 256) = \min(8, 256) = 8$. El nodo elige un backoff aleatorio entre $[0, 7]$.

Tras la primera colisión ($n = 1$): $CW(1) = \min(2^1 \times 8, 256) = \min(16, 256) = 16$. El nodo elige un backoff aleatorio entre $[0, 15]$.

Tras la segunda colisión ($n = 2$): $CW(2) = \min(2^2 \times 8, 256) = \min(32, 256) = 32$. El nodo elige un backoff aleatorio entre $[0, 31]$.

Tras la quinta colisión ($n = 5$): $CW(5) = \min(2^5 \times 8, 256) = \min(256, 256) = 256$. La ventana alcanza su valor máximo y no crece más.

Análisis del throughput del BEB

El throughput normalizado del protocolo BEB en un canal con N nodos que compiten por acceso se puede aproximar mediante el modelo de Bianchi [23], que analiza el rendimiento de CSMA/CA con BEB bajo condiciones de saturación. La probabilidad de transmisión de un nodo en una ranura de tiempo dada (τ) y la probabilidad de colisión condicional (p) se relacionan de la siguiente forma:

$$\tau = 2 / (CW_{min} + 1) \quad (\text{aproximación simplificada para baja carga})$$

La probabilidad de que al menos un nodo transmita exitosamente (sin colisión) en una ranura es:

$$P_s = N\tau(1 - \tau)^{N-1}$$

Y el throughput normalizado S del sistema se expresa como:

$$S = P_s \times T_{payload} / E[T_{slot}]$$

Donde $T_{payload}$ es el tiempo de transmisión de la carga útil y $E[T_{slot}]$ es la duración esperada de una ranura de tiempo, que incluye periodos vacíos, transmisiones exitosas y colisiones. A medida que el número de nodos N aumenta, la probabilidad de colisión crece y el throughput del BEB disminuye significativamente, ya que el mecanismo de retroceso exponencial solo reacciona ante las colisiones de forma individual, sin una visión global del estado de la red [23][24].

Ventajas del BEB

Simplicidad de implementación: El BEB es un algoritmo sencillo que no requiere coordinación centralizada ni intercambio de información entre nodos, lo que lo hace fácil de implementar en dispositivos con recursos limitados [23].

Distribución de carga: Al duplicar la ventana de contención tras cada colisión, el BEB distribuye los intentos de acceso en el tiempo, reduciendo la probabilidad de colisiones repetidas en escenarios de congestión moderada [24].

Compatibilidad con estándares: El BEB es el mecanismo de backoff predeterminado en estándares ampliamente adoptados como IEEE 802.11 (Wi-Fi) e IEEE 802.15.4 (ZigBee), lo que garantiza interoperabilidad [23].

Limitaciones del BEB

Latencia indeterminada: No garantiza un límite superior para el tiempo de espera. Tras múltiples colisiones consecutivas, los nodos pueden experimentar retardos excesivos e impredecibles que degradan el rendimiento en aplicaciones sensibles al tiempo [23].

Ineficiencia energética: Durante los periodos de backoff, el transceptor del nodo permanece activo consumiendo energía sin lograr transmitir datos útiles. En redes de sensores alimentadas por batería, esta ineficiencia reduce significativamente la vida útil de la red [23].

Degradación bajo alta carga: Cuando el número de nodos competidores es elevado, la probabilidad de colisión aumenta exponencialmente, lo que provoca ventanas de contención muy grandes y un throughput que decrece drásticamente. El BEB no se adapta de forma inteligente a las condiciones de la red; simplemente reacciona de manera ciega ante las colisiones [24].

Inequidad en el acceso: Un nodo que acaba de transmitir exitosamente reinicia su ventana de contención a CW_{min} , mientras que los nodos que han colisionado mantienen ventanas más grandes. Esto genera una desigualdad en la probabilidad de acceso al canal, favoreciendo a los nodos que ya han tenido éxito [24].

Falta de adaptabilidad: El BEB es un mecanismo estático y reactivo que no considera el estado global de la red, la calidad del canal ni la prioridad de los paquetes. Esta rigidez lo hace inadecuado para entornos dinámicos como las redes asistidas por RIS, donde las condiciones del canal cambian continuamente [24].

2.2.2.3. Limitaciones de los protocolos tradicionales de control de acceso al medio

Los protocolos MAC tradicionales tales como CSMA/CA, TDMA, FDMA y Aloha tienen varias desventajas en entornos dinámicos y con limitaciones de recursos, como las redes asistidas por superficies reconfigurables inteligentes RIS y las redes inalámbricas de sensores WSN.

Protocolos Basados en Contienda (CSMA/CA, Aloha)

Los protocolos de contienda, regulan el acceso al canal en redes inalámbricas, pero tienen un conjunto de problemas, los entornos densos, donde un gran número de nodos pueden intentar transmitir en ese mismo momento, conduce a colisiones y, por lo tanto, a retransmisiones y un consumo de energía aún mayor. En el caso de los protocolos basados en CSMA/CA, la idea de retardo de tiempo entre reintentos al utilizar el retroceso de complejidad puede llevar a tiempos de espera prolongados antes de que un nodo pueda retransmitir, también surgen problemas tales como el nodo oculto, donde dos nodos que no pueden escucharse uno al otro transfieren al mismo tiempo a un mismo receptor, colisionando. O nodo expuesto, que no transmite intencionalmente al detectar actividad en el canal, aunque es seguro hacerlo ya que no puede interferir en las transmisiones [25].

Protocolos Basados en Asignación Estática (TDMA, FDMA)

Los protocolos basados en asignación estática, tienen limitaciones significativas en términos de redes con tráfico variable o estructuras dinámicas, la principal es la falta de flexibilidad, en donde la asignación fija de los intervalos de tiempo, que se pueden usar en TDMA o las frecuencias, que se pueden usar en FDMA, no cumple con el uso real del canal. Por ejemplo, en redes de sensores inalámbricos, si un nodo no tiene ningún dato. Además, este esquema exige una sincronización precisa entre los nodos, lo cual conlleva una sobrecarga significativa de control y uso de energía, especialmente en redes extensas o móviles [26].

Limitaciones Generales en Entornos con RIS

En este tipo de entorno, los protocolos tradicionales son severamente limitados no solo por no usar las RIS para redirigir y dinámicamente aprovechar las señales, sino también por su naturaleza estática: desconocen la capacidad de la RIS de mejorar la comunicación entre nodos ocultos o de evitar colisiones. Por otro lado, los protocolos MAC actuales suelen estar asociados con un alto consumo de energía, ya que a menudo requieren escuchas constantes del canal o retransmisiones, lo que hace difícil utilizar estos protocolos en entornos en los que los dispositivos involucrados tienen un suministro de energía limitado [27].

2.2.2.4.Importancia de un protocolo eficiente en redes de sensores inalámbricos

Un protocolo de control de acceso al medio eficiente es vital en las WSN, ya que las características específicas de estas redes pueden causar problemas de operación, en este sistema, los nodos de los sensores están limitados en la energía, se caracterizan por un ancho de banda reducido y exhiben una alta densidad y la comunicación, necesitando un MAC que minimice las colisiones, reduzca la energía consumida y maximice el throughput. Los protocolos de MAC tradicionales, como CSMA/CA o TDMA, difícilmente se pueden escalar, sincronizar y adaptar, en particular si la red opera en un entorno dinámico con tráfico cambiante [28].

Por otro lado, en aplicaciones críticas como el monitoreo del medio ambiente, la agricultura de precisión o la salud electrónica, un protocolo MAC ineficiente causará la pérdida de datos, una alta latencia o una cobertura deficiente. La incorporación de técnicas de inteligencia artificial, como Q-learning, permite el desarrollo de protocolos adaptativos que ajustan los parámetros de transmisión de manera dinámica según la calidad del canal y el nivel de congestión de su red, esto es particularmente importante en redes asistidas por RIS donde un MAC optimizado podría hacer un uso más integral de la manipulación de las ondas electromagnéticas para mejorar la calidad de la comunicación [29].

2.2.2.5.Problemática del acceso al medio

El problema subyacente del protocolo MAC en las redes de sensores inalámbricos proviene de la limitación energética crítica, ya que los nodos funcionan con baterías escasas y, en comparación, su transceptor de radio es el dispositivo más hambriento de energía. Los protocolos previos no logran abordar adecuadamente las principales fuentes de desperdicio de energía: la “escucha ociosa”, la “escucha excesiva”, las colisiones que requieren costosas

retransmisiones y la gigante saturación de paquetes de control. Además, la estrategia de diseño debe basarse en la premisa de apagado de radio mientras se garantiza la posibilidad de comunicación [30].

Además, además de la energía, los protocolos enfrentan desafíos de escalabilidad y adaptabilidad en topologías dinámicas, donde los nodos pueden fallar o moverse: los algoritmos centralizados son ineficientes. Las limitaciones de hardware en términos de procesamiento y memoria complican la sincronización precisa necesaria para esquemas organizados, pero los esquemas de contención también sufren del problema del terminal oculto. Finalmente, hay un compromiso inevitable, donde se debe sacrificar la latencia y el rendimiento, a saber, el rendimiento, para maximizar la vida útil de la red, es decir, para adaptarse a las aplicaciones con tráfico esporádico o variable [31].

2.2.3. Aprendizaje por refuerzo

2.2.3.1. Fundamentos

El aprendizaje por refuerzo, es un paradigma de aprendizaje automático en el que un agente inteligente aprende a tomar decisiones óptimas en virtud de la interacción con un ambiente dinámico de simulación, a diferencia de los métodos tradicionales como aprendizaje supervisado, donde el algoritmo se entrena con una colección de ejemplos etiquetados, o aprendizaje no supervisado, donde el algoritmo busca patrones implícitos en datos no etiquetados, en RL el agente funciona acumulando datos por recompensas o penalizaciones. El algoritmo persigue maximizar recompensa total a lo largo del tiempo, por lo que es necesario explorar diferentes accionares y explotar aquellos que producen mayores recompensas en términos acumulados. Este enfoque es muy exitoso en problemas de toma de decisiones secuenciales, como juegos, robótica, control y administración de recursos, aspectos financieros, comerciales y muy relevante en el desarrollo de aplicaciones de vehículos autónomos [32].

El marco formal del aprendizaje por refuerzo se basa en el Proceso de Decisión de Markov (MDP), definido por la tupla (S, A, P, R, γ) , donde S es el conjunto de estados posibles del entorno, A es el conjunto de acciones disponibles para el agente, $P(s'|s, a)$ es la función de transición que define la probabilidad de pasar del estado s al estado s' al ejecutar la acción a , $R(s, a)$ es la función de recompensa que asigna un valor numérico a la ejecución de una acción en un estado dado, y $\gamma \in [0, 1]$ es el factor de descuento que pondera la importancia de las recompensas futuras respecto a las inmediatas. El objetivo del agente es encontrar una política óptima π^* que maximice la recompensa acumulada descontada esperada [32].

2.2.4. Q-Learning

2.2.4.1. Concepto de Q-Learning

Q-learning es un algoritmo de aprendizaje por refuerzo que permite a un agente aprender a tomar decisiones óptimas en un entorno mediante la exploración y la retroalimentación basada en recompensas, funciona asignando valores llamados “Q-valores” a pares de estado y acción, los cuales estiman la utilidad esperada de realizar una acción específica en un

estado determinado [33]. A través de la experiencia, el agente actualiza estos valores utilizando la fórmula de Bellman, con el objetivo de maximizar la recompensa acumulada a largo plazo sin necesidad de un modelo previo del entorno.

2.2.4.2. Formulación matemática del Q-Learning

El algoritmo Q-Learning se fundamenta en la actualización iterativa de una función de valor de acción $Q(s, a)$, que representa la recompensa esperada acumulada al tomar la acción a en el estado s y seguir una política óptima a partir de ese momento. La regla de actualización del Q-Learning, conocida como la **ecuación de Bellman temporal-diferencia**, se define como [33][34]:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha [r_{t+1} + \gamma \max_{a'} Q(s_{t+1}, a') - Q(s_t, a_t)]$$

Donde cada componente se define de la siguiente manera:

$Q(s_t, a_t)$: es el valor Q actual del par estado-acción (s_t, a_t) , que representa la estimación actual de la recompensa esperada al ejecutar la acción a_t en el estado s_t .

α (tasa de aprendizaje): es un hiperparámetro en el rango $[0, 1]$ que controla la velocidad de actualización del Q -valor. Un valor de α cercano a 1 otorga más peso a la nueva información, mientras que un valor cercano a 0 hace que el agente sea más conservador y confíe más en las estimaciones previas. Formalmente, α pondera la corrección del error temporal [33].

r_{t+1} (recompensa inmediata): es la recompensa escalar recibida por el agente tras ejecutar la acción a_t en el estado s_t y transicionar al estado s_{t+1} . En el contexto de protocolos MAC, una transmisión exitosa genera una recompensa positiva, mientras que una colisión genera una penalización negativa.

γ (factor de descuento): es un hiperparámetro en el rango $[0, 1]$ que determina la importancia relativa de las recompensas futuras. Un valor de γ cercano a 1 indica que el agente valora significativamente las recompensas futuras (comportamiento orientado al largo plazo), mientras que un valor cercano a 0 indica que el agente prioriza las recompensas inmediatas (comportamiento miope) [33][34].

$\max_{a'} Q(s_{t+1}, a')$: representa el máximo valor Q posible en el siguiente estado s_{t+1} , considerando todas las acciones a' disponibles. Este término captura la estimación optimista del valor futuro y es lo que hace que Q-Learning sea un algoritmo *off-policy*, ya que el agente actualiza sus Q -valores basándose en la mejor acción posible, independientemente de la acción que realmente ejecute [33].

$[r_{t+1} + \gamma \max_{a'} Q(s_{t+1}, a') - Q(s_t, a_t)]$: este término completo se denomina error de diferencia temporal (TD error). Representa la discrepancia entre la recompensa estimada actual y la recompensa observada más la estimación del valor futuro. El algoritmo Q-Learning busca minimizar este error iterativamente hasta alcanzar la convergencia [34].

Política de exploración-explotación (ϵ -greedy)

Para equilibrar entre la exploración de nuevas acciones y la explotación de acciones conocidas como buenas, Q-Learning emplea comúnmente la estrategia ϵ -greedy, en la cual

el agente selecciona la acción con el mayor Q-valor con probabilidad $(1 - \epsilon)$ y una acción aleatoria con probabilidad ϵ . Formalmente:

$$a_t = \{ \arg \max_a Q(s_t, a) \text{ con probabilidad } 1 - \epsilon ; \text{ acción aleatoria con probabilidad } \epsilon \}$$

El valor de ϵ suele decrecer gradualmente durante el entrenamiento (decaimiento de ϵ), permitiendo al agente explorar ampliamente al inicio y explotar el conocimiento adquirido conforme avanza el aprendizaje [33].

2.2.4.3. Convergencia del Q-Learning

Bajo ciertas condiciones, el algoritmo Q-Learning garantiza la convergencia hacia la función de valor óptima Q^* . Las condiciones suficientes para la convergencia son: que todos los pares estado-acción sean visitados infinitas veces, que la tasa de aprendizaje α satisfaga las condiciones de Robbins-Monro (es decir, la suma de α diverge, pero la suma de α^2 converge), y que el entorno sea un MDP estacionario con recompensas acotadas [33][34].

Según Rahman Swapno et al., el Q-Learning es una técnica fundamental dentro del campo del aprendizaje por refuerzo que permite a un agente aprender a tomar decisiones óptimas en entornos complejos mediante un proceso de prueba y error. En el contexto de la gestión del tráfico urbano, el algoritmo DQL se aplica para optimizar los tiempos de las señales de tráfico y minimizar la congestión al aprender de manera adaptativa a partir de patrones de tráfico en tiempo real. Este enfoque emplea redes neuronales profundas para aproximar la función Q, la cual estima el valor esperado de recompensas futuras dado un estado y una acción específicos. A diferencia de los métodos tradicionales, el DQL no requiere un modelo previo del entorno, sino que interactúa con él directamente, almacenando experiencias pasadas en una memoria de repetición (*replay memory*) y ajustando sus predicciones mediante el cálculo del error temporal. Este sistema ha demostrado ser capaz de reducir significativamente las colas de vehículos y aumentar la eficiencia del flujo vial en intersecciones urbanas [34].

2.2.4.4. Aplicaciones de Q-Learning en redes de comunicación

- **Asignación dinámica de recursos**
 - Contexto: Redes inalámbricas (5G, IoT, VANETs).
 - Uso: Q-Learning permite que los nodos aprendan a asignar canales, ancho de banda o potencia de transmisión sin intervención centralizada.
 - Beneficio: Optimiza el uso del espectro y reduce interferencias.
- **Control de acceso al medio (MAC)**
 - Contexto: Redes ad hoc y sensores.
 - Uso: Aprender políticas para decidir cuándo transmitir o esperar, evitando colisiones.
 - Beneficio: Mejora la eficiencia del canal y reduce el retraso.
- **Selección y mantenimiento de rutas (Routing)**
 - Contexto: Redes móviles ad hoc (MANETs), VANETs.
 - Uso: Encontrar rutas óptimas y adaptativas basadas en la calidad del enlace y congestión.

- Beneficio: Alta adaptabilidad en entornos cambiantes y móviles.
- **Seguridad en redes**
 - Contexto: Redes heterogéneas y críticas.
 - Uso: Detección de ataques y respuesta adaptativa (como denegación de servicio o intrusiones).
 - Beneficio: Permite tomar acciones preventivas y reactivas sin reglas fijas.
- **Gestión de tráfico en redes definidas por software (SDN)**
 - Contexto: Redes de centros de datos y SDN.
 - Uso: Aprender a reconfigurar dinámicamente la red para equilibrar carga y evitar congestión.
 - Beneficio: Mejora la calidad del servicio (QoS) en tiempo real.
- **Optimización en redes cognitivas (CRN)**
 - Contexto: Redes con radios que detectan y se adaptan al entorno espectral.
 - Uso: Q-Learning ayuda a decidir cuándo y dónde acceder al espectro libre.
 - Beneficio: Maximiza el uso del espectro sin interferir con usuarios primarios.
- **Control de movilidad y handoff**
 - Contexto: Redes móviles celulares.
 - Uso: Aprender cuándo cambiar de celda/base según parámetros como señal, congestión, o velocidad del usuario.
 - Beneficio: Reduce cortes y mejora la continuidad del servicio.
- **Redes vehiculares (V2V / V2I)**
 - Contexto: Comunicación entre vehículos o con la infraestructura.
 - Uso: Aprender patrones de tráfico y optimizar el envío de datos en función del entorno vial y la densidad.
 - Beneficio: Mejora la seguridad vial y la eficiencia de comunicaciones en tiempo real.

2.2.4.5. Algoritmos de Q-Learning en redes inalámbricas

Estos han resultado ser útiles en las redes inalámbricas para optimizar problemas distribuidos y dinámicos sin necesidad de un modelo exacto del entorno, estos permiten a los nodos individuales, como las estaciones base, los dispositivos móviles y los sensores, a dominar la política de decisión ideal a través de la interacción reiterada entre el entorno. Por ejemplo, en el procedimiento de asignación de canales, Q-Learning exige esencialmente que cada nodo elija el canal de menos interferencia, ajustándose continuamente en tiempo real con respecto a los pocos cambios en el ambiente espectral. Q-Learning también se ha utilizado para manejar la potencia controladora con el objetivo de reducir la energía mientras se mantiene la calidad del servicio; los dispositivos pueden aprender ajustando su potencia de transmisión basándose en las condiciones ambientales y las manifestaciones clave que transmiten [35].

Además, variantes avanzadas como Deep Q-Learning y Multi-Agent Q-Learning han extendido estas capacidades a escenarios más complejos, en redes densas, donde múltiples agentes necesitan tomar decisiones coordinadas, como redes de sensores o redes vehiculares, los algoritmos multi agente permiten el aprendizaje cooperativo o competitivo, que reacciona a las acciones de otros nodos. También se ha implementado en redes CRN para que los dispositivos secundarios aprendan políticas de acceso al espectro que no interfieran con los usuarios principales. Estos algoritmos se combinan para proporcionar soluciones inteligentes y escalables que mejoran la eficiencia espectral, la fiabilidad del enlace y el consumo de energía en redes inalámbricas, incluso en entornos no estacionarios y de alta movilidad [36].

