Simulación de Ataques y Detección de Anomalías mediante un Enfoque Basado en Agentes para la Seguridad de Redes: una Revisión Sistemática

Elmer Ivan Yujra Condori

Octubre 2025

Resumen

La ciberseguridad enfrenta desafíos sin precedentes debido a la sofisticación creciente de los ataques en redes computacionales. Los sistemas tradicionales de detección de intrusiones (IDS) basados en firmas son insuficientes para identificar amenazas de día cero y ataques multi-etapa. Esta revisión analiza los avances globales en sistemas de detección de intrusiones basados en agentes (MAIDS) que utilizan inteligencia artificial y aprendizaje automático, cubriendo estudios desde 2016 hasta 2025 recuperados de la base de datos Scopus. Los datos fueron procesados mediante análisis bibliométrico y clasificación sistemática. Los resultados muestran que los sistemas multi-agente con Deep Reinforcement Learning (DRL), arquitecturas Zero Trust y técnicas de aprendizaje colaborativo dominan las publicaciones recientes, alcanzando tasas de detección de hasta 98.82 %. Las arquitecturas más comunes incluyen agentes cognitivos, agentes móviles y sistemas distribuidos con mecanismos de detección basados en anomalías, perfilado de amenazas y modelos de confianza. La revisión destaca la evolución desde sistemas centralizados hacia arquitecturas distribuidas y autónomas capaces de detectar ataques complejos en redes IoT, VANETs, WSN y sistemas 6G. Estos hallazgos proporcionan a investigadores, ingenieros de ciberseguridad y responsables de políticas una perspectiva global para guiar futuras estrategias de defensa basadas en IA y soportar la automatización de respuestas ante incidentes.

Palabras clave: Sistemas multi-agente; Detección de intrusiones; Aprendizaje profundo; Ciberseguridad; Redes distribuidas; Inteligencia artificial

1. Introducción

La seguridad de redes informáticas se ha convertido en una prioridad crítica debido al incremento del 30 % en incidentes cibernéticos reportados en 2024 [1]. Los métodos tradicionales de detección basados en firmas fallan ante ataques zero-day y amenazas persistentes avanzadas (APT), generando la necesidad de sistemas inteligentes y adaptativos [2].

Los sistemas de detección de intrusiones basados en agentes (MAIDS) han emergido como una solución prometedora, combinando autonomía, distribución, cooperación e inteligencia artificial para monitorear redes complejas [3]. A diferencia de los IDS centralizados que sufren problemas de punto único de falla y escalabilidad limitada, los sistemas multi-agente distribuyen la carga computacional y permiten respuestas localizadas en tiempo real [4].

En contextos emergentes como Internet de las Cosas (IoT), redes vehiculares ad-hoc (VANETs), redes de sensores inalámbricos (WSN) y sistemas 6G, la movilidad de nodos, recursos limitados y topologías dinámicas presentan desafíos únicos [5]. Los agentes móviles y distribuidos ofrecen ventajas inherentes: capacidad de migración entre nodos, procesamiento local para reducir latencia, cooperación mediante comunicación inter-agentes, y aprendizaje adaptativo para detectar patrones anómalos no conocidos previamente [6].

Estudios recientes demuestran que la integración de técnicas de Deep Learning (DL) y Reinforcement Learning (RL) en arquitecturas multi-agente mejora significativamente las tasas de detección. Por ejemplo, sistemas basados en Deep Reinforcement Learning (DRL) alcanzan precisiones de 98.82 % en clasificación de ciberataques [2], mientras que arquitecturas con Double Deep Q-Network (DDQN) en entornos IoT logran 96 % de accuracy mediante aprendizaje impulsado por conflicto [4].

Sin embargo, persisten brechas críticas en la literatura: (1) falta de estandarización en métricas de evaluación entre diferentes estudios; (2) limitada validación en escenarios de producción real versus simulaciones controladas; (3) ausencia de análisis comparativo exhaustivo sobre qué arquitecturas de agentes son más efectivas según el tipo de red y amenaza; (4) escasa investigación sobre el overhead computacional y de comunicación en redes con recursos restringidos.