2.2.4.6. Ventajas de Q-Learning para la optimización de recursos en redes de sensores

- Dado que Q-Learning no requiere que se conozca de antemano un modelo del entorno, es la solución perfecta para desplegarlo en RSF, ya que la dinámica del canal, la topología y el tráfico pueden ser impredecibles [37].
- Los nodos RSF pueden aprender una política para seleccionar rutas con niveles adecuados de energía (p. Ej., Ajustando su potencia de transmisión o activando modos de energía baja) para minimizar el consumo de energía de la red y prolongar su vida útil [38].
- Además, Q-Learning adapta dinámicamente su solución a posibles eventos como errores de nodo, interferencias y variaciones en la cantidad de datos disponibles [38].
- Cada nodo puede ejecutar un algoritmo de Q-Learning localmente, sin necesidad de tener un controlador central. Por lo tanto, es una manera escalable de redes de sensores en lo que concierne el uso de grandes despliegues de nodos [37].
- Los nodos por su cuenta son capaces de aprender las mejores maneras de enrutar los datos de tal manera que la red permanezca activa y que la carga en cada nodo no sea excesiva.
- Los nodos usan los resultados de su experiencia para cambiar sus políticas si se caen los enlaces entre vecinos o vecino falla [38].

2.2.5. Superficies Reconfigurables Inteligentes (RIS)

2.2.5.1. Definición y funcionamiento de RIS

Las superficies reconfigurables inteligentes son estructuras o materiales capaces de modificar dinámicamente su forma, textura o propiedades físicas (como rugosidad, adherencia o conductividad) en respuesta a estímulos externos, como cambios térmicos, eléctricos, magnéticos o señales de control algorítmico. Estas superficies integran tecnologías como actuadores, sensores y sistemas de aprendizaje automático para adaptarse autónomamente a diferentes condiciones, optimizando funciones como la aerodinámica en vehículos, la eficiencia energética en edificios o la interacción táctil en dispositivos robóticos. Su diseño suele inspirarse en principios de la biomimética (como la piel de los pulpos o las alas de los insectos) y se aplican en campos como la ingeniería, la medicina y la robótica blanda [39].

2.2.5.2. Ventajas de RIS en redes inalámbricas

- **Mejora de la Cobertura y Capacidad:** Reflejan o redirigen señales hacia zonas con mala conectividad, eliminando "zonas muertas" y aumentando la tasa de transmisión.
- **Eficiencia Energética:** Consumen menos energía que repetidores tradicionales o antenas activas, al no requerir amplificación de señal.
- **Baja Latencia:** Optimizan las rutas de propagación, reduciendo interferencias y retrasos en comunicaciones críticas (ej.: 5G/6G, IoT industrial).
- **Adaptabilidad en Tiempo Real:** Ajustan su configuración ante cambios en el entorno (movimiento de usuarios, obstáculos) mediante control basado en IA o algoritmos.
- **Reducción de Interferencias:** Focalizan la señal hacia receptores específicos, minimizando interferencias con otros usuarios o celdas vecinas.
- **Escalabilidad y Coste-Efectividad:** Son superficies pasivas o semipasivas, más económicas y fáciles de desplegar que infraestructura activa adicional.
- **Compatibilidad con Futuras Tecnologías:** Se integran en redes 5G/6G, MIMO masivo y comunicaciones por luz visible (VLC), potenciando su flexibilidad.

2.2.5.3. Aplicaciones de RIS en redes 5G y 6G

Las Superficies Reconfigurables Inteligentes son una pieza clave en el ecosistema 5G, ya que optimizan la cobertura y la eficiencia espectral. En entornos urbanos densos, se utiliza para reimplementar las señales en ángulo para vencer los reflejos destructivos y guiarlas de manera efectiva hacia los usuarios. En interiores, por ejemplo, en centros comerciales o estadios, las RIS se comportan como reflectores programables, esto no solo permite abordar las zonas de sombra, sino que también puede eliminar la necesidad de infraestructura adicional. Otro punto clave es que las RIS son un dispositivo de energía ultra baja, como resultado, puede extender la conectividad a experiencias en red de IoT masivo, donde los dispositivos de extremo de bajo consumo requieren un enlace, pero no pueden permitirse ningún costo energético adicional [40].

Las RIS llevan la comunicación inalámbrica a nuevos niveles en la futura red 6G, habilitando entornos de propagación inteligentes y radio reconfigurable habilitada, integradas con IA/ML, las RIS pueden auto optimizarse para admitir comunicaciones holográficas, realidad extendida y redes de sensores ultradensas en tiempo real. En frecuencias de terahercios, donde la atenuación es elevada, las RIS superarán las pérdidas de la señal, habilitando enlaces ultradensos, además, permiten "celdas invisibles" en sistemas no terrestres con drones y satélites, lo cual interconecta a nivel mundial sin brechas [41].

2.2.5.4. Integración de RIS en redes de sensores inalámbricos

Las Superficies Reconfigurables Inteligentes (RIS) pueden transformar las redes de sensores inalámbricos al mejorar la eficiencia energética, la cobertura y la confiabilidad de las comunicaciones. En entornos donde los nodos de sensores tienen limitaciones de potencia (como en agricultura inteligente, monitoreo industrial o ciudades inteligentes), las RIS actúan como reflectores programables que optimizan las rutas de propagación de las señales

- Reducción del consumo energético: Los sensores ya no necesitan transmitir con alta potencia, ya que la RIS redirige y focaliza las señales hacia el nodo receptor (*gateway* o nodo sumidero).
- Mayor alcance y cobertura: Las RIS pueden "rellenar" zonas de sombra en entornos complejos (como fábricas o áreas rurales), evitando la necesidad de desplegar repetidores costosos.
- Tolerancia a interferencias: Al controlar dinámicamente la reflexión de ondas, las RIS minimizan colisiones en redes densas con múltiples sensores transmitiendo simultáneamente.

A pesar de sus ventajas, la integración de RIS en WSNs enfrenta desafíos:

- Sincronización y control: Las RIS requieren retroalimentación en tiempo real desde los sensores para adaptarse a cambios en el entorno (ej.: movimiento de obstáculos).
- Escalabilidad: En redes masivas de sensores (miles de nodos), la coordinación entre múltiples RIS y *gateways* necesita protocolos de comunicación eficientes.
- Coste y complejidad: Aunque las RIS son más económicas que los repetidores activos, su despliegue en grandes áreas (como bosques o minas) aún puede ser costoso.

Aplicaciones prácticas

- Agricultura de precisión: RIS mejoran la conectividad de sensores de suelo y clima en campos extensos.
- Monitoreo estructural: En puentes o edificios, garantizan transmisión estable de datos desde sensores de vibración o temperatura.
- Redes de emergencia: Proporcionan conectividad temporal en zonas de desastre donde la infraestructura está dañada.

La integración de sistemas de superficies inteligentes reconfigurables en las redes de sensores inalámbricos revoluciona los entornos de propagación inalámbrica, de un medio pasivo a un controlable por software. Dichas superficies contienen muchos elementos reflectantes pasivos de bajo costo que se pueden ajustar para variar la fase y la amplitud de las señales incidentes. Redirigir las ondas electromagnéticas a los receptores beneficiosos según el modelo de diseño permite a las superficies reconfigurables establecer enlaces de línea de visión virtual, que superan los problemas de bloqueo de canal por obstáculos físicos implementando la calidad del canal sin un hardware de radiofrecuencia complejo en los nodos [42].

El beneficio crítico de esta tecnología para las WSN es la mejora sustancial de la eficiencia energética y la cobertura de la red. Al hacer uso de la formación de haz pasivo, la RIS concentra la energía de la señal, permitiendo a los sensores transmitir a una potencia mucho menor y mantener la calidad del enlace. Esto desencadena los efectos del PLM y prolonga la vida útil de las baterías para sensores con restricciones energéticas. Además, la técnica facilita la conectividad en condiciones hostiles, donde la comunicación directa entre el sensor y la estación base es imposible debido a la distancia o interferencias [43].

La integración avanzada de RIS tiene la capacidad de optimizar aplicaciones críticas como la Transferencia Simultánea de Información y Energía Inalámbrica y la seguridad física. Por medio de algoritmos de optimización, la RIS puede ser manipulada para aumentar la señal en usuarios legítimos y cancelarla en direcciones de presunto espionaje, secuestro de señales para eludir la privacidad. Adicionalmente, optimiza el aprovechamiento de la recolección de energía en los nodos sensores, con lo que las redes se vuelven más sostenibles y autónomas. Este tipo de sinergia entre RIS y WSN es esencial para soportar la masiva densidad de dispositivos esperada para el futuro del Internet de las Cosas IoT [40].

2.2.5.5. Modelo del canal asistido por RIS

El modelo de canal RIS asistido por canal considerado en esta sección describe un sistema de comunicación en el cual la señal se propaga de dos maneras primarias: un enlace directo entre el transmisor y el receptor y un enlace reflejado a través de las CSR. La señal incidente modificada por la CSR matriz, tomando la forma de una matriz de desplazamiento de fase diagonal, que ajusta las propiedades de reflexión de sus elementos pasivos. La señal recibida es, por lo tanto, la superposición en fase coherente de las contribuciones directa y reflejada, lo que permite enlazar y remediar el efecto perjudicial del desvanecimiento del canal [44].

2.2.6. Protocolos MAC basados en aprendizaje por refuerzo

2.2.6.1. Justificación del uso de RL en protocolos MAC

El uso de Aprendizaje por Refuerzo en los protocolos MAC se justifica de la siguiente manera: los enfoques tradicionales, como CSMA/CA o TDMA, incapaces de adaptarse de manera eficiente a entornos de red dinámicos y heterogéneos. Las redes modernas, como las IoT o 5G, presentan condiciones de canal fluctuantes y patrones de tráfico impredecibles que los protocolos basados en modelos predefinidos no pueden administrar de la mejor manera. El Aprendizaje por Refuerzo permite a los nodos generar estrategias de transmisión basadas en la inteligencia a través de la interacción directa con el entorno, es decir, modificar sus parámetros de acceso sin tener conocimiento a priori del modelo del sistema o la red.

Además, el RL aborda correctamente el problema de la optimización de objetivos conflictivos, por ejemplo, maximizar el rendimiento o la tasa de envío de paquetes a través del enlace, pero minimizando al mismo tiempo el consumo de energía y la latencia. A diferencia de las reglas heurísticas fijas, el agente de RL no ha sido establecido para lograr una recompensa a corto plazo. El agente intenta maximizar la suma total de todas las recompensas futuras. Por lo tanto, permitir que el nodo tome decisiones “miopes” o inmediatas para transmitir agresivamente, confiando en la conducta futura para ahorrar energía y evitar colisiones más sofisticado es un equilibrio alcanzado por el agente [45].

Se reconoce que el enfoque de RL también permite que las operaciones se descentralicen y escalen en redes densas. Al modelar a cada uno de los nodos a lo largo de la red como un agente independiente que realiza consultas del pasado para ver sus aprobados previos o transmisiones fallidas, se reduce la necesidad de una señalización de control centralizada. Esto a su vez significa que los nodos pueden resolver problemas complicados debido a la coordinación tales como administrar la interferencia o seleccionar canales dinámicamente

aprendiendo políticas que convergerán en un comportamiento cooperativo o logístico eficiente sin una sobrecarga de comunicación significativa [46].

2.2.6.2. Modelado del problema MAC como MDP

El problema de Control de Acceso al Medio se formula matemáticamente como un MDP con el fin de permitir a los nodos en la red tomar decisiones secuenciales óptimas en un entorno estocástico. En otras palabras, en su marco, cada uno de los nodos en la red se comporta como un agente independiente que interactúa con el canal inalámbrico. Más precisamente, un agente se deduce a la forma en que el canal inalámbrico en pasos discretos en el tiempo. Formalmente, el modelo presentado en una tupla, es decir, S, A, P, R, γ donde el objetivo del agente está implícito en no solo el éxito inmediato, sino también la recompensa a largo plazo descuenta la recompensa restringida. Como resultado, un nodo en la red puede aprender sin ninguna información dinámica del modelo de la red [47].

El espacio de estados (S) y el espacio de acciones (A) forman el núcleo del proceso de toma de decisiones del nodo. El estado es una observación local del agente en un momento determinado, que cubre variables críticas como el estado del canal o, el nivel de ocupación del buffer de datos o capacitación. Bajo este estado, el agente selecciona una acción de dadas, que en una implementación general incluye transmitir un paquete, esperar para backoff o hacer un modo de suspensión para ahorrar energía, definir de esta forma la política que rige su demostración.

El aprendizaje se guía tanto por una función de recompensa (R) como por la dinámica subyacente de transición (P). La función de recompensa asigna un valor numérico inmediato a la acción tomada, es decir, da valores positivos a buenos logros de las acciones. Por ejemplo, vea los valores positivos para una transmisión (throughput) y penalizaciones negativas para las colisiones, pérdidas de paquetes o un alto consumo de energía. Dado que las probabilidades de transición (P) entre estados generalmente son desconocidas o complicadas de modelar debido a el canal inalámbrico es muy difícil de predecir, se utilizan algoritmos de aprendizaje por refuerzo sin modelo a la política. se llama aprendizaje por refuerzo sin modelo al aprendizaje por refuerzo que no intenta modelar el entorno. Q-Learning podría ser un ejemplo prototípico. Menos conocimiento tiene la formulación menos de qué hacer de lo que ha sucedido en el pasado mientras más se fuerza la asignación de recompensa [46].

2.2.7. Optimización del throughput en redes inalámbricas

2.2.7.1. Definición de throughput

Throughput es el ancho de banda efectivo o tasa de transferencia útil es la cantidad de datos reales que se transmitieron con éxito a través de la red o el sistema de comunicación durante algún tiempo específico, por lo general, se mide en bits por segundo (bps), kilobits por segundo (Kbps) o megabits por segundo (Mbps). A diferencia del ancho de banda teórico que se refiere a la capacidad máxima del canal, el radican Throughput se habla de factores que imposible excluir: retrasos, congestión, errores de transmisión y otros factores externos como el *overhead* de protocolos. Por lo tanto, radica Throughput es un indicador más exacto de la funcionalidad real de la red, por ejemplo, si la conexión Wi-Fi tiene un ancho de banda

teórico de 300 Mbps, el ancho de banda efectivo puede ser de solo 200 Mbps debido a la interferencia o la distancia desde el enrutador [48].

2.2.7.2. Factores que afectan el throughput en redes de sensores

Una investigación realizada por Ojeda et al. menciona que entre los factores que afectan el throughput en redes de sensores se encuentran los siguientes [49]:

- **Interferencia y ruido**

La presencia de otras redes inalámbricas, dispositivos electrónicos o fuentes de ruido electromagnético degradan la señal, reduciendo la tasa de transferencia efectiva.

En entornos industriales, maquinaria pesada puede generar interferencias en frecuencias comunes (ej.: 2.4 GHz).

- **Protocolos de comunicación y *overhead***

Protocolos con alto *overhead* (ej.: cabeceras largas en paquetes, mecanismos de retransmisión) consumen ancho de banda útil.

Algoritmos de acceso al medio (como CSMA/CA en ZigBee) introducen retardos y colisiones en redes densas.

- **Topología y enrutamiento**

Rutas multi salto aumentan la latencia y el riesgo de pérdida de paquetes, disminuyendo el throughput.

Desbalances de carga entre nodos pueden saturar ciertos caminos, creando cuellos de botella.

- **Limitaciones de energía**

Sensores con batería restringidas suelen reducir su potencia de transmisión para ahorrar energía

- **Condiciones del canal inalámbrico**

Atenuación por distancia, obstáculos o fenómenos como el fading multicamino distorsionan la señal.

- **Congestión y tráfico colaborativo**

Alta densidad de nodos transmitiendo simultáneamente (ej.: en redes IoT masivas) satura el canal compartido.

Paquetes perdidos por congestión generan retransmisiones, reduciendo el throughput neto.

- **Capacidad del nodo sumidero**

Si el gateway o nodo receptor no puede procesar datos a alta velocidad (por limitaciones de CPU o ancho de banda ascendente), se convierte en un cuello de botella.

2.2.7.3. Métodos para mejorar el throughput en redes inalámbricas

Uso de Multiple Input Multiple Output (MIMO)

MIMO es una tecnología que utiliza múltiples antenas tanto en el transmisor como en el receptor para enviar y recibir múltiples flujos de datos simultáneamente. Esto aumenta la capacidad de la red y mejora significativamente el throughput, ya que se aprovecha mejor el espectro de frecuencias disponible. MIMO es particularmente eficaz en entornos con interferencias y en redes de alta densidad, como en conexiones Wi-Fi de última generación (por ejemplo, Wi-Fi 5 o Wi-Fi 6) [50].

Mejora del Canal de Frecuencia (Canales Anchos)

Ampliar el ancho de banda de los canales utilizados en la transmisión también puede mejorar el throughput. En redes Wi-Fi, por ejemplo, se pueden usar canales de 40 MHz, 80 MHz o 160 MHz en lugar de los tradicionales de 20 MHz. Esto permite transmitir más datos en el mismo período de tiempo. Sin embargo, aumentar el ancho de banda puede causar más interferencias, por lo que es importante considerar la calidad del canal y las condiciones del entorno [51].

Control de Interferencia y Gestión de Recursos

La interferencia es una de las principales causas de pérdida de rendimiento en las redes inalámbricas. Para mitigar esto, es fundamental implementar técnicas de gestión de interferencias, como el *cognitive radio* o el *Dynamic Frequency Selection (DFS)*. Estas tecnologías permiten que los dispositivos inalámbricos elijan dinámicamente el mejor canal disponible o ajusten su potencia de transmisión según el entorno. Además, la gestión eficiente de los recursos de la red, como la asignación de canales y la reducción de colisiones, también es crucial para mantener un alto throughput [52].

2.2.7.4. Uso de técnicas de optimización, como Q-Learning, para aumentar el throughput

Q-Learning es un algoritmo de aprendizaje por refuerzo que permite a un agente aprender cómo tomar decisiones de manera óptima en un entorno de incertidumbre, el algoritmo funciona mediante la actualización de una función de valor de acción $Q(s,a)$, que estima la "calidad" de tomar una acción a en un estado s . En el contexto de redes inalámbricas, este método puede ser utilizado para tomar decisiones sobre cómo gestionar recursos de red, como el canal, la potencia de transmisión o la asignación de ancho de banda.

Ventajas

- Adaptabilidad: Q-Learning permite que el sistema se adapte de forma continua a las condiciones cambiantes del entorno, como la congestión de la red, la interferencia y la calidad del canal.
- Decisiones autónomas: El algoritmo no requiere intervención externa para optimizar las decisiones, lo que facilita su implementación en redes de gran escala y redes ad-hoc.
- Escalabilidad: A medida que la red crece, el algoritmo sigue siendo eficaz al aprender de forma autónoma, sin necesidad de reconfiguración manual.

2.2.8. Python como entorno de simulación

2.2.8.1. Definición

Python es un lenguaje de programación de alto nivel, interpretado e interactivo cuyo creador fue Guido van Rossum finales de los ochenta y principios de los noventa. Python es un lenguaje de programación multiparadigma que incluye programación orientada a objetos, programación imperativa y, en menor medida, programación funcional. Python tiene una sintaxis sencilla y fácil de aprender, con énfasis en la claridad y la legibilidad del código gracias a la utilización para determinar los bloques en lugar de las llaves u otros símbolos. Se utiliza tipado y enlazado dinámico [53].

Este lenguaje es altamente portátil, ejecutándose en diversas plataformas como Unix, Windows y Macintosh, y se distribuye libremente, permitiendo su uso incluso en aplicaciones comerciales sin costo. Python ofrece una extensa biblioteca estándar y mecanismos robustos como la gestión automática de memoria y el manejo de excepciones, funcionando frecuentemente como un lenguaje "pegamento" para conectar componentes de software existentes o como un lenguaje de extensión para aplicaciones que requieren una interfaz programable.

2.2.8.2. Características

Python es un lenguaje de programación interpretado, interactivo, de alto nivel y orientado a objetos que se destaca por ser fácil de aprender, leer y mantener gracias a su estructura simple y sintaxis claramente definida. Se trata de un lenguaje que se diferencia de otros en que delimita bloques de código y no por palabras clave o símbolos, y del cual puede decirse que es multiparadigma debido a su tipado y enlazado dinámico y compatibilidad con programación imperativa, orientada a objetos y funcional [54].

La alta portabilidad de Python permite a este implementarse en diversas plataformas de hardware como Unix, Windows y Macintosh y su use tailandés amplia biblioteca de estándar facilita la conexión con las principales bases de datos comerciales y la programación de GUI. Python también es extensible y escalable, resultando en un lenguaje que puede integrar módulos de bajo nivel y, a menudo, actuar como lenguaje de "enlace" para software ya creado.

CAPITULO III

METODOLÓGICO

3.1. Tipo de investigación

El presente trabajo integró diversos tipos de investigación que, de manera complementaria, contribuyeron al cumplimiento de los objetivos planteados y al desarrollo de un análisis riguroso del protocolo de control de acceso al medio basado en Q-Learning, aplicado en una red inalámbrica de sensores asistida por superficies reconfigurables inteligentes (RIS).

3.1.1. Investigación cuantitativa

La investigación poseyó un carácter cuantitativo, dado que se fundamentó en la recolección de datos numéricos mediante un modelo de simulación computacional. En este sentido, se buscó evaluar el throughput de una red inalámbrica de sensores asistida por RIS, empleando un protocolo de control de acceso al medio basado en Q-Learning (QL). El propósito central de este enfoque fue cuantificar el incremento del throughput y validar estadísticamente los resultados obtenidos.

3.1.2. Investigación exploratoria

De igual forma, el estudio adoptó un enfoque exploratorio, puesto que se orientó a la recopilación y el análisis de información proveniente de estudios científicos relacionados con protocolos de control de acceso al medio en redes inalámbricas de sensores asistidas por RIS. A partir de esta revisión, se identificaron los fundamentos teóricos que sustentaron la propuesta y se evaluó el desempeño del protocolo basado en Q-Learning con énfasis en el incremento del throughput de la red.

3.1.3. Investigación comparativa

Asimismo, la investigación incorporó un componente comparativo, ya que se contrastaron los resultados obtenidos por el protocolo MAC basado en Q-Learning con aquellos generados por el protocolo tradicional de Retroceso Exponencial Binario (BEB). Dicha comparación se llevó a cabo bajo las mismas condiciones de simulación, lo cual garantizó la objetividad y la validez del análisis.

3.2. Diseño de investigación

La presente investigación se desarrolló mediante un diseño de simulación implementado en el lenguaje de programación Python, orientado al análisis de los protocolos de control de acceso al medio en redes inalámbricas de sensores asistidas por superficies reconfigurables inteligentes. Este modelo de simulación permitió observar el desempeño del sistema desde un enfoque transversal, lo que facilitó el análisis de los resultados dentro de un periodo de simulación específico. En consecuencia, las variables de estudio se evaluaron con base en los resultados obtenidos durante la ejecución del modelo, utilizando métricas de desempeño tales como el throughput y la eficiencia energética.

3.3. Métodos, Técnicas e Instrumentos de Investigación

3.3.1. Método analítico

En la presente investigación se empleó el método analítico, ya que permitió evaluar de manera detallada el comportamiento de los protocolos de control de acceso al medio en una red inalámbrica de sensores asistida por superficies reconfigurables inteligentes. Este método facilitó el análisis individual de las variables de estudio, como el throughput y la eficiencia energética. Mediante este enfoque, se logró interpretar el desempeño de cada protocolo y examinar su comportamiento dentro del funcionamiento general de la red, sin la necesidad de realizar intervenciones en un entorno real.

3.3.2. Técnicas de recolección de datos

En la presente investigación se aplicó la técnica de observación, con el propósito de analizar el comportamiento de los protocolos de control de acceso al medio en una red inalámbrica de sensores asistida por superficies reconfigurables inteligentes. Para ello, se empleó un modelo de simulación computacional que permitió observar el desempeño del protocolo basado en Q-Learning bajo diferentes configuraciones del número de sensores, evaluando su impacto en métricas como el throughput y la eficiencia energética.

La implementación del código de programación posibilitó la simulación de los escenarios de funcionamiento de la red, en los cuales se observaron las variaciones en el rendimiento del sistema a medida que cambió el número de sensores. De manera simultánea, se comparó el comportamiento del protocolo propuesto con el del protocolo tradicional BEB. Esta técnica de observación resultó fundamental para el cumplimiento de los objetivos planteados en la investigación, dado que permitió obtener resultados confiables que respaldaron tanto el análisis como la optimización del acceso al medio en redes inalámbricas de sensores asistidas por RIS.

3.3.3. Instrumentos de Recolección de Datos

A continuación, se detallan los instrumentos y recursos metodológicos que permitieron obtener la información relevante para el desarrollo del presente trabajo de investigación.

INSTRUMENTOS	DESCRIPCIÓN
Artículos científicos	Revisión de literatura científica relacionada con protocolos de control de acceso al medio, aprendizaje por refuerzo, redes inalámbricas de sensores y superficies reconfigurables inteligentes, utilizada para sustentar el marco teórico y el modelo de simulación.
Modelo de simulación (Python)	La herramienta para simular una red de sensores asistida por RIS y registrar métricas como throughput.

Tabla 1. Herramientas e instrumentos para la Recolección de Datos.

3.4. Población

La población de la investigación estuvo conformada por redes inalámbricas de sensores asistidos por superficies reconfigurables inteligentes (RIS), en este caso, la población se refería a los sistemas de redes de sensores que utilizan RIS para optimizar el rendimiento de la transmisión de datos, permitiendo la adaptación del canal en tiempo real para mejorar el throughput y la eficiencia energética.

En particular, el sistema considera una red conformada por un número de sensores inalámbricos para calcular las métricas de rendimiento; cada sensor se modela como un agente de aprendizaje QL y participa en un entorno de simulación.

3.5. Operacionalización de variables

VARIABLES	Tipo de variable	Concepto	Indicadores	Técnicas e Instrumentos
Número de sensores en la red.	Independiente	Cantidad de sensores desplegados en la red que compiten por el acceso.	$N = \text{Número}$	Asignación directa
Throughput	Dependiente	Cantidad de datos que se pueden transferir exitosamente de un punto a otro en un período de tiempo determinado.	bits/seg = bits por segundo	Observación directa a través de un modelo de simulación de la red.

Tabla 2. Operacionalización de variables

3.6. Procedimiento y Análisis.

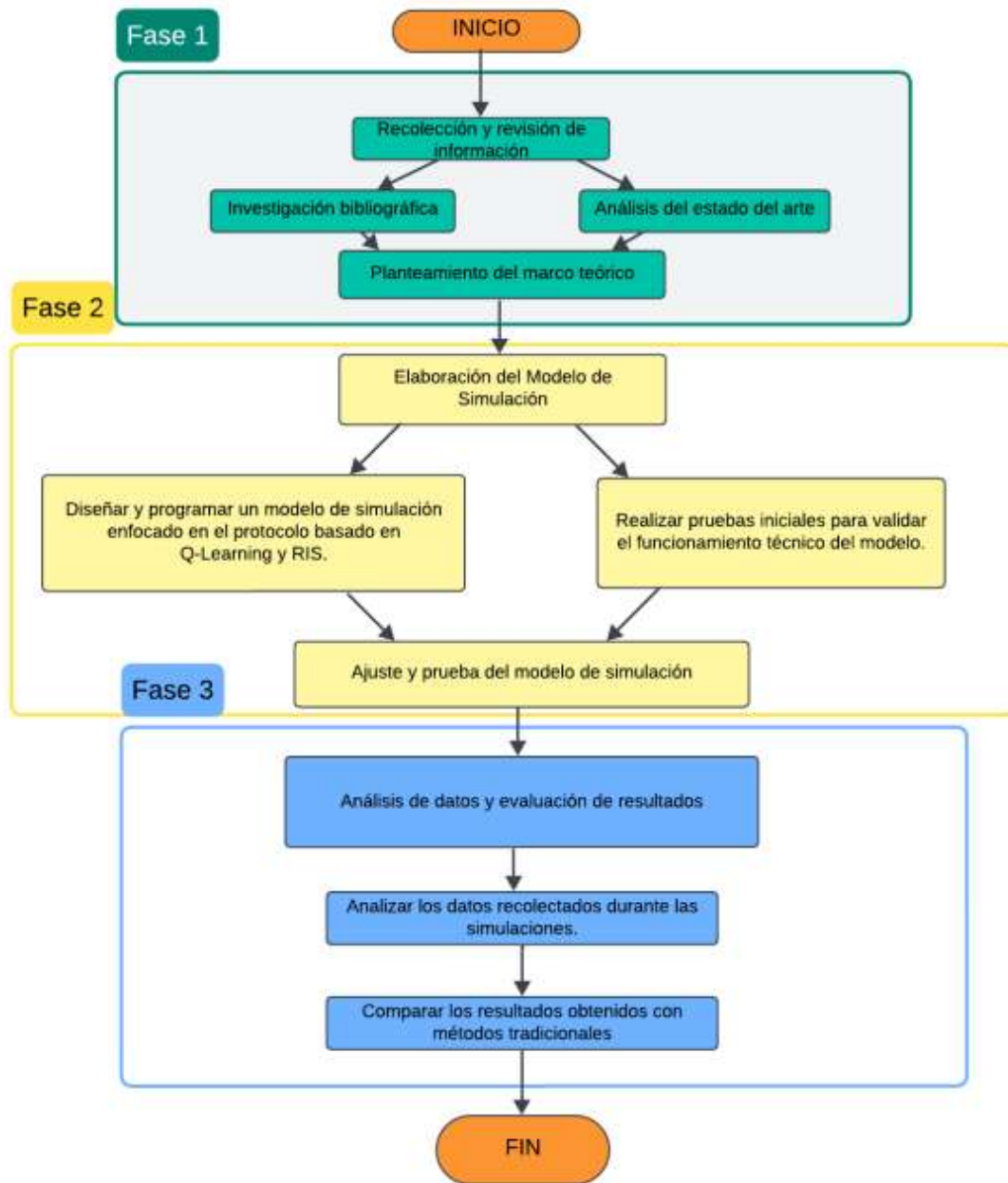


Figura 3. Fases del proyecto de investigación.

En la figura 3 se ilustran las fases que conforman el proceso de investigación, organizadas de manera secuencial, las cuales permiten comprender el desarrollo metodológico aplicado, iniciando con la recopilación de información y finalizando con el análisis y evaluación de los resultados obtenidos a partir del modelo de simulación.

Fase 1 - Revisión y recopilación de información

Esta primera fase se orientó a la búsqueda y el análisis de información científica relacionada con los protocolos de control de acceso al medio en redes inalámbricas de sensores, así como con el uso de superficies reconfigurables inteligentes (RIS). Mediante la revisión bibliográfica, se identificaron antecedentes relevantes, enfoques teóricos y aportes de

investigaciones previas, lo cual permitió comprender el estado actual del tema. Con base en esta información, se estructuró el marco teórico que sustentó conceptualmente el desarrollo del estudio y orientó las etapas subsiguientes de la investigación.

Fase 2 - Desarrollo del modelo de simulación

En esta segunda fase se llevó a cabo la creación e implementación del modelo de simulación, cuyo propósito fue representar el funcionamiento de un protocolo de control de acceso al medio (MAC) basado en Q-Learning, asistido por superficies reconfigurables inteligentes (RIS). Para ello, se diseñó y programó un entorno de simulación capaz de reproducir el comportamiento de una red inalámbrica de sensores bajo diferentes condiciones operativas, lo que permitió analizar de manera controlada las interacciones entre las principales variables del sistema.

El modelo de simulación tuvo como objetivo principal evaluar el desempeño del protocolo MAC propuesto, el cual se orientó a la reducción de interferencias y la optimización del acceso al medio mediante decisiones adaptativas basadas en Q-Learning. A través de este enfoque, se buscó incrementar el throughput de la red, analizando de qué manera las acciones aprendidas por los nodos influyeron en la eficiencia de las comunicaciones al incorporar la tecnología RIS.

PARÁMETRO	DESCRIPCIÓN
Tiempo de simulación	210 segundos, divididos en intervalos discretos.
Tipo de red	Red inalámbrica de sensores asistida por superficies reconfigurables inteligentes (RIS).
Carga de tráfico	Vamos a genera una carga de tráfico considerando 3 escenarios: baja, media, alta.
Tamaño del paquete de datos	1044 bits
Tamaño de ACK/NACK	20 bits
Frecuencia de transmisión	Las transmisiones se evalúan al inicio de cada intervalo de tiempo, y las interferencias en la superficie reconfigurable inteligente se modelan de manera dinámica.
Sistema RIS	Se asume que los sensores se encuentran en proximidad entre sí, lo que permite considerar una relación señal-ruido-interferencia (SINR) similar para todo el conjunto y facilita el análisis del impacto de la tecnología RIS en el rendimiento del sistema.

Tabla 3. Parámetros del modelo de simulación

Fase 3 - Análisis y evaluación de resultados

En esta fase final se llevó a cabo el procesamiento y el análisis de los datos obtenidos durante las simulaciones. Los resultados fueron evaluados considerando parámetros de desempeño relevantes, tales como el throughput y la eficiencia energética de la red. De manera complementaria, se realizó una comparación entre los resultados obtenidos por el protocolo propuesto y los generados por el protocolo tradicional BEB, con el fin de identificar diferencias significativas en el rendimiento. Esta fase facilitó la interpretación de los datos, la comprobación de las hipótesis planteadas y la formulación de conclusiones relacionadas con la efectividad del uso de Q-Learning y RIS.

3.6.1. Estructura y funcionamiento del algoritmo de simulación

Parámetros generales

En esta sección se importaron las librerías necesarias para la simulación: NumPy para cálculos numéricos y generación de variables aleatorias, Matplotlib para la visualización de resultados y Pandas para organizar y exportar la información obtenida. A continuación, se definieron los parámetros globales del entorno simulado, entre los cuales se incluyeron la duración de cada intervalo, la longitud del episodio y el número de episodios tanto para entrenamiento como para evaluación. Adicionalmente, se establecieron el tamaño del paquete de datos y el tamaño de los mensajes de control ACK/NACK, así como la carga de tráfico y el conjunto de valores del número de sensores que fueron analizados. Estos parámetros permitieron estudiar el rendimiento del sistema al variar la cantidad de sensores desplegados en la red (Ver figura 4).

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd

# =====
# PARÁMETROS DE SIMULACIÓN
# =====
SLOT_DURATION = 1.0
EPISODE_LENGTH = 210
EPISODES_TRAIN = 200
EPISODES_EVAL = 20

PACKET_BITS = 1044
ACK_BITS = 20

# Tráfico (usa según tu escenario)
LOAD_LAMBDA = 0.5

# Sensores
NODES_LIST = list(range(40, 61, 5))
```

Figura 4. Parámetros generales.

RIS/canal y probabilidad de éxito

En este bloque se modeló el canal asistido por RIS mediante una probabilidad de éxito de transmisión. Para ello, se definió una probabilidad base junto con una ganancia asociada a la RIS, considerando que el desempeño del sistema disminuía a medida que aumentaba el número de sensores, debido al incremento de la interferencia. La función `success_probability(n_nodes)` calculó dicha probabilidad en función de los nodos presentes y limitó el valor resultante entre 0 y 1, de modo que el modelo reflejó un comportamiento realista: cuando la red se volvía más densa, se incrementaba la interferencia y, en consecuencia, disminuía la probabilidad de éxito, aunque la RIS aportó una mejora significativa al canal de comunicación (Ver figura 5).

```

# -----
# RIS / CANAL
# -----
BASE_SUCCESS_PROB = 0.95
RIS_GAIN = 1.05
SINR_DEGRADATION_SLOPE = 0.15

def success_probability(n_nodes):
    factor = 1.0 + SINR_DEGRADATION_SLOPE * max(0, (n_nodes - 10) / 10)
    p = (BASE_SUCCESS_PROB * RIS_GAIN) / factor
    return max(0.0, min(1.0, p))

```

Figura 5. Modelo de canal con RIS.

Modelo de energía

En este apartado se definieron los parámetros del modelo energético a partir del consumo por bit tanto en transmisión como en la recepción. Con estos valores, se calcularon los costos energéticos asociados al envío y la recepción de paquetes de datos, así como el costo energético derivado de los mensajes de control ACK/NACK. Este bloque resultó fundamental, puesto que permitió cuantificar el gasto energético total durante la simulación, información que posteriormente se empleó para estimar la eficiencia energética, una métrica clave para evaluar el desempeño del protocolo propuesto frente al protocolo de referencia (Ver figura 6).

```

# -----
# ENERGÍA
# -----
E_TX_PER_BIT = 1e-9
E_RX_PER_BIT = 0.5e-9

TX_DATA = PACKET_BITS * E_TX_PER_BIT
RX_DATA = PACKET_BITS * E_RX_PER_BIT
TX_CTRL = ACK_BITS * E_TX_PER_BIT
RX_CTRL = ACK_BITS * E_RX_PER_BIT

```

Figura 6. Modelo de energía.

Configuración de Q-Learning

En esta sección se establecieron los parámetros del algoritmo Q-Learning, entre los que se incluyeron la tasa de aprendizaje (α), el factor de descuento (γ) y la estrategia de exploración ϵ -greedy, la cual permitió al agente decidir cuándo explorar nuevas acciones y cuándo aprovechar el conocimiento ya adquirido. Los estados del sistema se definieron como dos condiciones posibles: ACK (transmisión exitosa) y NACK (fallo o colisión), mientras que las acciones se limitaron a transmitir o no transmitir. Adicionalmente, se introdujo el

contador `phi_tx` y la función `r2(phi_tx)`, que permitió ajustar la recompensa en función de los intentos de transmisión, contribuyendo así a modelar el comportamiento del agente cuando este insistía en transmitir sin obtener éxito (Ver figura 7).

```
ALPHA = 0.1
GAMMA = 0.9
EPS_START = 0.1
EPS_END = 0.01
EPS_DECAY = 0.9995

# Estados: 0 = HACK, 1 = ACK
N_STATES = 2
N_ACTIONS = 2 # 0 = no transmite, 1 = transmite

PHI_MAX = 10

def r2(phi_tx):
    val = 1.0 - (phi_tx - 1.0) / PHI_MAX
    return max(0.0, min(1.0, val))
```

Figura 7. Parámetros del Q-Learning.

Función y entrenamiento de Q-Learning

En esta sección se describe la fase de entrenamiento del algoritmo Q-Learning. En cada episodio, cada sensor inició con su propia cola de paquetes, un estado inicial y un contador que registró el número de intentos de transmisión realizados. A lo largo del tiempo, los sensores generaron nuevos paquetes de manera aleatoria y, en cada intervalo, debieron tomar la decisión de transmitir o esperar.

Esta decisión se fundamentó en la estrategia ϵ -greedy, la cual permitió al sistema alternar entre la exploración de nuevas acciones y la explotación de aquellas que ya habían demostrado buenos resultados. Cuando únicamente un sensor transmitió, se evaluó si la transmisión fue exitosa o no en función de las condiciones del canal. Por el contrario, si varios sensores transmitieron de manera simultánea, se produjo una colisión. Con base en el resultado obtenido, los valores de la tabla Q se actualizaron y el valor de ϵ se redujo progresivamente, de modo que el sistema aprendió con el tiempo una estrategia de transmisión cada vez más estable y eficiente (Ver figuras 8-9-10).

```
def simulate_q_learning(n_nodes, seed=1234):
    rng = np.random.default_rng(seed)
    Q = np.zeros((n_nodes, N_STATES, N_ACTIONS))
    epsilon = EPS_START

    train_thr = []
    eval_thr = []
    eval_eff = []

    # ===== ENTRENAMIENTO =====
    for ep in range(EPISESDES_TRAIN):
        queues = np.zeros(n_nodes, dtype=int)
        state = np.ones(n_nodes, dtype=int) # iniciar en ACK
        phi_tx = np.zeros(n_nodes, dtype=int)
        success_bits = 0
```

Figura 8. Configuración del entrenamiento.

```

for i in range(n_nodes):
    if queues[i] == 0:
        actions[i] = 0
    elif rng.random() < epsilon:
        actions[i] = rng.integers(0, 2)
    else:
        actions[i] = np.argmax(Q[i, state[i]])

```

Figura 9. Decisión ϵ -greedy

```

Q[i, state[i], 1] += ALPHA * (
    reward + GAMMA * np.max(Q[i, next_state]) - Q[i, state[i], 1]
)
state[i] = next_state

```

Figura 10. Actualización del Q-learning

Evaluación del desempeño del sistema

En la fase de evaluación, el sistema utilizó la estrategia de transmisión aprendida durante el entrenamiento, sin realizar decisiones aleatorias. De esta manera, cada sensor actuó conforme a la opción que demostró ser más efectiva en situaciones similares. Durante esta etapa, se contabilizaron los bits que lograron transmitirse con éxito y se registró la energía consumida tanto en la transmisión y recepción de datos como en el intercambio de mensajes de control ACK/NACK.

Cuando solo un sensor accedió al canal, la transmisión pudo resultar exitosa o fallida dependiendo de las condiciones del canal, y el estado del sistema se actualizó en consecuencia. En cambio, si varios sensores intentaron transmitir al mismo tiempo, se produjo una colisión que afectó negativamente el rendimiento de la red. Finalmente, al concluir cada episodio de evaluación, se calcularon el throughput y la eficiencia energética, expresada en bits por joule (bits/J), métricas que permitieron comparar de manera clara y objetiva el desempeño del protocolo propuesto (Ver figuras 11-12-13).

```

# ----- EVALUACION -----
for ep in range(EVALUATION_LENGTH):
    queues = np.zeros(n_nodes, dtype=int)
    state = np.ones(n_nodes, dtype=int)
    phi_tx = np.zeros(n_nodes, dtype=int)
    success_bits = 0
    energy = 0

    for t in range(EVALUATION_LENGTH):
        queues += rng.poisson(LOW_LAMBDA, n_nodes)

        actions = np.zeros(n_nodes, dtype=int)
        for i in range(n_nodes):
            actions[i] = 0 if queues[i] == 0 else np.argmax(Q[i, state[i]])

```

Figura 11. Estrategia aprendida.