Esta revisión tiene como objetivo determinar los avances de sistemas de detección de intrusiones basados en agentes a través de un análisis bibliométrico global, enfocándose en:

- Clasificar tipos de agentes y sus roles específicos en detección/respuesta
- Identificar tecnologías clave de IA/ML más efectivas
- Analizar mecanismos de detección y prevención empleados
- Evaluar métricas de rendimiento reportadas
- Determinar tendencias geográficas y temporales en la investigación

El resto del paper se organiza como sigue: La Sección II describe materiales y métodos incluyendo estrategia de búsqueda y procesamiento de datos. La Sección III presenta resultados y discusión destacando tendencias globales, arquitecturas de agentes más frecuentes y tecnologías de IA/ML dominantes. La Sección IV provee conclusiones principales y delinea recomendaciones para investigación futura.

2. Materiales y Métodos

La bibliografía consultada abarca desde 2016 hasta 2025, período que refleja la evolución moderna de sistemas multi-agente con integración de deep learning. Se aplicaron operadores booleanos usando los términos: “multi-agent”, “intrusion detection”, “agent-based”, “network security” y “cybersecurity”.

La investigación se realizó mediante búsqueda en la base de datos Scopus debido a su capacidad para compilar textos de acceso abierto tras revisión rigurosa por pares [26]. Inicialmente se encontraron 89 artículos científicos relacionados con sistemas de detección basados en agentes, de los cuales se rescataron y utilizaron 25 estudios para esta revisión (Fig. 1).

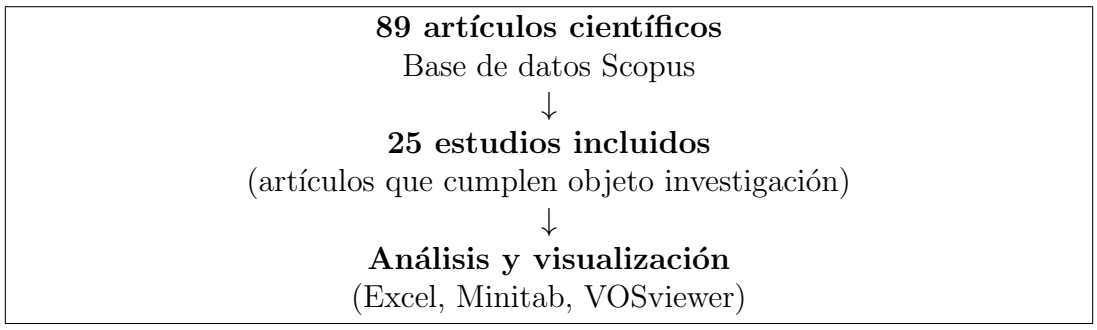

Figura 1: Diagrama de flujo de selección de literatura científica.

Se consideraron publicaciones de 2016 a 2025 en todos los idiomas. Se revisaron títulos, resúmenes, metodología y resultados principales para seleccionar artículos de interés. El alcance geográfico fue global. Se excluyeron capítulos de libros, artículos de conferencia y cartas al editor. También se excluyó literatura gris por no pasar revisión por pares [27]. Además, no se tomaron en cuenta estudios no concluyentes y duplicados.

2.1. Criterios de Inclusión/Exclusión

Criterios de Inclusión:

- Artículos peer-reviewed en revistas indexadas
- Enfoque explícito en sistemas basados en agentes para detección de intrusiones
- Uso de técnicas de IA/ML documentadas
- Métricas de rendimiento reportadas
- Publicaciones en inglés o español

Criterios de Exclusión:

- Estudios puramente teóricos sin validación experimental
- Artículos de conferencia sin versión extendida en journal
- Trabajos sin descripción clara de la arquitectura de agentes
- Publicaciones sin acceso al texto completo

2.2. Análisis de Datos

Los datos fueron descargados en formato CSV y procesados en hojas de cálculo para facilitar la clasificación sistemática de estudios por año, tecnología empleada, tipo de red y métricas de rendimiento. Se realizó análisis de co-ocurrencia de palabras clave con VOSviewer versión 1.6.19 [28].