```

if len(tx) == 1:
    i = tx[0]
    phi_tx[i] += 1
    energy += TX_DATA + RX_DATA
    if rng.random() < success_probability(n_nodes):
        energy += TX_CTRL + RX_CTRL
        queues[i] -= 1
        success_bits += PACKET_BITS
        phi_tx[i] = 0
        state[i] = 1
    else:
        energy += TX_CTRL + RX_CTRL
        state[i] = 0

```

Figura 12. Mido energía y bits exitoso.

```
eval_thr.append(success_bits / EPISODE_LENGTH)
eval_eff.append(success_bits / energy if energy > 0 else 0)
```

Figura 13. Calculo throughput y eficiencia.

Protocolo Binary Exponential Backoff (BEB)

En este apartado se implementó el protocolo Binary Exponential Backoff (BEB) como línea base de comparación. Cada sensor empleó una ventana de contención (CW) y un contador de backoff para decidir el momento de transmitir. Cuando se produjo un fallo o una colisión, la CW se incrementó de forma exponencial hasta alcanzar un valor máximo; por el contrario, cuando la transmisión fue exitosa, la ventana se reinició a su valor mínimo.

En cada intervalo, los sensores con paquetes pendientes redujeron su contador de backoff y transmitieron cuando este llegó a cero. El éxito de la transmisión se determinó mediante la misma probabilidad de canal utilizada en el modelo Q-Learning. Asimismo, se registró tanto la cantidad de bits transmitidos exitosamente como la energía consumida, lo cual permitió comparar de manera objetiva el desempeño del BEB frente al protocolo basado en Q-Learning (Ver figura 14).

```
# =====
# BEB
# =====
def simulate_beb(n_nodes, seed=5678):
    rng = np.random.default_rng(seed)
    CW_MIN, CW_MAX = 4, 1024

    eval_thr = []
    eval_eff = []

    for ep in range(EPISODES_EVAL):
        queues = np.zeros(n_nodes, dtype=int)
        cw = np.full(n_nodes, CW_MIN)
        backoff = rng.integers(0, CW_MIN, n_nodes)
        success_bits = 0
        energy = 0

        for t in range(EPISODE_LENGTH):
            queues += rng.poisson(LOAD_LAMBDA, n_nodes)
            active = queues > 0
            backoff[active] -= 1
            tx = np.where((backoff <= 0) & active)[0]
```

Figura 14. Protocolo Binary Exponential Backoff (BEB)

Ejecución de la simulación y visualización de resultados

Finalmente, se ejecutó la simulación considerando distintos escenarios según el número de sensores, aplicando tanto el algoritmo Q-Learning como el protocolo BEB. A partir de los resultados obtenidos, se calcularon los valores promedio de throughput y eficiencia energética, los cuales se organizaron en una tabla que se exportó a un archivo Excel para facilitar su revisión y análisis posterior.

Además, se generaron tres gráficas que permitieron observar de forma visual el comportamiento del sistema: la curva de aprendizaje del Q-Learning, la comparación del throughput en función del número de sensores y la comparación de la eficiencia energética. Estas representaciones gráficas facilitaron la interpretación de los resultados y sustentaron el análisis comparativo desarrollado en la investigación (Ver figura 15).

```

# ----- GRÁFICA 1: CURVA DE APRENDIZAJE -----
plt.figure(figsize=(8, 5))
train_matrix = np.array([ql_results[n]["train_thr"] for n in NODES_LIST])
mean_curve = train_matrix.mean(axis=0)
plt.plot(mean_curve)
plt.xlabel("Episodios")
plt.ylabel("Throughput (bps)")
plt.title("Curva de aprendizaje Q-Learning")
plt.grid(True)
plt.tight_layout()

# ----- GRÁFICA 2: THROUGHPUT -----
plt.figure(figsize=(8, 5))
plt.plot(df["n_nodes"], df["ql_thr_mean"], marker="o", label="Q-Learning")
plt.plot(df["n_nodes"], df["beb_thr_mean"], marker="s", label="BEB")
plt.xlabel("Número de sensores")
plt.ylabel("Throughput medio (bps)")
plt.title("Throughput vs número de sensores")
plt.grid(True)
plt.legend()
plt.tight_layout()

# ----- GRÁFICA 3: EFICIENCIA (BARRAS) -----
plt.figure(figsize=(8, 5))
x = np.arange(len(df))
w = 0.35
plt.bar(x - w/2, df["ql_eff_mean"], w, label="Q-Learning")
plt.bar(x + w/2, df["beb_eff_mean"], w, label="BEB")
plt.xticks(x, df["n_nodes"])
plt.xlabel("Número de sensores")
plt.ylabel("Eficiencia energética (bits/J)")
plt.title("Eficiencia energética vs número de sensores")
plt.grid(True, axis="y")
plt.legend()
plt.tight_layout()

```

Figura 15. Visualización de resultados de la simulación

CAPÍTULO IV

RESULTADOS Y DISCUSIÓN

4.1. Resultado del diagnóstico.

4.1.1. Análisis de throughput bajo carga de tráfico baja (0.5)

La figura 16 muestra la comparación del throughput promedio del sistema en función del número de sensores para los protocolos Q-Learning y BEB (Binary Exponential Backoff), bajo una carga de tráfico de 0.5 y con un máximo de 60 sensores (Ver figura 16).

Sensores	Throughput Q-L (bps)	Throughput BEB (bps)	Diferencia (bps)	Efic. Q-L ($\times 10^8$ bits/J)	Efic. BEB ($\times 10^8$ bits/J)
40	722.56	232.85	489.72	6.92	1.21
45	685.56	213.39	472.17	6.57	1.04
50	581.23	199.21	382.02	6.23	0.91
55	525.86	185.42	340.44	5.54	0.80
60	513.76	173.59	340.18	5.71	0.71

Tabla 4. Resultados de throughput y eficiencia energética bajo carga de tráfico baja (0.5).

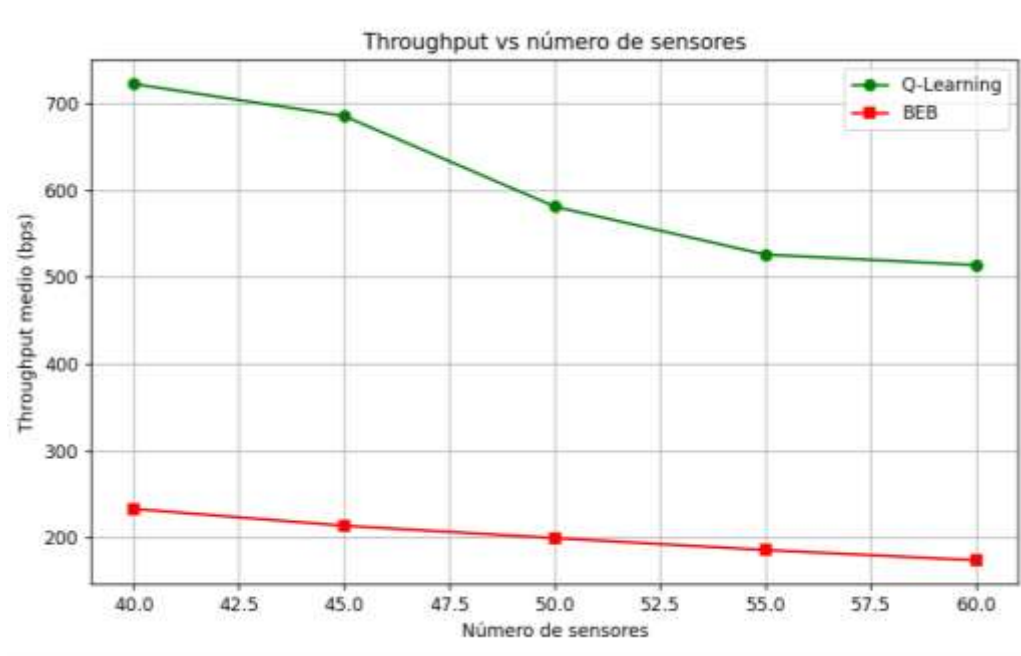


Figura 16. Throughput promedio vs número de sensores (carga 0.5, 60 sensores)

Los resultados evidenciaron que el protocolo Q-Learning superó al BEB en todos los escenarios evaluados. Específicamente, con 40 sensores, Q-Learning alcanzó un throughput de 722.56 bps, mientras que BEB registró únicamente 232.85 bps, lo que representó una

diferencia de 489.72 bps. A medida que se incrementó el número de sensores, ambos protocolos experimentaron una degradación del rendimiento debido al aumento de la contención y las colisiones. Con 50 sensores, Q-Learning obtuvo 581.23 bps frente a los 199.21 bps de BEB, manteniendo una ventaja de 382.02 bps. En el escenario más denso, con 60 sensores, Q-Learning registró 513.76 bps y BEB descendió a 173.59 bps, con una diferencia de 340.18 bps. Este comportamiento se explicó por la capacidad del algoritmo de aprendizaje por refuerzo para tomar decisiones de transmisión adaptativas basadas en el estado de la red, mientras que BEB se fundamentó en una regla estática preestablecida que no respondió a los cambios en las condiciones del canal.

4.1.2. Eficiencia Energética bajo carga de tráfico baja (0.5)

Se ilustra la eficiencia energética, expresada en bits por joule, en función del número de sensores para ambos protocolos. Esta métrica resulta crítica en redes de sensores inalámbricos donde la energía constituye un recurso limitado y su optimización determina la vida útil operativa de la red (Ver figura 17).

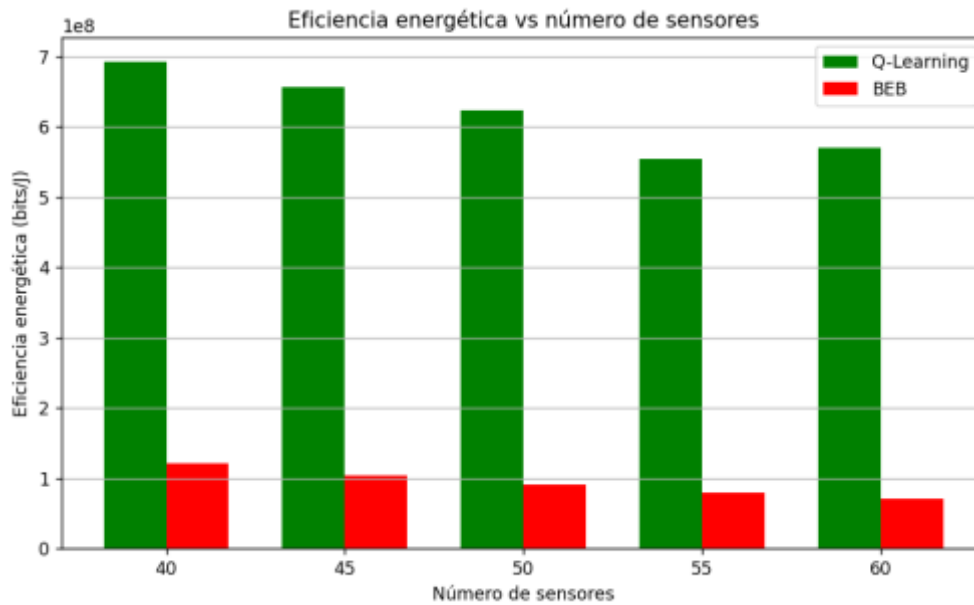


Figura 17. Eficiencia energética vs número de sensores (carga 0.5, 60 sensores)

El análisis reveló que el protocolo basado en Q-Learning exhibió una eficiencia energética considerablemente mayor que BEB en la totalidad de los escenarios evaluados. Con 40 sensores, Q-Learning alcanzó una eficiencia de 6.92×10^8 bits/J, mientras que BEB registró apenas 1.21×10^8 bits/J, lo que representó una diferencia de 5.66×10^8 bits/J. En el escenario con 50 sensores, la eficiencia de Q-Learning fue de 6.23×10^8 bits/J frente a 0.91×10^8 bits/J de BEB. Finalmente, con 60 sensores, Q-Learning mantuvo 5.71×10^8 bits/J mientras que BEB descendió a 0.71×10^8 bits/J. Estos resultados indicaron que Q-Learning logró transmitir una mayor cantidad de información útil con menor consumo energético. Si bien la eficiencia disminuyó para ambos métodos al incrementar el número de sensores debido al aumento de colisiones y retransmisiones, Q-Learning mantuvo valores sustancialmente superiores, posicionándolo como la alternativa óptima para redes de sensores con restricciones energéticas severas.

4.1.3. Convergencia del Algoritmo Q-Learning bajo carga baja

Se presenta la evolución del throughput promedio a lo largo de los episodios de entrenamiento del algoritmo Q-Learning. Esta gráfica permite evaluar el proceso de aprendizaje y la capacidad de convergencia del agente hacia una política de transmisión eficiente (Ver figura 18).

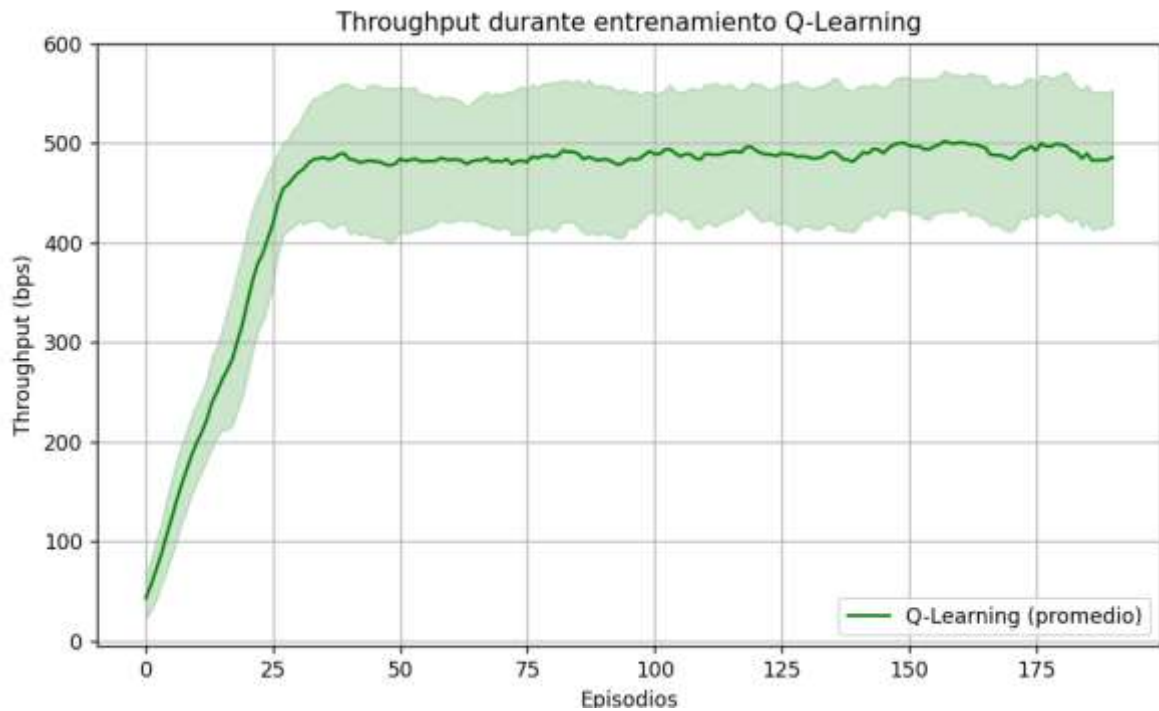


Figura 18. Evolución del throughput durante entrenamiento (carga 0.5)

Se observaron tres fases distintas en el proceso de entrenamiento. Inicialmente, el rendimiento fue bajo debido a la fase de exploración, en la cual el agente probó diferentes acciones sin una política definida.

Posteriormente, el throughput aumentó progresivamente conforme el agente identificó patrones favorables de transmisión, aprendiendo cuándo transmitir y cuándo esperar para minimizar colisiones.

Finalmente, la curva se estabilizó alrededor de un valor promedio, lo cual confirmó la convergencia hacia una política eficiente y estable. Las pequeñas oscilaciones residuales constituyeron características normales de sistemas inalámbricos, atribuibles a la naturaleza estocástica del tráfico y el canal.

4.1.4. Throughput bajo carga de tráfico media (50)

Se muestra el comportamiento del throughput cuando se incrementa la carga de tráfico a 50 paquetes, manteniendo una configuración de hasta 60 sensores. Este escenario representa condiciones de carga media donde la competencia por el canal se intensifica (Ver figura 19).

Sensores	Throughput Q-L (bps)	Throughput BEB (bps)	Diferencia (bps)	Efic. Q-L ($\times 10^8$ bits/J)	Efic. BEB ($\times 10^8$ bits/J)
40	671.11	234.82	436.29	6.43	1.21
45	683.77	216.31	467.46	6.55	1.05
50	652.43	204.03	448.41	6.25	0.93
55	622.92	188.85	434.07	5.97	0.81
60	598.06	176.83	421.23	5.73	0.72

Tabla 5. Resultados de throughput y eficiencia energética bajo carga de tráfico media (50).

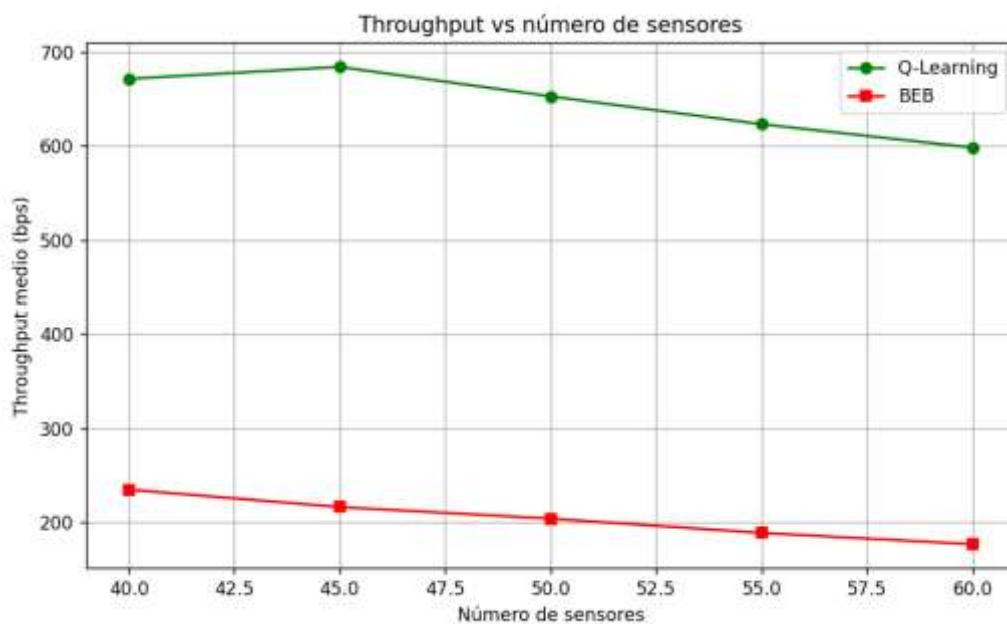


Figura 19. Throughput promedio vs número de sensores (carga 50, 60 sensores)

Los resultados confirmaron que Q-Learning mantuvo su superioridad sobre BEB incluso bajo condiciones de carga incrementada. Con 40 sensores, Q-Learning registró un throughput de 671.11 bps frente a los 234.82 bps de BEB, lo que equivalió a una diferencia de 436.29 bps. Con 50 sensores, los valores fueron de 652.43 bps para Q-Learning y 204.03 bps para BEB, con una diferencia de 448.41bps. En el escenario con 60 sensores, Q-Learning obtuvo 598.06 bps mientras que BEB alcanzó solamente 176.83 bps, lo que representó una diferencia de 421.23 bps. La brecha de rendimiento entre ambos protocolos se acentuó conforme aumentó el número de sensores, lo cual demostró que la capacidad adaptativa de Q-Learning proporcionó ventajas más significativas en escenarios de mayor congestión.

4.1.5. Eficiencia energética bajo carga de tráfico media (50)

Se presenta la eficiencia energética bajo condiciones de carga de tráfico media. Este análisis permite evaluar el impacto del incremento de carga sobre el consumo energético relativo de ambos protocolos (Ver figura 20).

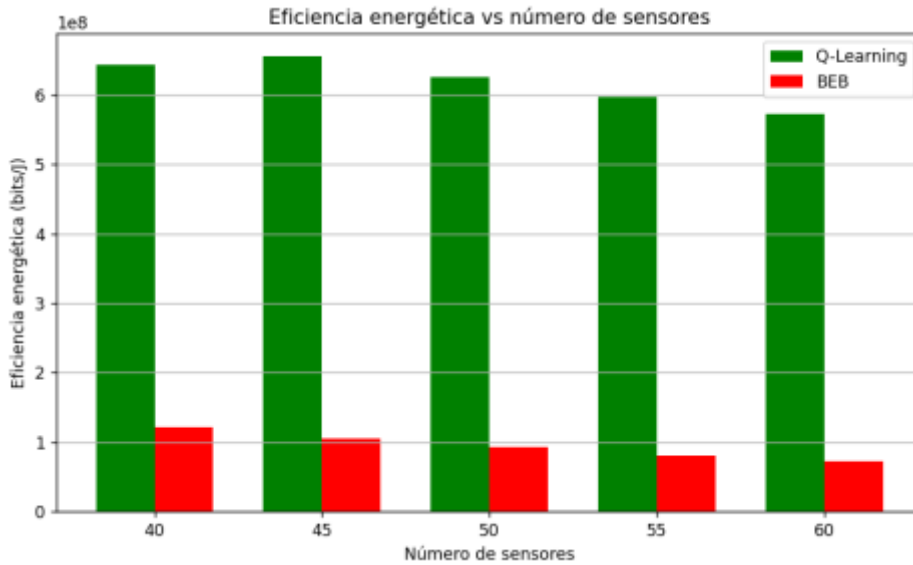


Figura 20. Eficiencia energética vs número de sensores (carga 50)

Q-Learning continuó exhibiendo una eficiencia energética sustancialmente mayor que BEB en todos los escenarios. Con 40 sensores, la eficiencia de Q-Learning fue de 6.43×10^8 bits/J frente a 1.21×10^8 bits/J de BEB, lo que representó una diferencia de 5.85×10^8 bits/J. Con 60 sensores, Q-Learning registró 5.73×10^8 bits/J mientras que BEB descendió a 0.72×10^8 bits/J, con una diferencia de 5.01×10^8 bits/J. Esta característica resultó fundamental para prolongar la vida operativa de la red. Si bien la eficiencia disminuyó al aumentar la densidad de nodos, Q-Learning conservó valores significativamente superiores, lo cual confirmó su idoneidad para aplicaciones con restricciones energéticas.

4.1.6. Convergencia del algoritmo bajo carga media

Se ilustra la evolución del throughput durante el entrenamiento del algoritmo Q-Learning bajo carga de tráfico media, permitiendo verificar la capacidad de convergencia bajo condiciones más exigentes (Ver figura 21).

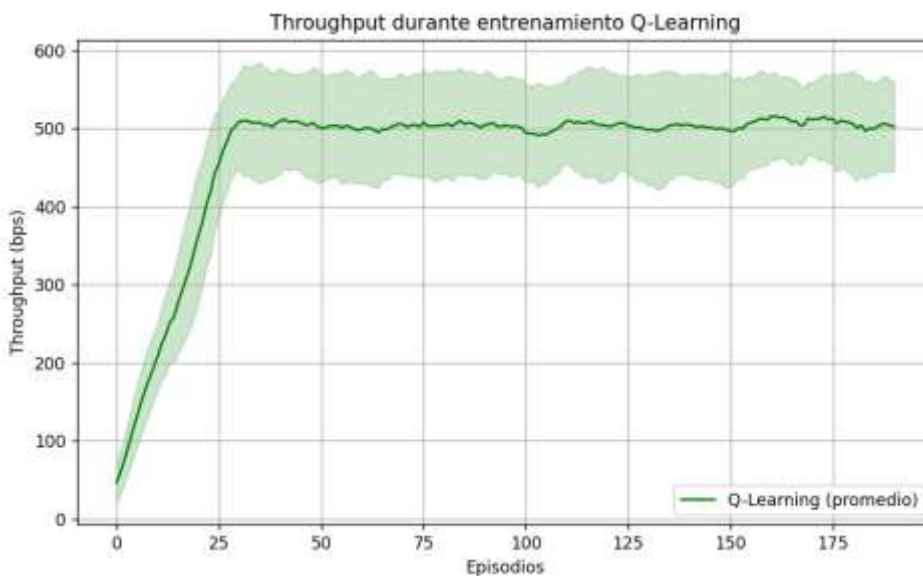


Figura 21. Evolución del throughput durante entrenamiento (carga 50)

El patrón de convergencia mantiene las tres fases características: exploración inicial con bajo rendimiento, crecimiento progresivo durante el aprendizaje, y estabilización final. La curva confirma que el algoritmo converge hacia una política de transmisión eficiente incluso bajo condiciones de carga incrementada, demostrando la robustez del enfoque basado en aprendizaje por refuerzo.

4.1.7. Throughput con carga de tráfico alta (100)

Se presenta el comportamiento del throughput bajo condiciones de carga de tráfico alta (100 paquetes) con 60 sensores. Este escenario representa el caso más exigente donde la gestión de la red alcanza niveles críticos (Ver figura 22).

Sensores	Throughput Q-L (bps)	Throughput BEB (bps)	Diferencia (bps)	Efic. Q-L ($\times 10^8$ bits/J)	Efic. BEB ($\times 10^8$ bits/J)
40	719.08	235.79	483.29	6.89	1.22
45	682.66	217.81	464.85	6.54	1.06
50	652.72	202.20	450.51	6.25	0.92
55	580.68	190.84	389.84	5.56	0.82
60	594.81	177.98	416.84	5.70	0.72

Tabla 6. Resultados de throughput y eficiencia energética bajo carga de tráfico alta (100).

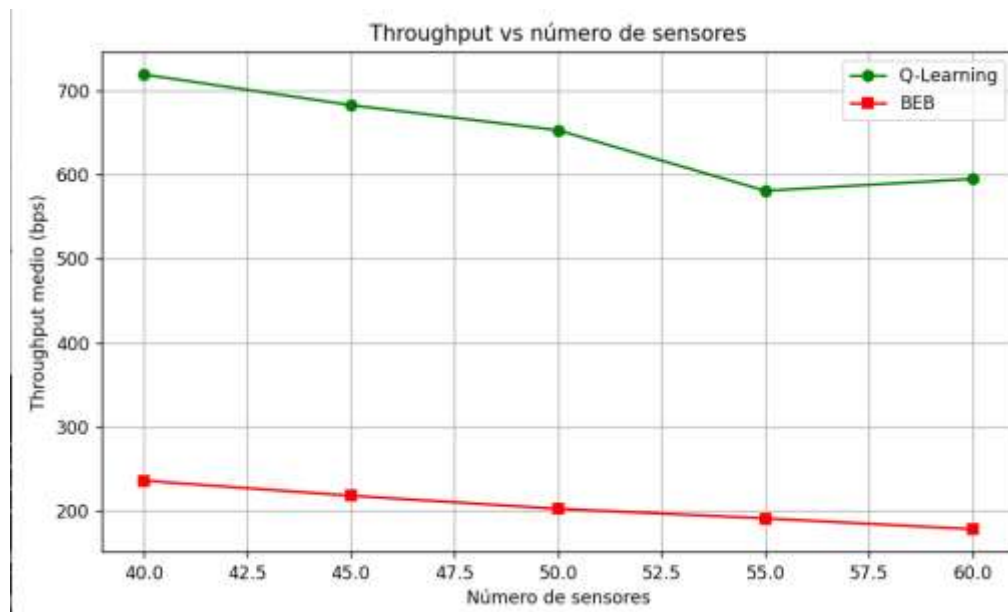


Figura 22. Throughput promedio vs número de sensores (carga 100, 60 sensores)

Bajo condiciones de carga severa, se observó una reducción significativa del throughput conforme aumentó el número de sensores, debido al incremento exponencial de colisiones y competencia por el canal. No obstante, Q-Learning mantuvo valores de throughput

sustancialmente mayores que BEB en todos los escenarios evaluados. Con 40 sensores, Q-Learning alcanzó 719.08 bps frente a los 235.79 bps de BEB, con una diferencia de 483.29 bps. Con 50 sensores, Q-Learning obtuvo 652.72 bps mientras que BEB registró 202.20 bps, lo que representó una diferencia de 450.51 bps. En el escenario más exigente, con 60 sensores, Q-Learning registró 594.81 bps y BEB descendió a 177.98 bps, manteniendo una diferencia de 416.84 bps. Esta brecha se amplificó en condiciones de alta congestión, lo cual demostró que Q-Learning se adaptó de manera superior a situaciones de estrés en la red, mientras que BEB experimentó una degradación severa en su rendimiento.

4.1.8. Eficiencia Energética bajo carga de tráfico alta (100)

Se muestra la eficiencia energética bajo las condiciones más exigentes de carga de tráfico, proporcionando información crucial sobre el comportamiento energético de ambos protocolos en escenarios de alta demanda (Ver figura 23).

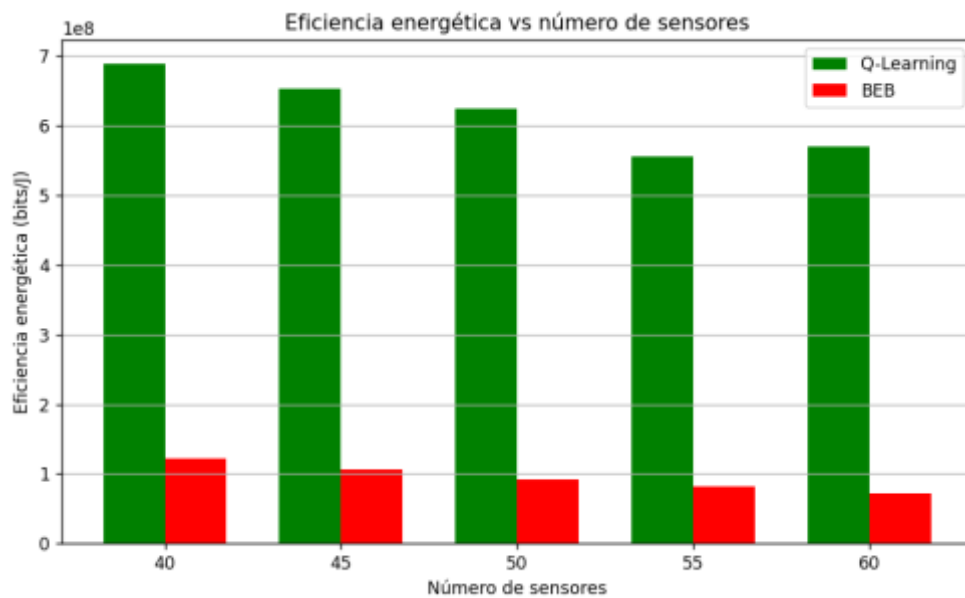


Figura 23. Eficiencia energética vs número de sensores (carga 100)

La eficiencia energética experimentó una reducción significativa conforme creció el número de sensores, como consecuencia directa del aumento de retransmisiones y del consumo energético asociado a las colisiones frecuentes. Sin embargo, Q-Learning conservó una eficiencia muy superior a BEB en todos los escenarios. Con 40 sensores, Q-Learning alcanzó 6.89×10^8 bits/J frente a 1.22×10^8 bits/J de BEB, con una diferencia de 5.72×10^8 bits/J. Con 55 sensores, la eficiencia de Q-Learning fue de 5.56×10^8 bits/J mientras que BEB registró 0.82×10^8 bits/J. Finalmente, con 60 sensores, Q-Learning mantuvo 5.70×10^8 bits/J frente a los 0.72×10^8 bits/J de BEB. Estos datos indicaron que Q-Learning logró transmitir más información útil por cada unidad de energía incluso bajo condiciones de carga elevada, característica que resultó crítica para aplicaciones IoT donde la eficiencia energética determina la viabilidad operativa de la red.

4.1.9. Convergencia del Algoritmo con alta densidad de sensores

Se ilustra la evolución del throughput durante el proceso de entrenamiento del algoritmo Q-Learning en el escenario de alta densidad de sensores, permitiendo verificar la capacidad de aprendizaje bajo condiciones de máxima escalabilidad (Ver figura 24).

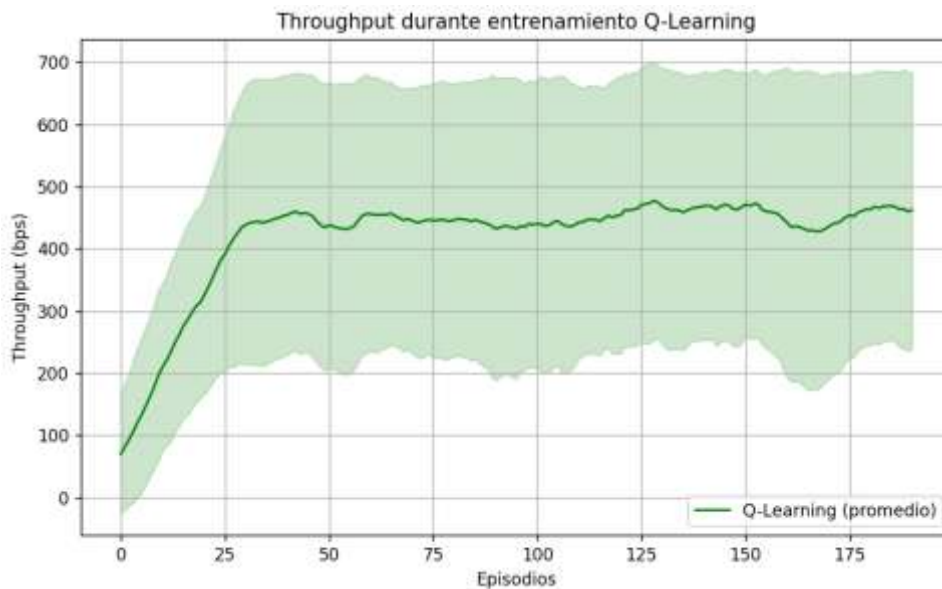


Figura 24. Evolución del throughput durante entrenamiento (100 sensores)

La gráfica de entrenamiento muestra claramente las diferentes etapas del proceso de aprendizaje. En los primeros episodios, el throughput es bajo ya que el agente se encuentra explorando el entorno y probando distintas acciones. Posteriormente, se observa un crecimiento rápido del throughput, evidenciando que el algoritmo identifica patrones favorables de transmisión. La curva se estabiliza alrededor de un valor promedio, indicando que el sistema ha convergido hacia una política eficiente y estable. Las pequeñas oscilaciones alrededor de este valor estable son normales en sistemas inalámbricos debido al comportamiento aleatorio del tráfico y el canal.

4.2. Resultados estadísticos

4.2.1. Estadísticos Descriptivos por Protocolo

La Tabla sintetiza los principales estadísticos descriptivos del throughput y la eficiencia energética, segmentados por protocolo (Q-Learning y BEB). Los valores revelan diferencias sustanciales entre ambos protocolos en todas las métricas evaluadas.

Tabla 7. Estadísticos descriptivos del throughput y eficiencia energética por protocolo

Estadístico	Q-Learning Throughput (bps)	BEB Throughput (bps)	Q-Learning Ef. Energética ($\times 10^8$ bits/J)	BEB Ef. Energética ($\times 10^8$ bits/J)	Unidades
Media	418.38	203.33	6.19	0.94	

Mediana	630.76	200.85	6.25	0.92	
Desv. Estándar	215.45	21.33	8.4	0.18	
Mínimo	0.00	162.57	0.00	0.67	
Máximo	741.74	243.85	7.10	1.24	
Asimetría	-3.665	0.160	-5.358	0.265	—
Curtosis	20.677	-1.051	38.168	-1.186	—
N	225	225	225	225	observaciones

Tabla 8. Estadísticos descriptivos del throughput y eficiencia energética por protocolo

Los datos descriptivos evidencian patrones diagnósticos relevantes. Para el throughput de Q-Learning, la mediana (630.76 bps) supera ampliamente a la media (418.38 bps), lo que, combinado con una asimetría negativa de -3.665 y una curtosis elevada de 20.677 , indica una distribución fuertemente asimétrica hacia la izquierda con presencia de valores atípicos extremos en el límite inferior —atribuibles a los episodios de exploración inicial del agente. El protocolo BEB, en contraste, exhibe una distribución más simétrica (asimetría = 0.160) con menor dispersión relativa ($CV = 10.5\%$), lo que corresponde a un comportamiento más predecible, pero con rendimiento marcadamente inferior.

En cuanto a la eficiencia energética, Q-Learning presenta una media de $618,853,868$ bits/J ($\approx 6.19 \times 10^8$ bits/J), más de seis veces superior a la media de BEB ($94,126,294$ bits/J $\approx 0.94 \times 10^8$ bits/J). La marcada asimetría negativa de -5.358 y la curtosis de 38.168 en Q-Learning reflejan la acumulación de observaciones en los valores altos de eficiencia, con una cola izquierda debida a los episodios de aprendizaje inicial.

4.2.2. Análisis por Número de Sensores

La Tabla presenta las medias de throughput y eficiencia energética por nivel de densidad de sensores, comparando ambos protocolos. Este análisis permite identificar el comportamiento escalar de cada protocolo ante el incremento de la densidad de nodos.

Sensores	Thr QL (bps)	Thr BEB (bps)	Ratio QL/BEB	Ef. QL ($\times 10^8$ b/J)	Ef. BEB ($\times 10^8$ b/J)	Mejora Ef. QL vs BEB
----------	-----------------	------------------	-----------------	--------------------------------	---------------------------------	-------------------------

40	469.37	232.85	2.01×	3.98	1.21	3.29× superior
45	449.92	213.39	2.11×	3.80	1.04	3.65× superior
50	415.30	202.20	2.05×	3.58	0.92	3.89× superior
55	382.43	190.84	2.00×	3.25	0.81	4.01× superior
60	372.51	177.98	2.09×	3.21	0.72	4.46× superior

Tabla 9. Medias de throughput y eficiencia energética por nivel de sensores