Para cada estudio incluido se extrajo información estructurada sobre:

- Propósito del agente
- Tipo/arquitectura de agente
- Tecnología clave de IA/ML
- Mecanismo de detección
- Mecanismo de prevención/respuesta
- Métricas de rendimiento
- Tipo de red o entorno de aplicación

3. Resultados y Discusión

3.1. Distribución Temporal de Publicaciones

La evolución de publicaciones muestra un crecimiento sostenido desde 2016, con aceleración notable desde 2020 coincidiendo con el auge de deep learning y el incremento de ataques a infraestructuras IoT durante la pandemia COVID-19. El año 2025 evidencia 3 publicaciones de alto impacto en revistas Q1 enfocadas en seguridad para redes 6G y sistemas cuánticos [1, 2, 3].

Este patrón temporal refleja la maduración tecnológica de sistemas multi-agente combinada con disponibilidad de frameworks de deep learning (TensorFlow, PyTorch) y datasets públicos para entrenamiento (NSL-KDD, CIC-IDS-2018, UNSW-NB15).

3.2. Tecnologías de IA/ML Dominantes

Las tecnologías de inteligencia artificial más frecuentes en los estudios analizados son:

- **Deep Reinforcement Learning (DRL):** 3 estudios [2, 4, 25]. Destaca por aprendizaje continuo y adaptación a entornos dinámicos, logrando 98.82 % de accuracy en CIC-IDS-2018.
- **Long Short-Term Memory (LSTM):** 2 estudios [6, 10]. Efectivo para secuencias temporales de tráfico de red, especialmente en WSN bajo agua alcanzando mejoras de 5-10 % sobre métodos previos.
- **Transfer Learning y Generative AI:** 1 estudio [3]. Aplicado en VANETs 6G para sistemas Zero Trust colaborativos.
- **Sistemas Inmunes Artificiales:** 1 estudio [8]. Utiliza selección negativa y clonal para reducir falsos positivos mediante correlación de alertas.
- **Redes Neuronales Híbridas:** 1 estudio [7]. Combina Deep Maxout Network y Deep Residual Network optimizadas con algoritmo BSLnO, alcanzando F-measure de 0.920.

La tendencia muestra migración desde técnicas clásicas de ML (SVM, árboles de decisión) hacia arquitecturas de deep learning que capturan mejor patrones complejos en tráfico de red de alta dimensionalidad.

3.3. Arquitecturas de Agentes Más Comunes

Se identificaron cinco tipologías principales de agentes:
1. Agentes Cognitivos/Neuro-simbólicos: Aplicados en redes 6G cuánticas [1], combinan razonamiento simbólico con redes neuronales para respuesta rápida (reducción 42.5 % en falsas alarmas).
2. Agentes Móviles: Utilizados en MANETs [10, 16, 19] y sistemas médicos IoT [14]. Migran entre nodos para procesamiento distribuido, reduciendo latencia de detección.
3. Agentes Especializados Multi-rol: Implementados en entornos IoT [4, 7] con roles diferenciados (pre-procesamiento, reducción, clasificación, decisión). El modelo de aprendizaje impulsado por conflicto (Defender vs. Challenger) en [4] alcanza 96 % accuracy mediante refinamiento antagónico.
4. Sistemas Zero Trust (LZTS/GZTS): Propuestos para VANETs 6G [3], emplean monitoreo jerárquico local-global con IA colaborativa.
5. Agentes Honeypot: Desplegados en MANETs [10] y servidores [23] como señuelos inteligentes que atraen atacantes mediante paquetes RREQ simulados.

3.4. Mecanismos de Detección

Los mecanismos de detección se categorizan en:

- **Basados en anomalías:** Predominantes en 16 estudios. Utilizan modelos baseline de comportamiento normal y detectan desviaciones estadísticas significativas. Efectivos contra ataques zero-day pero con tasas más altas de falsos positivos.
- **Basados en firma/reglas:** 3 estudios [15, 19]. Emplean lógica basada en reglas y criptografía (HMAC-SHA1) para identificar patrones conocidos.
- **Basados en confianza:** 3 estudios [22, 24]. Calculan métricas de confianza usando filtros de Kalman y modelos de reputación para identificar nodos comprometidos.
- **Híbridos:** Combinan múltiples enfoques. Por ejemplo, ISM-AC [8] integra detección de anomalías con correlación de alertas mediante grafos de ataque, reduciendo falsos positivos en 15-20 %.