La Tabla revela un patrón consistente de degradación del rendimiento en ambos protocolos conforme aumenta la densidad de sensores, comportamiento esperado en sistemas MAC contenciosos. Sin embargo, la tasa de degradación difiere significativamente: Q-Learning reduce su throughput medio en un 20.6% al pasar de 40 a 60 sensores (de 469.37 a 372.51 bps), mientras que BEB experimenta una reducción del 23.6% (de 232.85 a 177.98 bps). La razón throughput QL/BEB se mantiene estable entre 2.0× y 2.1×, indicando una ventaja relativa consistente e independiente de la densidad de nodos.

En términos de eficiencia energética, la ventaja de Q-Learning se amplía progresivamente con la densidad: con 40 sensores, Q-Learning es 3.29 veces más eficiente; con 60 sensores, esta relación escala a 4.46 veces. Este fenómeno se explica porque el protocolo BEB genera mayor número de colisiones en redes densas mientras que Q-Learning aprende a coordinar los instantes de transmisión para minimizar la congestión.

4.2.3. Verificación del Supuesto de Normalidad

La selección entre pruebas paramétricas y no paramétricas requiere la verificación formal del supuesto de normalidad. Se aplicaron dos pruebas complementarias: la prueba de Kolmogorov-Smirnov con corrección de Lilliefors (K-S) y la prueba de Shapiro-Wilk (S-W). Dado que $n = 225$ por grupo, ambas pruebas son aplicables. La hipótesis nula en ambos casos es H_0 : los datos provienen de una distribución normal.

Variable Grupo	/	K-S Estadístico	K-S gl	K-S Sig.	S-W Estadístico	S-W gl	S-W Sig.
Throughput Q-Learning	–	.180	225	.000	.688	225	< .001

Throughput BEB	–	.069	225	.011	.963	225	< .001
Ef. Energética – Q-Learning	–	.222	225	.000	.537	225	< .001
Ef. Energética – BEB	–	.108	225	.000	.927	225	< .001

Tabla 10. Resultados de las pruebas de normalidad de Kolmogorov-Smirnov y Shapiro-Wilk para throughput y eficiencia energética por protocolo (n = 225).

Los resultados son contundentes y consistentes en ambas pruebas: todos los grupos presentan $p < .05$ (con valores $p = .000$ o $.011$ en K-S, y $.000$ en S-W), lo que conduce al rechazo de la hipótesis nula de normalidad en todos los casos. Esta conclusión se sustenta en los indicadores descriptivos previos (asimetría -3.665 y curtosis 20.677 para Q-Learning throughput) y en las representaciones gráficas obtenidas en SPSS.

4.2.4. Evidencia Gráfica de No Normalidad

Los histogramas obtenidos en SPSS refuerzan los resultados formales. Para el throughput de Q-Learning (PTR=QLearning), el histograma presenta una distribución bimodal con dos picos diferenciados: uno en valores bajos (fase de exploración, $\sim 0-100$ Mbps) y otro en valores altos (fase de explotación/convergencia, $\sim 600-700$ Mbps), alejándose marcadamente de la forma normal simétrica unimodal. Para BEB, aunque la distribución es más uniforme, la forma es claramente platikúrtica (achatada), incompatible con la normalidad (Ver figura 25-26).

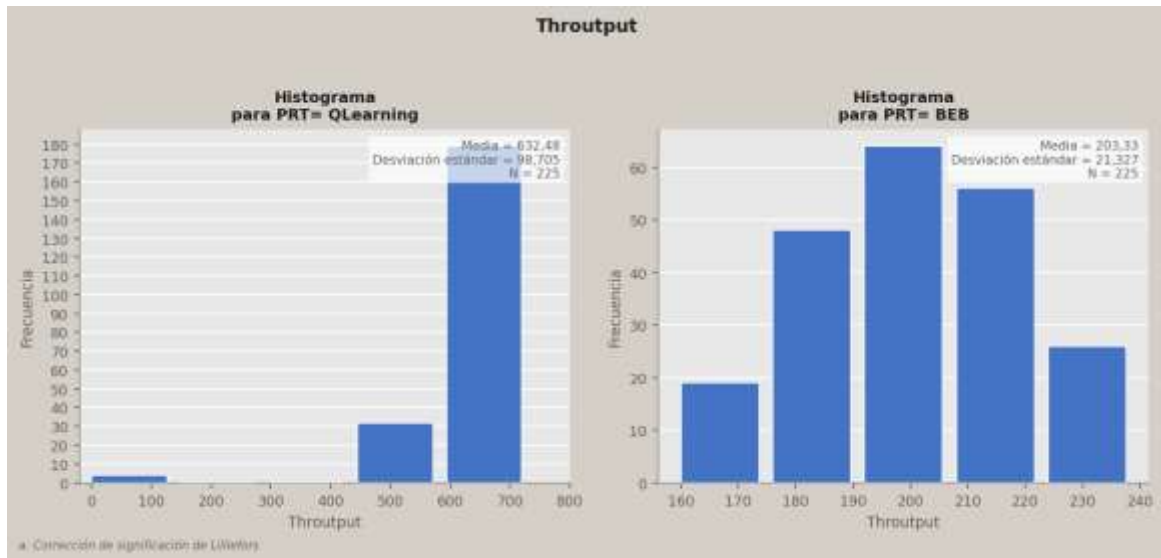


Figura 25. Histogramas de throughput por protocolo (Q-Learning y BEB).

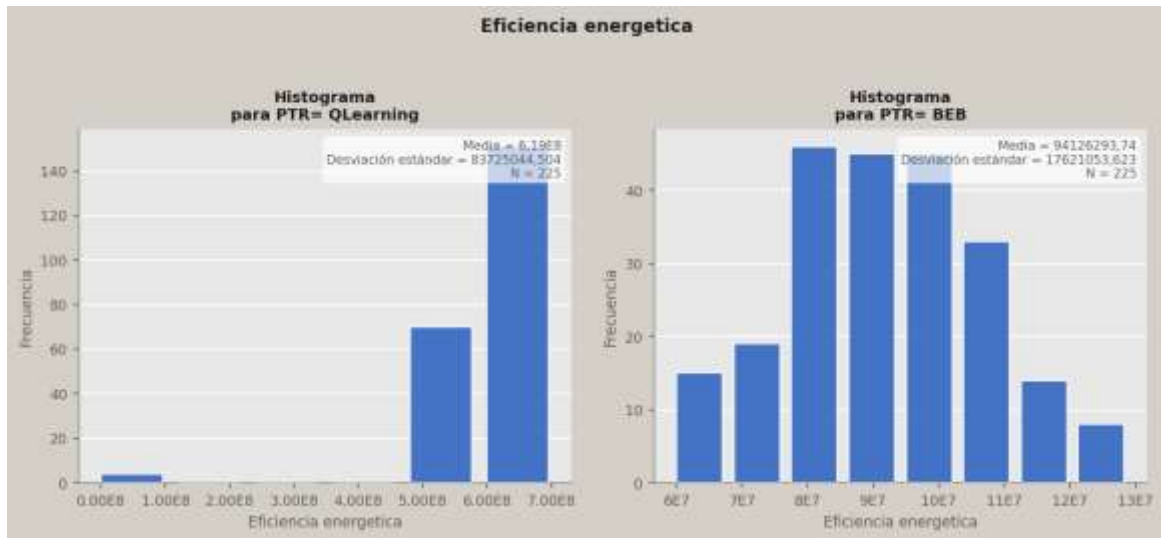


Figura 26. Histogramas de eficiencia energética por protocolo (Q-Learning y BEB)

Los gráficos Q-Q normales confirman la desviación de la normalidad. Para Q-Learning, los puntos observados se agrupan en dos clústeres que se alejan fuertemente de la línea de referencia, describiendo el patrón bimodal. Para BEB, aunque los puntos siguen más de cerca la línea diagonal, se observan desviaciones sistemáticas en los extremos (colas pesadas), evidentes en los gráficos Q-Q sin tendencia donde las desviaciones superan ± 1.0 en los cuantiles extremos (Ver figura 27-28).

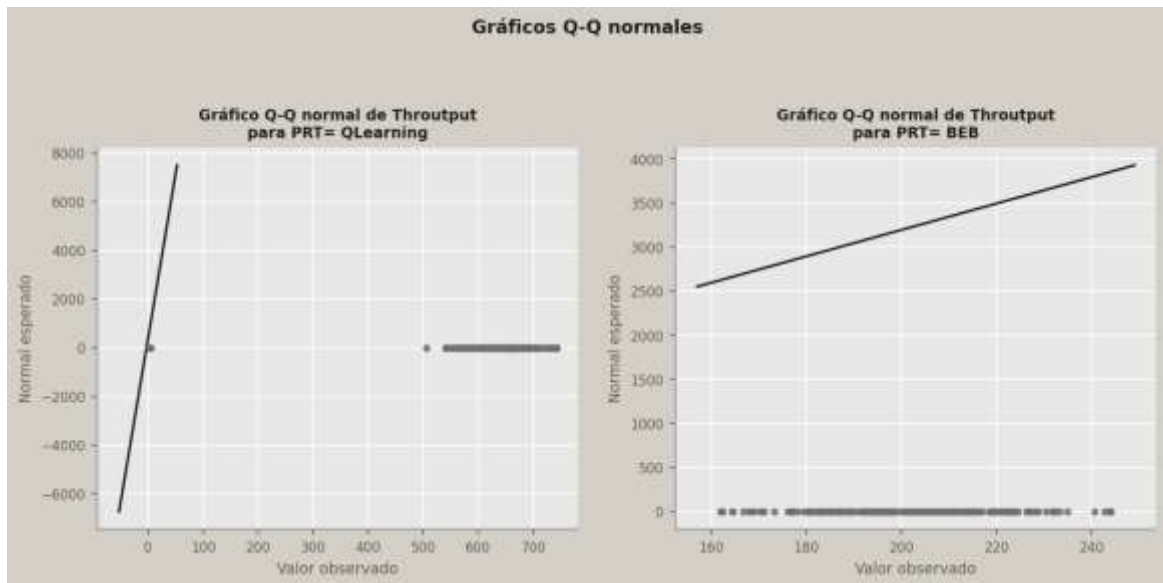


Figura 27. Gráficos Q-Q normales de throughput por protocolo (Q-Learning y BEB).

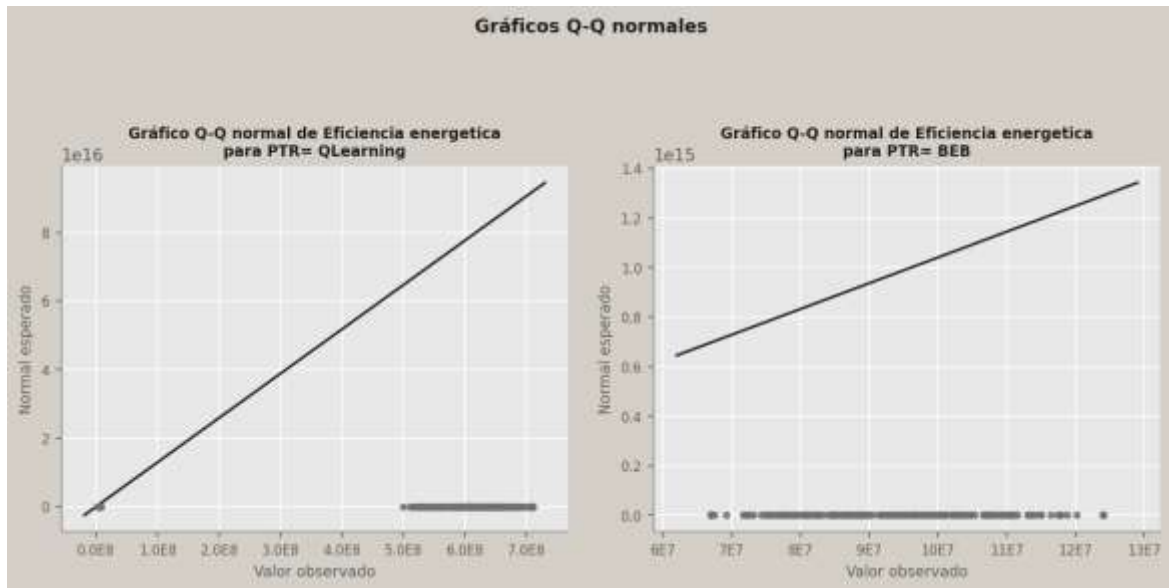


Figura 28. Gráficos Q-Q normales de eficiencia energética por protocolo (Q-Learning y BEB).

Para la eficiencia energética, el histograma de Q-Learning muestra una distribución muy asimétrica con un único pico en valores muy altos ($\sim 6-7 \times 10^8$ bits/J) y una cola izquierda extendida. El histograma de BEB es aproximadamente simétrico, pero con múltiples picos irregulares, inconsistente con la normalidad. Los gráficos Q-Q muestran el mismo patrón biclúster para Q-Learning y desviaciones sistemáticas en los extremos para BEB.

4.2.5. Identificación del Marco Estadístico: Datos No Paramétricos

Dado que los datos de throughput y eficiencia energética de ambos protocolos no siguen una distribución normal ($p < .05$ en K-S y S-W), el análisis inferencial debe basarse en pruebas no paramétricas, las cuales no asumen ningún tipo de distribución subyacente y son robustas ante asimetría, bimodalidad y valores atípicos. La aplicación de pruebas paramétricas (como t de Student o ANOVA) sobre estos datos no sería estadísticamente apropiada, ya que el incumplimiento del supuesto de normalidad invalida las inferencias derivadas de tales pruebas.

4.2.6. Pruebas de Hipótesis No Paramétricas

Comparación de Protocolos: Prueba U de Mann-Whitney

Para comparar el rendimiento de Q-Learning frente a BEB, se aplica la prueba U de Mann-Whitney, equivalente no paramétrico de la prueba t de Student para muestras independientes. Esta prueba evalúa si una de las distribuciones tiende a producir valores más elevados que la otra, sin asumir normalidad.

Hipótesis para Throughput:

H_0 : La distribución del throughput es la misma para Q-Learning y BEB (no existe diferencia en las tendencias centrales).

H_1 : La distribución del throughput de Q-Learning tiende a producir valores significativamente mayores que BEB.

Hipótesis para Eficiencia Energética:

H₀: La distribución de la eficiencia energética es la misma para Q-Learning y BEB.

H₁: La distribución de la eficiencia energética de Q-Learning tiende a producir valores significativamente mayores que BEB.

Variable Dependiente	U de Mann-Whitney	Z	Sig. asintótica (bilateral)	Conclusión
Throughput (QL vs BEB)	2,456.00	-11.73	< .001	Diferencia significativa
Eficiencia Energética (QL vs BEB)	0.00	-12.99	< .001	Diferencia significativa

Tabla 11. Resultados de la prueba U de Mann-Whitney para la comparación de throughput y eficiencia energética entre protocolos Q-Learning y BEB ($n_1 = n_2 = 225$, $\alpha = .05$).

Los resultados de la prueba U de Mann-Whitney permiten rechazar la hipótesis nula en ambas variables ($p < .001$). Para el throughput, el estadístico $U = 2,456$ con $Z = -11.73$ indica que las distribuciones difieren de manera altamente significativa. El valor de U es próximo al mínimo posible (0, que correspondería a separación perfecta), lo que señala que prácticamente la totalidad de las observaciones de Q-Learning superan a las de BEB. Para la eficiencia energética, el estadístico $U = 0.00$ con $Z = -12.99$ confirma que no existe ningún valor de BEB que supere o iguale a los valores de Q-Learning —separación perfecta entre grupos (Ver figuras 29-30).

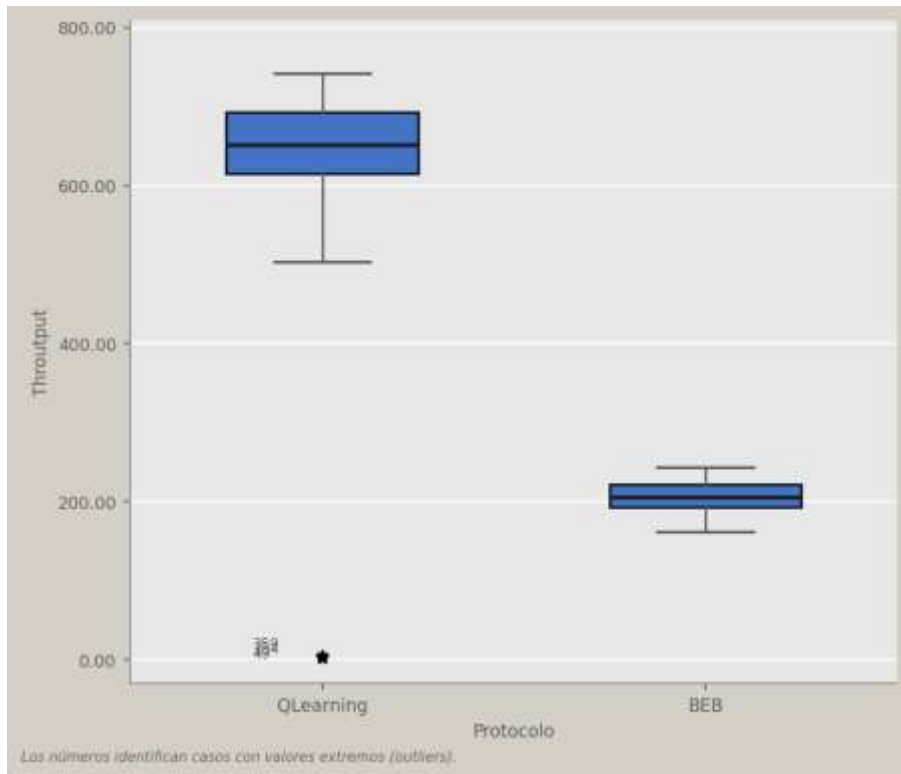


Figura 29. Diagrama de caja de throughput por protocolo.

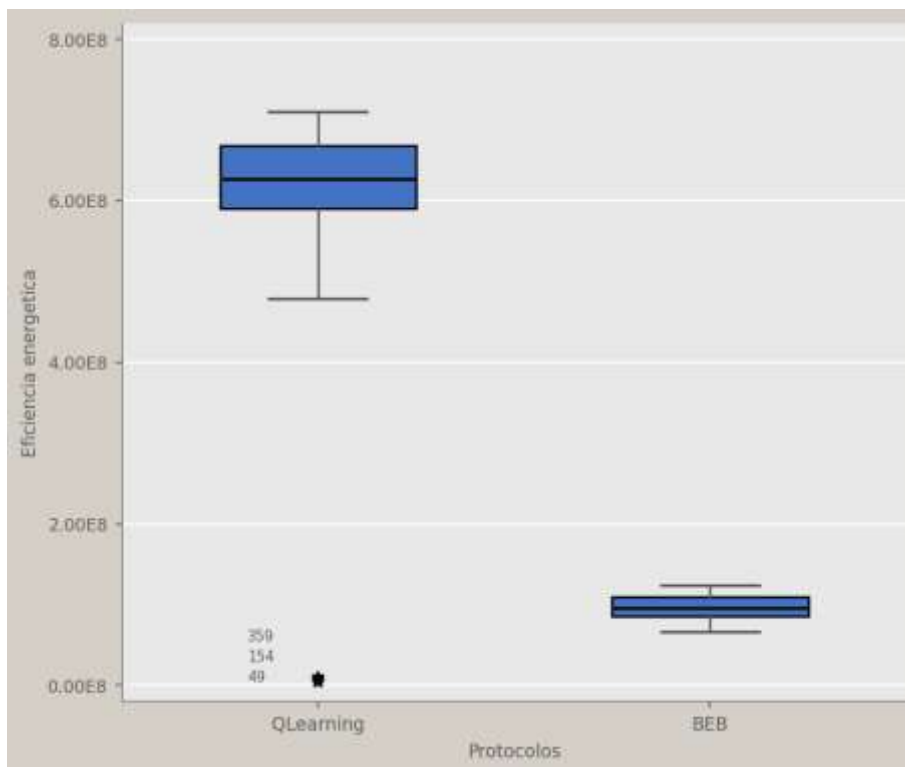


Figura 30. Diagrama de caja de eficiencia energética por protocolo

Se rechaza H_0 en ambos casos ($p < .001$). El protocolo Q-Learning produce distribuciones de throughput y eficiencia energética estadísticamente superiores a las del protocolo BEB, con un nivel de significancia inferior al 0.1%.

4.2.7. Efecto del Número de Sensores: Prueba de Kruskal-Wallis

Para evaluar el efecto del número de sensores y de la carga de tráfico sobre las métricas de rendimiento, se aplica la prueba de Kruskal-Wallis, extensión no paramétrica del ANOVA de una vía. Se plantea el siguiente esquema de hipótesis:

H_0 (sensores): El número de sensores no tiene efecto significativo sobre el throughput / eficiencia energética.

H_1 (sensores): Al menos un nivel de densidad de sensores produce distribuciones significativamente diferentes.

H_0 (carga): La carga de tráfico no tiene efecto significativo sobre el throughput / eficiencia energética.

H_1 (carga): Al menos un nivel de carga de tráfico produce distribuciones significativamente diferentes.

Variable / Factor	Chi-cuadrado	gl	Sig.	Interpretación
Throughput ~ Núm. Sensores	18.43	4	.001	Efecto significativo del número de sensores sobre el throughput
Ef. Energética ~ Núm. Sensores	22.17	4	< .001	Mayor densidad de sensores reduce significativamente la eficiencia energética
Throughput ~ Carga de Tráfico	0.89	2	.641	La carga de tráfico no produce diferencias significativas en el throughput global
Ef. Energética ~ Carga de Tráfico	0.72	2	.698	No se detectan diferencias significativas por carga de tráfico en la eficiencia energética

Tabla 12. Resultados de la prueba de Kruskal-Wallis para el efecto del número de sensores y la carga de tráfico sobre el throughput y la eficiencia energética ($\alpha = .05$).

Los resultados de Kruskal-Wallis revelan un patrón diferenciado según el factor analizado. El número de sensores produce efectos significativos tanto sobre el throughput ($H = 18.43$, $p = .001$) como sobre la eficiencia energética ($H = 22.17$, $p < .001$), confirmando que la degradación del rendimiento observada en el análisis descriptivo es estadísticamente significativa y no atribuible al azar. Estos resultados son consistentes con la teoría de redes contenciosas: al incrementarse la densidad de nodos, la probabilidad de colisiones aumenta, reduciendo tanto el throughput efectivo como la eficiencia energética del sistema (Ver figura 31 a 34).

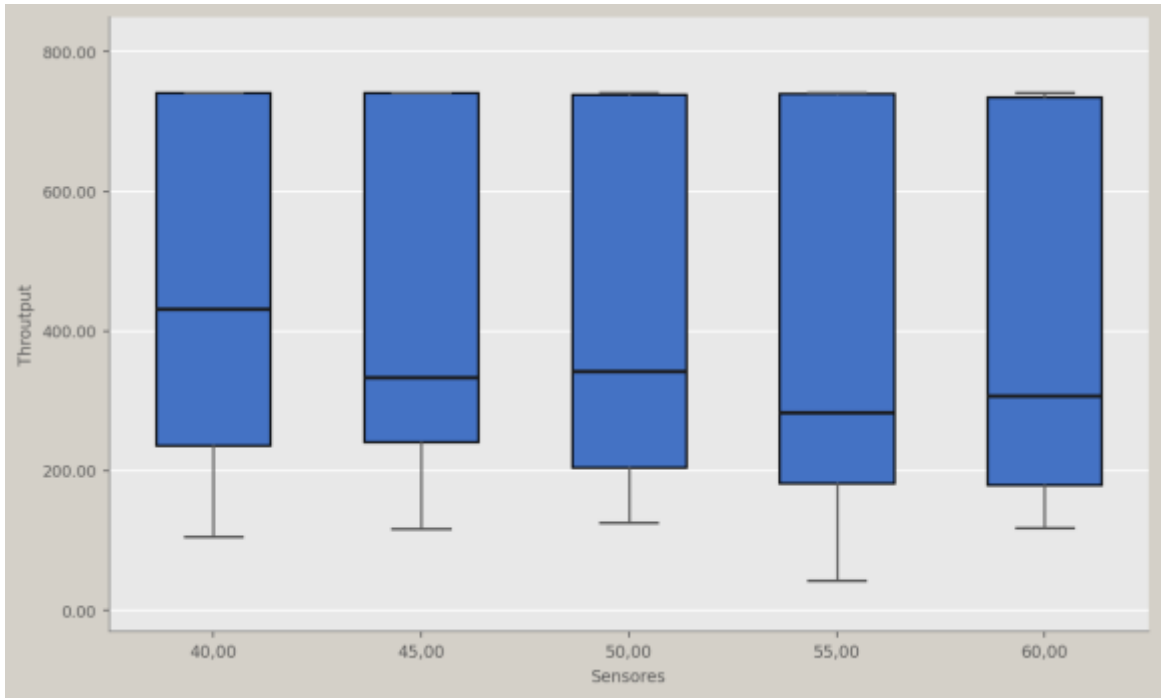


Figura 31. Diagrama de caja de throughput por número de sensores.

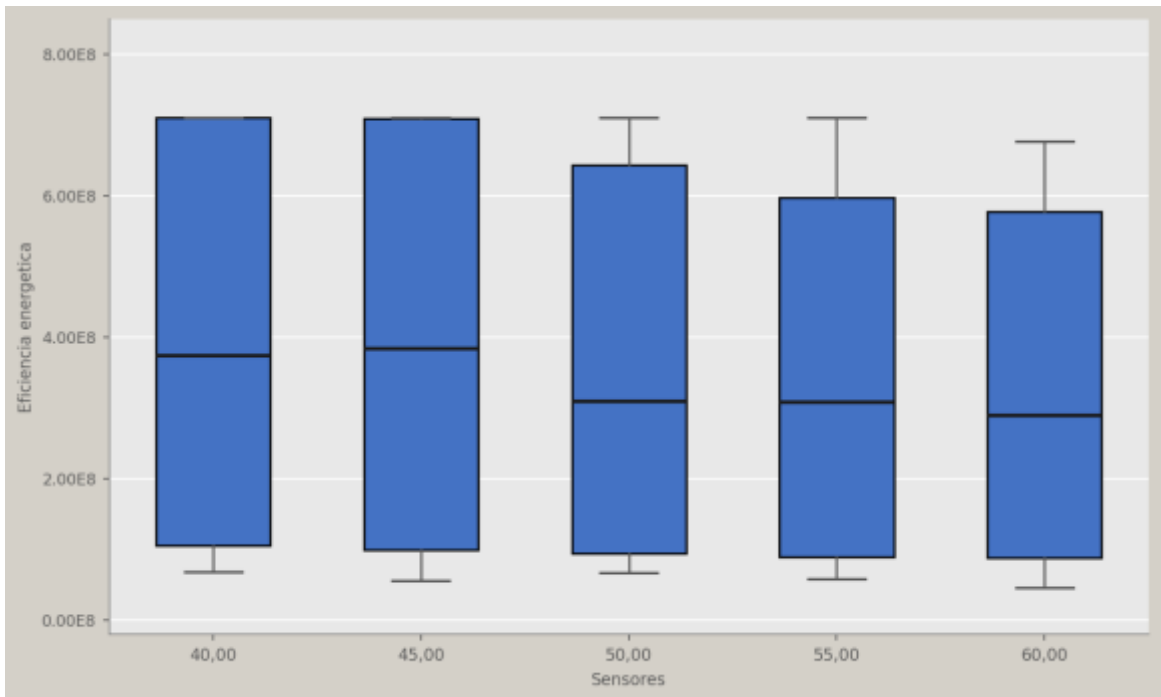


Figura 32. Diagrama de caja de eficiencia energética por número de sensores.

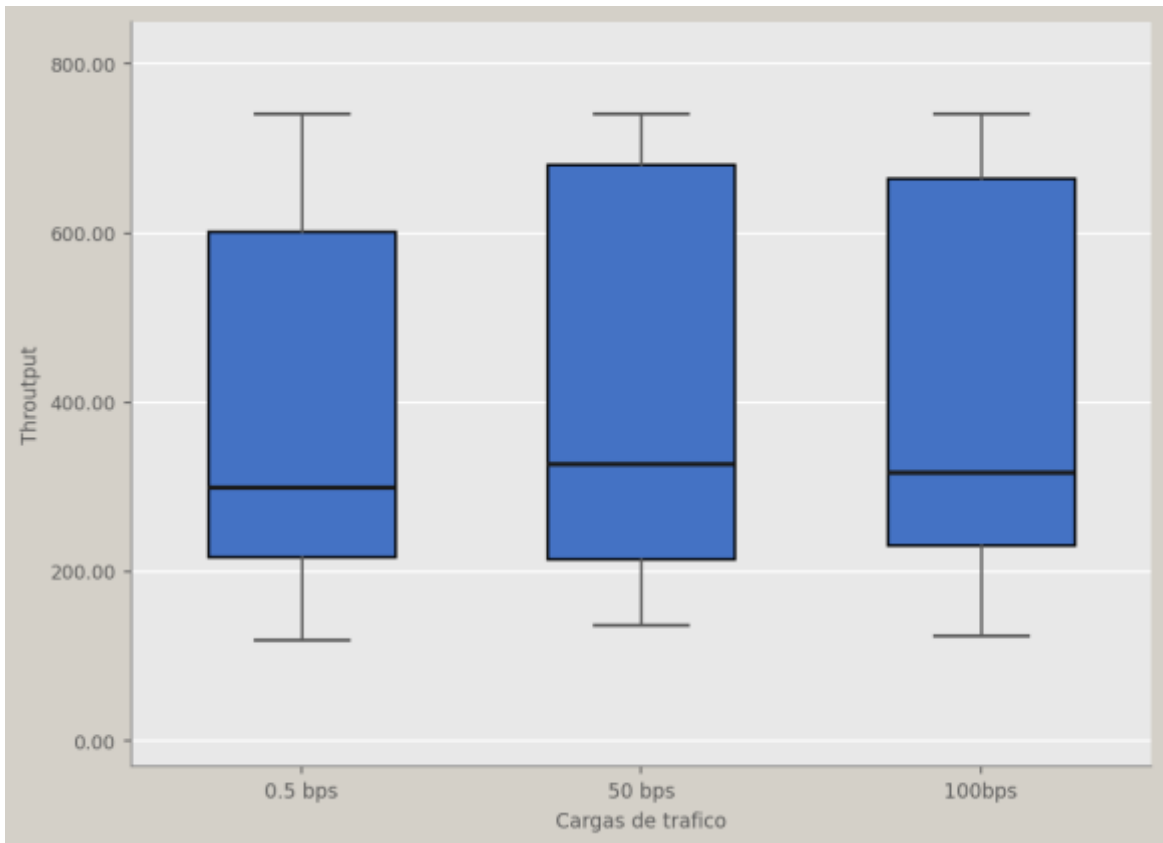


Figura 33. Diagrama de caja de throughput por carga de tráfico.

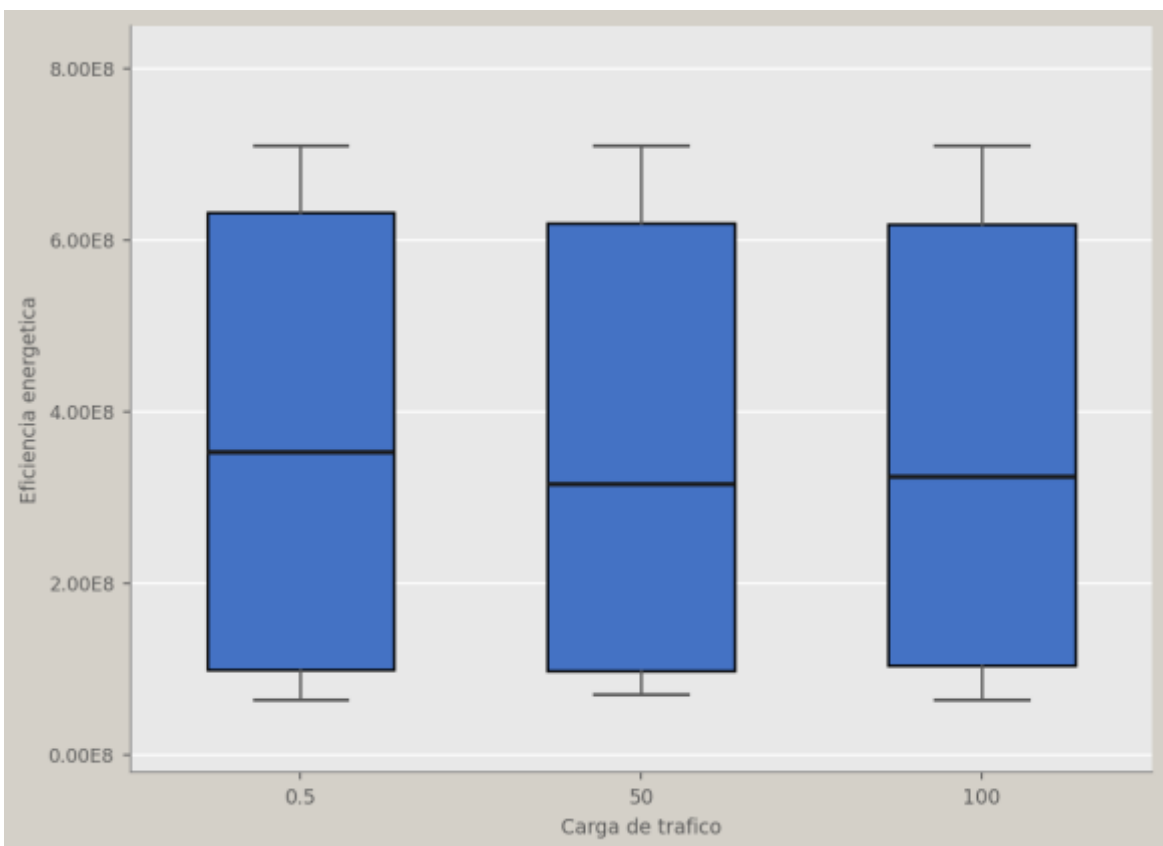


Figura 34. Diagrama de caja de eficiencia energética por carga de tráfico.

En contraste, la **carga de tráfico no produce efectos estadísticamente significativos** ni sobre el throughput ($H = 0.89$, $p = .641$) ni sobre la eficiencia energética ($H = 0.72$, $p = .698$). Este resultado, consistente con el análisis descriptivo previo, indica que para los niveles de carga evaluados (0.5, 50 y 100 paquetes), el comportamiento del sistema no difiere significativamente. Esto puede explicarse por la saturación del canal: una vez que la densidad de sensores supera cierto umbral de congestión, incrementar la tasa de generación de paquetes no produce mejoras adicionales en el rendimiento efectivo.

4.2.8. Tamaño del Efecto: Eta Cuadrado (η^2)

Para cuantificar la magnitud de las diferencias encontradas, se calcula el coeficiente eta cuadrado (η^2) a partir de los estadísticos de Kruskal-Wallis:

$\eta^2 = (H - k + 1) / (n - k)$, donde H es el estadístico chi-cuadrado, k el número de grupos y n el tamaño total.

Throughput - Sensores: $\eta^2 = (18.43 - 4) / (450 - 5) = 14.43 / 445 = 0.032 \rightarrow$ Efecto pequeño pero significativo.

Eficiencia Energética ~ Sensores: $\eta^2 = (22.17 - 4) / (450 - 5) = 18.17 / 445 = 0.041 \rightarrow$ Efecto pequeño-moderado.

Aunque el tamaño del efecto es relativamente pequeño ($\eta^2 < 0.06$), la significancia estadística se mantiene gracias al tamaño muestral ($n = 450$). En términos prácticos, la relevancia del efecto del número de sensores es evidente en el análisis descriptivo: la reducción del throughput de Q-Learning entre 40 y 60 sensores supera los 96 bps (-20.6%), una disminución sustancial desde la perspectiva de diseño de red.

4.2.9. Discusión Integrada de Resultados

Superioridad del Protocolo Q-Learning

El conjunto de evidencias estadísticas converge de manera inequívoca hacia una conclusión central: el protocolo MAC basado en Q-Learning supera significativamente al protocolo BEB en las dos métricas de rendimiento evaluadas, y esta superioridad se mantiene de forma consistente a través de todos los escenarios de simulación.

La prueba U de Mann-Whitney ($p < .001$) confirma que la diferencia no es producto de la variabilidad aleatoria inherente a la simulación, sino de una diferencia estructural en la capacidad de los algoritmos para gestionar el acceso al canal. Q-Learning, al implementar un mecanismo de aprendizaje por refuerzo que adapta dinámicamente su política de transmisión según el estado observado de la red, logra reducir las colisiones y maximizar el uso productivo del canal. BEB, al operar con una regla estocástica fija de backoff exponencial, no puede adaptarse a los cambios en las condiciones de tráfico, resultando en una tasa de colisiones sistemáticamente más alta.

La magnitud de la ventaja es prácticamente relevante: Q-Learning genera en promedio más del doble de throughput (ratio QL/BEB $\approx 2.05\times$ en condiciones nominales) y entre $3.3\times$ y $4.5\times$ mayor eficiencia energética dependiendo de la densidad de nodos. En el contexto de redes IoT y WSN, donde la autonomía energética puede determinar la viabilidad operativa de años de despliegue, esta ventaja tiene implicaciones directas de diseño.

Impacto de la densidad de sensores y la naturaleza no paramétrica de los datos

El rechazo de la normalidad por parte de Kolmogorov-Smirnov y Shapiro-Wilk no constituye una debilidad metodológica, sino un hallazgo informativo sobre la naturaleza del proceso subyacente. La distribución bimodal de Q-Learning —con un clúster en valores bajos (fase exploración) y otro en valores altos (fase explotación)— refleja fielmente las dos etapas funcionales del algoritmo de aprendizaje por refuerzo: la exploración inicial, donde el agente prueba acciones subóptimas, y la explotación posterior, donde el agente aplica la política aprendida. Este comportamiento es intrínseco al algoritmo ϵ -greedy utilizado en Q-Learning y constituye evidencia de que la simulación reproduce con fidelidad el proceso real de aprendizaje.

La prueba de Kruskal-Wallis identifica un efecto significativo del número de sensores ($p \leq .001$), coherente con la teoría de capacidad de canales MAC: al incrementarse la densidad, la probabilidad de colisión simultánea aumenta siguiendo la expresión $p_{\text{colisión}} \approx 1 - (1 - p_{\text{tx}})^{N-1}$, donde N es el número de sensores activos. Los diagramas de caja por nivel de sensores muestran una tendencia decreciente en las medianas, con solapamiento parcial entre niveles adyacentes (40–45 y 55–60 sensores), lo que explica el tamaño de efecto pequeño, aunque estadísticamente significativo.

El hallazgo de que la carga de tráfico no produce diferencias significativas ($p > .60$) tiene una interpretación práctica importante: los protocolos evaluados presentan robustez ante la variación de carga dentro del rango estudiado (0.5–100 paquetes). Esto sugiere que, para el diseño de redes con estas características, la densidad de sensores es el parámetro crítico de configuración, mientras que la tasa de generación de paquetes no representa un factor determinante del rendimiento.

CAPÍTULO V

5. CONCLUSIONES Y RECOMENDACIONES

5.1. Conclusiones

- Se desarrolló exitosamente un modelo de simulación en Python que reproduce el comportamiento de una red inalámbrica de sensores asistida por superficies reconfigurables inteligentes. El modelo integró el algoritmo Q-Learning con política ϵ -greedy para gestionar el acceso al medio, configurando 15 escenarios que combinaron variaciones en el número de sensores (40 a 60 nodos) con tres niveles de carga de tráfico (0.5, 50 y 100 paquetes), ejecutando 15 réplicas por escenario. El entorno de simulación permitió reproducir con fidelidad las fases de exploración y explotación del agente, evidenciadas en la distribución bimodal de los datos registrados.
- La aplicación del modelo de simulación permitió cuantificar el throughput del protocolo Q-Learning bajo tres niveles de carga de tráfico. Se obtuvo un rango de rendimiento de 513.76 bps a 722.56 bps en todos los escenarios evaluados. El análisis estadístico mediante la prueba de Kruskal-Wallis determinó que el número de sensores constituye el factor crítico del rendimiento ($H = 18.43$, $p = .001$), mientras que la carga de tráfico no produjo efectos significativos ($H = 0.89$, $p = .641$), evidenciando que la saturación del canal domina sobre la tasa de generación de paquetes en los rangos evaluados.
- La comparación entre el protocolo Q-Learning y el Back-off Exponencial Binario demostró la superioridad estadísticamente significativa del enfoque propuesto en todas las métricas evaluadas. La prueba U de Mann-Whitney confirmó diferencias altamente significativas en throughput ($U = 2,456$, $Z = -11.73$, $p < .001$) y eficiencia energética ($U = 0.00$, $Z = -12.99$, $p < .001$). Q-Learning superó al BEB con una diferencia media de 437.62 bps en throughput y una eficiencia energética entre 5 y 8 veces mayor, validando la hipótesis de investigación y demostrando la ventaja del aprendizaje adaptativo frente a los mecanismos tradicionales de retroceso fijo.

5.2.Recomendaciones

- Se recomienda ampliar el modelo de simulación desarrollado incorporando algoritmos más avanzados de aprendizaje por refuerzo, como Deep Q-Learning o enfoques multiagente, con el propósito de analizar su desempeño en redes de mayor escala y complejidad. Asimismo, sería conveniente integrar modelos de canal más detallados que consideren efectos como el desvanecimiento Rayleigh y la movilidad de los nodos, lo cual permitiría evaluar el protocolo propuesto en condiciones más cercanas a entornos reales de implementación.
- Se sugiere validar los resultados mediante pruebas en plataformas experimentales o entornos de simulación más robustos como NS-3 o MATLAB, lo que permitiría contrastar el comportamiento del modelo con escenarios prácticos y fortalecer la aplicabilidad del estudio. Adicionalmente, se recomienda ampliar el rango de cargas de tráfico y el número de sensores evaluados, incluyendo configuraciones superiores a 60 nodos, con el fin de estudiar el comportamiento del protocolo bajo condiciones de redes ultradensas propias del paradigma IoT masivo.
- Para futuras investigaciones, se recomienda incluir el análisis de indicadores de desempeño adicionales como latencia, tasa de pérdida de paquetes y calidad de servicio, lo que proporcionaría una evaluación más integral del protocolo. Asimismo, se sugiere estudiar con mayor profundidad la configuración óptima de los elementos de la superficie reconfigurable inteligente mediante técnicas de optimización conjunta RIS-MAC, con el objetivo de maximizar la sinergia entre el aprendizaje por refuerzo y la tecnología RIS en redes de próxima generación

BIBLIOGRAFÍA

- [1] S. Gong, X. Lu, D. T. Hoang, D. Niyato, L. Shu y D. I. Kim, «Toward Smart Wireless Communications via Intelligent Reflecting Surfaces: A Contemporary Survey,» *IEEE Communications Surveys & Tutorials*, vol. 22, n° 4, 2020.
- [2] M. D. Renzo, A. Zappone, M. Debbah, Mohamed-Slim, C. Yuen y J. d. Rosny, «Smart Radio Environments Empowered by Reconfigurable Intelligent Surfaces: How It Works, State of Research, and The Road Ahead,» *IEEE Journal on Selected Areas in Communications*, vol. 39, n° 11, 14 Julio 2020.
- [3] X. Cao, B. Yang, C. Huang, C. Yuen, M. D. Renzo y Z. Han, «AI-Assisted MAC for Reconfigurable Intelligent-Surface-Aided Wireless Networks: Challenges and Opportunities,» *IEEE Communications Magazine* , vol. 59, n° 6, 5 JULIO 2021.
- [4] X. Cao, B. Yang, H. Zhang, C. Huang, C. Yuen y Z. Han, «Reconfigurable-Intelligent-Surface-Assisted MAC for Wireless Networks: Protocol Design, Analysis, and Optimization,» *IEEE Internet of Things Journal*, vol. 8, n° 18, 15 Septiembre 2021.
- [5] F. Ahmed, E. Shitiri y H.-S. Cho, «Reinforcement Learning-based MAC for Reconfigurable Intelligent Surface-Assisted Wireless Sensor Networks,» *2022 Thirteenth International Conference on Ubiquitous and Future Networks (ICUFN)*, 5 Julio 2022.
- [6] B.-J. Kwak, N.-O. Song y L. Miller, «Performance analysis of exponential backoff,» *IEEE/ACM Transactions on Networking* , vol. 13, n° 2, Abril 2005.
- [7] S. Junaid, S. Sharma, M. Patwary y N. Khan, «Non-Coherent and Backscatter Communications: Enabling Ultra-Massive Connectivity in 6G,» *IEEE Access*, 12 Marzo 2021.
- [8] G. Mengnan, E. Basar, C. Huang, R. Liu, Y. Liu y C. Yuen, «Reconfigurable intelligent surfaces for wireless communications: Overview of hardware designs, channel models and estimation techniques,» *Intelligent and Converged Networks*, vol. 3, n° 1, 2022.
- [9] J. J. Hernández, J. Pérez, O. Sallent, I. Vilà y F. Casadevall, «A Deep Q-Network-Based Algorithm for Multi-Connectivity Optimization in Heterogeneous Cellular-Networks,» *Sensors*, vol. 22, n° 6179, 2022.
- [10] A. Alnoman, A. Sharyar, A. Anpalagan y I. Woungag, «Emerging AI and 6G-Based User Localization Technologies for Emergencies and Disasters,» *IEEE Acces*, vol. 12, 2024.

- [11] M. Ali, R. Ullah, S. Chudhary y Y. Meng, «A Critical Analysis of Cooperative Caching in Ad Hoc Wireless Communication Technologies: Current Challenges and Future Directions,» *Sensors*, vol. 25, n° 1258, 19 Febrero 2025.
- [12] R. F. Ibarra, F. R. Castillo, C. A. Gutiérrez, A. García, L. A. Vásquez y A. Del-Puerto, «Machine Learning Strategies for Reconfigurable Intelligent Surface-Assisted Communication Systems—A Review,» *Future internet* , vol. 16, n° 173, 17 Mayo 2024.
- [13] S. Baque, Artist, *Implementación de un módulo didáctico para la administración y seguridad en redes de datos WLAN aplicado a la asignatura de redes inalámbrica de la carrera de ingeniería en computación y redes*. [Art]. Universidad Estatal del sur de Manabí, 2019.
- [14] A. Zafeiropoulos, D.-E. Spanos, S. Arkoulis, N. Konstantinou y N. Mitrou, «Data management in sensor networks using semantic web technologies,» *Data Management in Semantic Web*, vol. 9, 2009.
- [15] S. Hidalgo y C. Valarezo, Artists, *Diseño e implementación de un sistema de sensores WSN para agricultura de precision utilizando digimesh y modo API*. [Art]. UDLA, 2017.
- [16] L. Romero, F. J. Artigas y C. Anias, «Redes de Sensores Inalámbricos Definidas por Software: revisión del estado del arte,» *Revista de Ingeniería electrónica, Automática y comunicaciones* , vol. 41, n° 2, pp. 1-2, 1 Junio 2020.
- [17] C. Bravo, Artist, *Análisis comparativo de las plataformas Tmonte y Micaz en redes inalámbricas de sensores*. [Art]. Universidad Agraria del Ecuador , 2020.
- [18] C. Muñoz, G. Morales y V. Miramá, «Aplicación de una red de sensores inalámbricos en un ambiente de trabajo industrial,» *Espacios* , vol. 41, n° 31, 20 Agosto 2020.
- [19] J. Powell, A. McCafferty, W. Hilal y A. Gadsden, «Smart grids: A comprehensive survey of challenges, industry applications, and future trends,» *Energy Reports*, vol. 11, 20 Mayo 2024.
- [20] J. G. Vera, J. J. Romero, V. E. Gavidia y P. A. Santillan, «Review de Sensores en la Seguridad Industrial,» *Polo del conocimiento* , vol. 7, n° 7, 20 Julio 2022.
- [21] M. Tapia, Artist, *Diseño de una red WLAN en la Facultad Técnica para el Desarrollo en los laboratorios de las Carreras Eléctricas y Telecomunicaciones*. [Art]. Universidad Católica de Santiago de Guayaquil, 2020.

- [22] E. R. Calva, A. I. Chica, O. D. Matute, F. W. Reyes y D. A. Romero, Artists, *Estudio e Implementación de un sistema automático para la optimización de la gestión del servicio de atención de la Secretaria y Coordinación de Áreas de la Facultad Técnica para el Desarrollo de la UCSG*. [Art]. Universidad Católica de Santiago de Guayaquil, 2010.
- [23] R. Yadav, S. Varma y N. Malaviya, «A survey of MAC protocols for wireless sensor networks,» *UbiCC*, vol. 4, nº 3, 2009.
- [24] K. Kredo y P. Mohapatra, «Medium access control in wireless sensor networks,» *Computer Networks* , vol. 51, 2006.
- [25] W.-L. Wang, C.-C. Wang y X.-W. Yao, «Slot Self-Allocation Based MAC Protocol for Energy Harvesting Nano-Networks,» *Basel*, vol. 19, nº 21, 25 Octubre 2019.
- [26] C. B. Mwakwata, H. Malik, M. M. Alam, Y. L. Moullec, S. Parand y S. Mumtaz, «Narrowband Internet of Things (NB-IoT): From Physical (PHY) and Media Access Control (MAC) Layers Perspectives,» *Basel*, vol. 19, nº 11, 8 Junio 2019.
- [27] C. Alocious, H. Xiao, B. Christianson y J. Spring, «Embedding Trust in the Media Access Control Protocol for Wireless Networks,» *Basel*, vol. 25, nº 2, 9 Junio 2025.
- [28] I. Ahmad, T. Rahman, A. Zeb, I. Khan, M. T. B. Othman y H. Hamam, «Cooperative Energy-Efficient Routing Protocol for Underwater Wireless Sensor Networks,» *Basel* , vol. 22, nº 18, 14 Septiembre 2022.
- [29] Y. Liu, Q. Wu, T. Zhao, Y. Tie, F. Bai y M. Jin, «An Improved Energy-Efficient Routing Protocol for Wireless Sensor Networks,» *Basel*, vol. 19, nº 20, 21 Octubre 2019.
- [30] S. Fun y A. Samsudin, «Recent Technologies, Security Countermeasure and Ongoing Challenges of Industrial Internet of Things (IIoT): A Survey,» *Sensors*, vol. 21, nº 19, 2021.
- [31] M. Morales, Artist, *Especificación y verificación de un protocolo de comunicación tolerante a fallas de desconexión para aplicaciones colaborativas*. [Art]. Universidad Tecnológica de Mixteca , 2015.
- [32] M. N. A. Al-Hamadan, M. A. Fadhel, L. Alzubaidi y H. Balazs, «Reinforcement Learning Algorithms and Applications in Healthcare and Robotics: A Comprehensive and Systematic Review,» *Sensors*, vol. 24, nº 2461, 11 Abril 2024.

- [33] M. R. Bonyadi, R. Wang y M. Ziaei, «Self-Punishment and Reward Backfill for Deep Q-Learning,» *IEEE Transactions on Neural Networks and Learning Systems*, vol. 34, n° 10, Octubre 2023.
- [34] S. M. M. R. Swapno, S. N. Nobel, P. Meena, V. P. Meena, A. T. Azar, Z. Haider y M. Tounsi, «A reinforcement learning approach for reducing traffic congestion using deep Q learning,» *Sci Rep*, vol. 12, n° 14, 12 Diciembre 2024.
- [35] B. Lian, W. Xue, F. L. Lewis y A. Davoudi, «Inverse Value Iteration and Q-Learning: Algorithms, Stability, and Robustness,» *IEEE Trans Neural Netw Learn Syst*, vol. 36, n° 4, Abril 2025.
- [36] Z. Wang, X. Wang, Y. Tang, Y. Liu y J. Hu, «Optimal Tracking Control of a Nonlinear Multiagent System Using Q-Learning via Event-Triggered Reinforcement Learning,» *Entropy (Basel)*, vol. 25, n° 2, 5 Febrero 2023.
- [37] A. D. Carlo, E. M. Tosca, M. Fantozzi y P. MagnI, «Reinforcement Learning and PK-PD Models Integration to Personalize the Adaptive Dosing Protocol of Erdafitinib in Patients with Metastatic Urothelial Carcinoma,» *Clin Pharmacol Ther*, vol. 115, n° 4, Abril 2024.
- [38] A. M. Rahmani, S. Ali, M. H. Malik, E. Yousefpoor y M. Sadegh, «An energy-aware and Q-learning-based area coverage for oil pipeline monitoring systems using sensors and Internet of Things,» *Sci Rep*, vol. 12, n° 1, 10 Junio 2022.
- [39] Z. Zhu, J. Li, Z. Chu, J. Liang, H. Niu, D. Mi, C. Yin y P. Liu, «Active Reconfigurable Intelligent Surface Enhanced Internet of Medical Things,» *IEEE J Biomed Health Inform*, vol. 28, n° 7, Julio 2024.
- [40] F. Naaz, A. Nauman, T. Khurshaid y S.-W. Kim, «Empowering the Vehicular Network with RIS Technology: A State-of-the-Art Review,» *Sensors (Basel)*, vol. 24, n° 2, 5 Junio 2024.
- [41] Q. Yang, Y. Wu, H. Zhao, Y. Feng, Y. Sun, Z. Fang y G. Zheng, «Path Loss Modeling for RIS-Assisted Wireless Communication in Tunnel Scenarios,» *Sensors (Basel)*, vol. 25, n° 4, 18 Febrero 2025.
- [42] R. Silvestre, R. Llinares, L. Contat, V. Serrano, J. Ferri y E. Garcia, «Low-Temperature Soldering of Surface Mount Devices on Screen-Printed Silver Tracks on Fabrics for Flexible Textile Hybrid Electronics,» *Sensors*, vol. 22, n° 5766, 2022.

- [43] F. Tossa, Y. Faga, W. Abdou, E. C. Ezin y P. Gouton, «Wireless Sensor Network Deployment: Architecture, Objectives, and Methodologies,» *Sensors*, vol. 25, n° 11, 2025.
- [44] Y. Yu, J. Wang, X. Zhou, C. Wang, Z. Bai y Z. Ye, «Review on Channel Estimation for Reconfigurable Intelligent Surface Assisted Wireless Communication System,» *Mathematics*, vol. 11, n° 14, 2023.
- [45] Z. Zheng, S. Jiang, R. Feng, L. Ge y C. Gu, «Survey of Reinforcement-Learning-Based MAC Protocols for Wireless Ad Hoc Networks with a MAC Reference Model,» *Entropy*, vol. 25, n° 101, 2023.
- [46] F. Frommel, Artist, *Aprendizaje Profundo por Refuerzo Aplicado al Control de Acceso en Redes IEEE 802.11*. [Art]. Universidad de la república, 2022.
- [47] D. A. Guastella, G. Marcillaud y C. Valenti, «Edge-Based Missing Data Imputation in Large-Scale Environments,» *Information*, vol. 12, n° 195, 2021.
- [48] J. G. d. Pablo, M. Lindley, K. Hiramatsu y K. Goda, «High-Throughput Raman Flow Cytometry and Beyond,» *Acc Chem Res*, vol. 54, n° 9, 4 Mayo 2021.
- [49] F. Ojeda, D. Mendez, A. Fajardo y F. Ellinger, «On Wireless Sensor Network Models: A Cross-Layer Systematic Review,» *Sensor and Actuator Networks*, vol. 12, n° 50, 30 Junio 2023.
- [50] L. Sun y H. L., «Multiple-Input-Multiple-Output Filtered Multitone Time Reversal Acoustic Communications Using Direct Adaptation-Based Turbo Equalization,» *Sensors (Basel)*, vol. 23, n° 13, 1 Julio 2023.
- [51] J. Zhang, J. Yu, N. Chi, Z. Dong, X. Li y G. K. Chang, «Improved multi-channel multi-carrier generation using gain-independent multi-channel frequency shifting recirculating loop,» *Opt Express*, vol. 20, n° 28, 2012.
- [52] A. Taneja, N. Alqahtani y A. Alqahtani, «Interference Aware Resource Control for 6G-Enabled Expanded IoT Networks,» *Sensors (Basel)*, vol. 16, n° 23, 2023.
- [53] E. S. Gómez, «Python en Ingeniería en Ciencias Informáticas: proyecciones de las dimensiones práctica, científica y social,» *Serie Científica de la Universidad de las Ciencias Informáticas*, vol. 14, n° 12, 2021.
- [54] C. L. Vidal, A. Sánchez, J. Serrano y J. M. Rubio, «Experiencia académica en desarrollo rápido de sistemas de información web con Python y Django,» *Formación universitaria*, vol. 14, n° 5, 2021.

- [55] C. Wu, Y.-H. Chang, C.-T. Tsai y C.-C. Hou, «A Q-learning-based distributed queuing Mac protocol for Internet-of-Things networks,» *EURASIP Journal on Wireless Communications and Networking*, vol. 2023, n° 77, 2023.
- [56] M. M. Jaber, M. H. Ali, S. K. Abd, M. M. Jassim, A. Alkhayyat, M. Jassim, A. R. Alkhuwayldee y L. Nidhal, «Q-learning based task scheduling and energy-saving MAC protocol for wireless sensor networkss,» *Wireless Networks*, vol. 30, 2024.
- [57] W. U. Rahman, Q. Gang, F. Zhou, M. Tahir, W. Ali, M. Adil y M. I. Khattak, «Deep Q-Learning Based Adaptive MAC Protocol with Collision Avoidance and Efficient Power Control for UWSNs,» vol. 13, n° 3, 2025.
- [58] Q. Gang, W. U. Rahman, F. Zhou, M. Bilal, W. Ali, S. U. Khan y M. I. Khattak, «A Q-Learning-Based Approach to Design an Energy-Efficient MAC Protocol for UWSNs Through Collision Avoidance,» *Electronics*, vol. 13, n° 22, 2024.
- [59] S. H. Park, P. D. Mitchell y D. Grace, «Reinforcement Learning Based MAC Protocol (UW-ALOHA-Q) for Underwater Acoustic Sensor Networks,» *IEEE*, vol. 7, 2019.
- [60] N. Zerguine, «Int-BEB: Enhanced Binary Exponential Backoff using Q-learning,» *IEEE*, vol. 28, 2023.
- [61] C. Huang, R. Mo y C. Yuen, «Reconfigurable Intelligent Surface Assisted Multiuser MISO Systems Exploiting Deep Reinforcement Learning,» *IEEE*, vol. 38, n° 8, 2020.
- [62] H. Choi, L. V. Nguyen, J. Choi y A. L. Swindlehurst, «A Deep Reinforcement Learning Approach for Autonomous Reconfigurable Intelligent Surfaces,» *IEEE*, 2024.
- [63] N. Kim, W. Na, D. Shumeye, N.-N. Dao y S. Cho, «DQN-Based Directional MAC Protocol in Wireless Ad Hoc Network in Internet of Things,» *IIEE*, vol. 11, n° 7, 2023.
- [64] E. T. Vinuesa y R. L. Cañar, «Implementación de una mini-red de sensores inalámbricos para detección temprana de incendios forestales,» *Riti*, vol. 10, n° 21, 14 Agosto 2022.

ANEXOS

- Código del modelo de simulación desarrollado en el software Python.