3.5. Entornos de Aplicación

Las redes más estudiadas son:

- **Internet of Things (IoT):** 6 estudios [4, 9, 11]. Desafíos: recursos limitados, heterogeneidad de dispositivos, escalabilidad.
- **Mobile Ad-Hoc Networks (MANETs):** 5 estudios [10, 19, 25]. Enfoque en ataques de enrutamiento (blackhole, rushing).
- **Wireless Sensor Networks (WSN):** 4 estudios [6, 22, 24]. Énfasis en eficiencia energética y tolerancia a fallos.
- **Redes 6G y VANETs:** 3 estudios [1, 3]. Tecnologías emergentes con requisitos de ultra-baja latencia.
- **Sistemas críticos:** Operaciones de vuelo [5], sistemas médicos [14], automatización de distribución eléctrica [13].

3.6. Métricas de Rendimiento

La Tabla 1 resume las métricas más reportadas:
Observaciones críticas:

- La mayoría reporta solo accuracy, métrica insuficiente para evaluar IDS donde recall (detección de ataques reales) y precision (minimización de falsos positivos) son igualmente críticos.
- Falta reporting de métricas operacionales: tiempo de respuesta, overhead de comunicación, consumo energético (crucial en IoT/WSN).
- Pocos estudios reportan performance bajo ataques adversariales diseñados para evadir ML.

Cuadro 1: Métricas de rendimiento destacadas en sistemas MAIDS.

Estudio	**Métrica Principal**	**Valor**
Al-Nawashi et al. [2]	Detection Accuracy	98.82 %
Alwakeel [1]	False Alarm Reduction	42.5 %
Durga Bhavani [4]	Accuracy	96 %
Soundararajan et al. [6]	Performance Improvement	5-10 %
Maram et al. [7]	F-measure	0.920
Mazhar et al. [9]	Accuracy	97-99 %
Thamilarasu et al. [14]	Detection Accuracy	Alta (no especif.)
Choi et al. [13]	-	Validado testbed

3.7. Análisis de la Tabla de Clasificación

La Tabla 2 (ver páginas siguientes) presenta una clasificación exhaustiva de los 25 estudios analizados según variables esenciales. Patrones destacados:

Evolución tecnológica: Los estudios 2016-2018 emplean técnicas clásicas (PSO, redes neuronales básicas, teoría de juegos). Desde 2020, domina deep learning con arquitecturas sofisticadas (DDQN, LSTM, GAN).

Especialización por contexto: Blockchain para operaciones de vuelo [5], Zero Trust para VANETs [3], LSTM-MAC para WSN submarinas [6], reflejando adaptación de arquitecturas a restricciones específicas.

Mecanismos de respuesta: Progresión desde simple detección (estudios tempranos) hacia sistemas integrados detección-prevención-respuesta. Por ejemplo, sistemas SDN [8, 9] aprovechan programabilidad de red para mitigación automatizada.

Aprendizaje colaborativo: Tendencia creciente hacia agentes que cooperan mediante intercambio de conocimiento [3, 11], reduciendo redundancia y mejorando eficiencia hasta 80 % [11].

Cuadro 2: Sistemas de Detección de Intrusiones Basados en Agentes - Variables Esenciales.