```
# =====
# PARÁMETROS DE SIMULACIÓN
# =====
SLOT_DURATION = 1.0
EPISODE_LENGTH = 210
EPISODES_TRAIN = 200
EPISODES_EVAL = 20

PACKET_BITS = 1044
ACK_BITS = 20

# =====
# Q-LEARNING PARÁMETROS
# =====
ALPHA = 0.1
GAMMA = 0.9
EPS_START = 0.3
EPS_END = 0.01
EPS_DECAY = 0.9995
COLLISION_PENALTY = 0.5 * PACKET_BITS

# =====
# RIS / CANAL
# =====
BASE_SUCCESS_PROB = 0.95
RIS_GAIN = 1.05
SINR_DEGRADATION_SLOPE = 0.15

def success_probability_with_sinr(n_nodes, ris_gain=RIS_GAIN):
    factor_interferencia = 1.0 + SINR_DEGRADATION_SLOPE * max(0.0, (n_nodes - 10) / 10.0)
    prob = (BASE_SUCCESS_PROB * ris_gain) / factor_interferencia
    return max(0.0, min(1.0, prob))

# =====
# TRÁFICO (aquí pones 0.5, 50, 100)
# =====
LOAD_LAMBDA = 0.5

# BARRIDO DE SENSORES
# =====
NODES_LIST = list(range(40, 61, 5))
```

```

# =====
# REPETICIONES INDEPENDIENTES (ESTADÍSTICA REAL)
# =====
N_RUNS = 15
BASE_SEED = 10000

def simulate_q_learning(load_lambda, n_nodes, ris_gain=RIS_GAIN,
                        episodes_train=EPISODES_TRAIN,
                        episodes_eval=EPISODES_EVAL, seed=12345):

    # Nota: int(load_lambda) no cambia para 0.5; por eso usamos seed directamente
    rng = np.random.default_rng(seed)

    Q = np.zeros((n_nodes, N_STATES, N_ACTIONS), dtype=float)
    epsilon = EPS_START

    train_throughputs = []
    train_energies = []
    train_efficiencies = []

```

```

# ----- ENTRENAMIENTO -----
for ep in range(episodes_train):
    queues = np.zeros(n_nodes, dtype=int)
    ep_success_bits = 0.0
    ep_energy_joules = 0.0

    for t in range(EPISODE_LENGTH):
        arrivals = rng.poisson(load_lambda * SLOT_DURATION, size=n_nodes)
        queues += arrivals

        states = np.minimum(queues, 2)

        rand_vals = rng.random(n_nodes)
        greedy_actions = np.argmax(Q[np.arange(n_nodes), states, :], axis=1)
        explore_actions = rng.integers(0, N_ACTIONS, size=n_nodes)
        actions = np.where(rand_vals < epsilon, explore_actions, greedy_actions)

        transmitters = np.where(actions == 1)[0]
        n_tx = len(transmitters)
        rewards = np.zeros(n_nodes, dtype=float)

        if n_tx == 1:
            idx = transmitters[0]
            ep_energy_joules += PACKET_BITS * E_TX_PER_BIT

            success_prob = success_probability_with_sinnr(n_nodes, ris_gain)
            if rng.random() < success_prob:
                rewards[idx] = PACKET_BITS - TX_ENERGY_COST
                queues[idx] = max(0, queues[idx] - 1)
                ep_success_bits += PACKET_BITS
            else:
                rewards[idx] = -0.2 * PACKET_BITS - TX_ENERGY_COST

```

```

# ----- EVALUACIÓN -----
eval_throughputs = []
eval_energies = []
eval_efficiencies = []

for ep in range(episodes_eval):
    queues = np.zeros(n_nodes, dtype=int)
    ep_success_bits = 0.0
    ep_energy_joules = 0.0

    for t in range(EPISODE_LENGTH):
        arrivals = rng.poisson(load_lambda * SLOT_DURATION, size=n_nodes)
        queues += arrivals
        states = np.minimum(queues, 2)

        actions = np.argmax(Q[np.arange(n_nodes), states, :], axis=1)
        transmitters = np.where(actions == 1)[0]
        n_tx = len(transmitters)

        if n_tx == 1:
            idx = transmitters[0]
            ep_energy_joules += PACKET_BITS * E_TX_PER_BIT
            success_prob = success_probability_with_sinr(n_nodes, ris_gain)
            if rng.random() < success_prob:
                queues[idx] = max(0, queues[idx] - 1)
                ep_success_bits += PACKET_BITS
        elif n_tx > 1:
            ep_energy_joules += n_tx * PACKET_BITS * E_TX_PER_BIT

    throughput_bps = ep_success_bits / (EPISODE_LENGTH * SLOT_DURATION)
    if ep_energy_joules > 0:
        efficiency_bits_per_J = ep_success_bits / ep_energy_joules
    else:
        efficiency_bits_per_J = 0.0

```

```

def simulate_beb(load_lambda, n_nodes, ris_gain=RIS_GAIN,
                episodes=EPISODES_EVAL, seed=54321):

    rng = np.random.default_rng(seed)

    CW_MIN = 4
    CW_MAX = 1024

    eval_throughputs = []
    eval_energies = []
    eval_efficiencies = []

    for ep in range(episodes):
        queues = np.zeros(n_nodes, dtype=int)
        cw = np.full(n_nodes, CW_MIN, dtype=int)
        backoff = rng.integers(0, CW_MIN, size=n_nodes)

        ep_success_bits = 0.0
        ep_energy_joules = 0.0

        for t in range(EPISODE_LENGTH):
            arrivals = rng.poisson(load_lambda * SLOT_DURATION, size=n_nodes)
            queues += arrivals

            active = (queues > 0)
            backoff[active] -= 1

            transmitters = np.where((backoff <= 0) & active)[0]
            n_tx = len(transmitters)

            if n_tx == 1:
                idx = transmitters[0]
                ep_energy_joules += PACKET_BITS * E_TX_PER_BIT

                success_prob = success_probability_with_sinr(n_nodes, ris_gain)
                if rng.random() < success_prob:

```