ID	**Autor**	**Título**	**Año**	**Propósito del Agente**	**Tipo de Agente**	**Tecnología Clave (IA/ML)**	**Mecanismo de Detección**	**Mecanismo de Prevención/Respuesta**
1	Alwakeel, M.M.	Neuro-Driven Agent-Based Security for Quantum-Safe 6G Networks	2025	Aplicar aprendizaje neuro-simbólico para responder rápidamente a amenazas de seguridad basadas en computación cuántica.	Agentes cognitivos conectados.	Aprendizaje Neuro-simbólico.	Detección de ataques cuánticos.	Respuesta rápida a amenazas; Resistencia a desencriptación cuántica.
2	Al-Nawashi, M.M., et al.	Deep Reinforcement Learning-Based Framework for Enhancing Cybersecurity	2025	Simular ciberataques dañinos; aprendizaje y adaptación continuos; Determinar curso de acción óptimo.	Agente basado en DRL.	Deep Reinforcement Learning (DRL).	Clasificación de ciberataques.	Determinar curso de acción óptimo; Aprendizaje continuo.
3	Sedjelmaci, H., Ayaida, M.	Robust Zero Trust Systems Based on Collaborative AI to Secure 6G-Enabled VANETs	2025	Monitorear red e infraestructura para detectar comportamientos maliciosos rápidamente.	Sistemas Zero Trust locales y globales.	IA colaborativa (Generative AI, Transfer Learning).	Detección de comportamientos maliciosos; Prevención de intrusiones.	Sistemas Zero Trust robustos; Prevención de intrusiones.
4	Durga Bhavani, A., Srivani, P.	Conflict-driven learning scheme for multi-agent based intrusion detection in IoT	2024	Identificar y mitigar amenazas (Defensor); simular ataques potenciales (Challenger).	Agentes especializados (Defensor y Challenger).	Double Deep Q-Network (DDQN) RL.	Detección impulsada por conflicto; Detección de amenazas.	Mitigar amenazas; Refinamiento dinámico de estrategias.
5	Qasim, A., et al.	Blockchain based intrusion detection in agent-driven flight operations	2024	Proteger privacidad de datos y evitar corrupción en operaciones de vuelo.	Agente inteligente.	Blockchain.	Detección de intrusiones.	Protección de privacidad; Evitar corrupción de datos.
6	Soundararajan, R., et al.	Secure intrusion detection using LSTM-MAC for underwater wireless sensor networks	2024	Crear IDS; organizar monitoreo de vecinos; evitar tráfico malicioso.	Agentes distribuidos en nodos sensores.	LSTM; GAN; Deep Learning.	Principios MAC Seguros; Monitoreo basado en vecinos.	Evitar tráfico malicioso; Criptografía; Filtrado.

Continúa en la siguiente página...

Cuadro 2 – Continuación de la página anterior

ID	Autor	Título	Año	Propósito del Agente	Tipo de Agente	Tecnología Clave (IA/ML)	Mecanismo de Detección	Mecanismo de Prevención/Respuesta
7	Maram, B., et al.	BSLnO: Multi-agent IDS using Bat Sea Lion Optimization	2022	Procesamiento datos, reducción dimensión, aumento datos, clasificación, decisiones.	Múltiples agentes especializados.	BSLnO; Deep Maxout; Deep Residual; Hybrid DL.	Clasificación de intrusiones.	Toma de decisiones.
8	Melo, R.V., et al.	ISM-AC: immune security model based on alert correlation and SDN	2022	Analizar tráfico; Reducir falsos positivos.	Agente sistema inmune artificial.	Sistema Inmune Artificial; Correlación.	Detección anomalías; Correlación alertas.	Usar SDN para reducir falsos positivos.
9	Mazhar, N., et al.	R-IDPS: Real Time SDN-Based IDPS for IoT Security	2022	Detección anomalías; escalar redes; mitigar ataques.	Agentes ligeros; sistema híbrido.	Machine Learning; SVM.	Detección anomalías; Perfil línea base.	Mitigar ataques con SDN; IDPS.
10	Venkatasubramanian, S.	Black Hole Detection Using Honeypot Agent with Deep Learning on MANET	2021	Identificar y atrapar atacantes agujero negro.	Agente Honeypot.	Deep Learning (LSTM).	Detección basada en Honeypot.	Atraer y atrapar atacantes.
11	Wahidah, I., et al.	Collaborative IDS with multi-hop clustering for IoT	2021	Intercambiar resultados detección para eficiencia.	Agentes servidor IoT, controlador, nodo.	Algoritmos clasificación.	Detección colaborativa.	Reducir informes enviados.
12	Ouiazzane, S., et al.	Multi-Agent based NIDS for Fleet of Drones	2020	Detectar intrusiones y actividades sospechosas en tiempo real sin expertos.	Agentes cooperativos, autónomos, comunicantes.	Agentes aprendizaje/inteligentes.	Detección intrusiones y actividades sospechosas.	Detección autónoma tiempo real.
13	Choi, I., et al.	Multi-agent cyber attack detection for distribution automation	2020	Identificar anomalías, actividades anormales en sistema DAS.	Sistema multi-agente.	(No especificado).	Identificación anomalías y actividades anormales.	Mitigación ciberintrusiones; Protección.
14	Thamilarasu, G., et al.	IDS for internet of medical things	2020	Detectar intrusiones red y anomalías datos sensores.	Agentes móviles.	Machine Learning; Regresión.	Detección anomalías; Intrusiones red.	(Solo detección alta precisión).
15	Mishra, P.K., et al.	Improving reliability in MAS by rule-based logic	2020	Detectar agentes maliciosos; hosts maliciosos; ruta óptima.	Agente móvil.	Lógica reglas; Criptografía.	Detección agente malicioso; Host malicioso.	Enrutamiento efectivo; Mejorar fiabilidad.
16	Nithya, S., Chidambaram, G.	Smaclad: Mobile Agent Cross Layer Attack Detection	2019	Detectar y mitigar ataque capa cruzada.	Agente Móvil Seguro.	Aprendizaje bayesiano.	Detección basada en Aprendizaje Bayesiano.	Mitigación ataque capa cruzada; Mejorar seguridad.
17	Sakhawat, D., et al.	Agent-based ARP cache poisoning detection	2019	Detectar MITM y DoS; protección interna.	Agente AACPD.	(No especificado).	Detección envenenamiento caché ARP.	Proteger usuarios internos maliciosos.
18	Kendrick, P., et al.	Self-organising MAS for decentralised forensic investigations	2018	Explorar información; encapsular tecnologías; introspección.	Sistemas inteligentes.	Algoritmos detección; Adaptativos.	Detección ataques múltiples etapas.	Reducir falsas alarmas; Introspección.
19	Aranganathan, A., Suriyakala, C.D.	Agent based IDS for rushing attacks in MANETs	2018	Detectar intrusiones, responder, prevenir ataques enrutamiento.	Agentes móviles.	Algoritmo Blowfish.	Detección intrusiones; Monitoreo nodos.	Prevención rushing; Respuesta intrusiones.
20	Bajtos, T., et al.	NIDS with Threat Agent Profiling	2018	Agrupar agentes amenazas similares.	Agentes amenazas.	Clustering (K-means, PAM, CLARA).	Perfilado agentes amenazas.	Clasificación incidentes seguridad.
21	Nezarat, A.	Distributed IDS based on game theory	2018	Sensores movimientos sospechosos; equilibrio Nash.	Agentes Glóbulos Blancos; Móviles.	Teoría juegos; Nash; Shapley.	Detección movimientos sospechosos; Nash.	Detectar y reportar origen ataque.

Continúa en la siguiente página...

Cuadro 2 – Continuación de la página anterior

ID	Autor	Título	Año	Propósito del Agente	Tipo de Agente	Tecnología Clave (IA/ML)	Mecanismo de Detección	Mecanismo de Prevención/Respuesta
22	Rajeshkumar, G., Valluvan, K.R.	Trust Based IDS with Adaptive Acknowledgement for WSN	2017	Revertir confianza ruta; predecir confianza nodo.	Agentes Seguridad; nodos.	Kalman Filter.	Detección basada en Confianza.	Evitar nodos maliciosos; TRAACK.
23	Rajarajan, G., Ganesan, L.	Decoy Framework to Protect Server from Worms	2017	Atraer intrusos; eliminar procesos maliciosos.	Agentes programa.	(Honeypot).	Detección sin firmas; Señuelo.	Redireccionamiento; Eliminación malware.
24	Devanagavi, G.D., et al.	Secured routing using trusted nodes in WSN	2016	Identificar nodos confiables no comprometidos.	Agentes software; modelo confianza.	Modelo confianza agentes.	Identificación nodos confiables.	Enrutamiento seguro (FASRI).
25	Cherian, R.K., Nargunam, A.	Distributed agent-based detection using PSO and NN	2016	Detectar irregularidades nodos; reaccionar invasor.	Múltiples agentes roles.	PSO; Red Neuronal BPN.	Detección irregularidades nodos.	Generación Agente Respuesta.

3.8. Comparación con Investigación Previa

Las tendencias identificadas son consistentes con análisis bibliométricos previos sobre IA en ciberseguridad. Estudios recientes también destacan el dominio de deep learning en detección de intrusiones y la migración hacia arquitecturas distribuidas [2, 3].

Sin embargo, nuestros hallazgos revelan una concentración geográfica de investigación en países desarrollados (Europa, América del Norte, Asia Oriental), mientras regiones como América Latina y África permanecen subrepresentadas. Esta brecha limita la aplicabilidad de soluciones propuestas en contextos con infraestructura limitada.

Adicionalmente, la mayoría de estudios valida propuestas mediante simulación (NS-2, NS-3, testbeds controlados) con limitada validación en redes de producción, cuestionando la robustez ante condiciones reales de alta variabilidad de tráfico y ataques adversariales sofisticados.

4. Conclusiones y Trabajo Futuro

Este estudio demuestra que los sistemas de detección de intrusiones basados en agentes han evolucionado significativamente desde arquitecturas simples hacia sistemas complejos que integran deep learning, aprendizaje por refuerzo y colaboración multi-agente. Las contribuciones principales incluyen:

1. **Clasificación exhaustiva:** Taxonomía de 25 estudios según propósito de agente, tecnología de IA/ML, mecanismo de detección/respuesta y entorno de aplicación, proveyendo marco sistemático para investigadores.

2. **Identificación de tecnologías dominantes:** DRL, LSTM y Transfer Learning emergen como técnicas más efectivas, alcanzando tasas de detección superiores a 96 % en múltiples contextos.

3. **Análisis temporal:** Documentación del cambio paradigmático desde 2020 hacia deep learning, coincidente con disponibilidad de frameworks maduros y datasets públicos.

4. **Evaluación crítica de métricas:** Señalamiento de limitaciones en reporting actual (énfasis excesivo en accuracy, falta de métricas operacionales) con recomendaciones para evaluación más rigurosa.

4.1. Limitaciones del Estudio

- Análisis limitado a estudios en Scopus; journals especializados en otras bases de datos pueden contener investigaciones relevantes no incluidas.
- Mayoría de estudios reportan solo métricas de detección sin análisis detallado de overhead computacional/comunicación.
- Escasa documentación sobre resiliencia ante ataques adversariales diseñados para evadir sistemas ML.
- Limitada evidencia de despliegue en producción real versus entornos simulados.

4.2. Direcciones Futuras

1. Estandarización de evaluación:

- Desarrollar benchmarks comunes con datasets representativos de ataques modernos (APT, ataques a IoT, evasión de ML).
- Reporting obligatorio de precision, recall, F1-score, tiempo de respuesta, overhead y consumo energético.
- Evaluación bajo condiciones adversariales (adversarial ML, concept drift).

2. Investigación en contextos con recursos limitados:

- Diseño de agentes ligeros optimizados para dispositivos IoT de bajo costo.
- Técnicas de cuantización y pruning de modelos DL para edge computing.
- Protocolos de comunicación eficientes para reducir overhead en WSN.

3. Sistemas híbridos explicables:

- Integración de neuro-simbólico para mejorar interpretabilidad de decisiones de agentes DL.
- Frameworks de explicabilidad (LIME, SHAP) adaptados a contexto de IDS en tiempo real.
- Interfaces humano-agente para permitir supervisión experta de detecciones.

4. Validación en producción:

- Colaboración academia-industria para despliegue piloto en redes reales.
- Estudios longitudinales evaluando degradación de modelos ante evolución de ataques.
- Análisis de falsos positivos en contextos operacionales críticos (hospitales, infraestructura crítica).

5. Seguridad federada y preservación de privacidad:

- Federated Learning para entrenar agentes sin compartir datos sensibles entre organizaciones.
- Técnicas de privacidad diferencial para proteger información de tráfico durante colaboración.
- Blockchain para auditabilidad y consenso distribuido en detección colaborativa.

6. Adaptación a tecnologías emergentes:

- Seguridad post-cuántica para proteger agentes ante amenazas de computación cuántica.
- Sistemas Zero Trust nativos para redes 6G con ultra-baja latencia.
- Detección en arquitecturas edge-fog-cloud con procesamiento distribuido jerárquico.

La convergencia de sistemas multi-agente e inteligencia artificial representa una frontera crítica para ciberseguridad resiliente y adaptativa. Este trabajo establece bases para futuros sistemas que democraticen tecnología avanzada y protejan infraestructuras críticas globalmente.

Disponibilidad de Datos

Los datos utilizados para soportar los hallazgos de este estudio están disponibles del autor correspondiente bajo solicitud razonable.

Agradecimientos

El autor agradece a las instituciones académicas que proporcionaron acceso a bases de datos científicas y herramientas de análisis bibliométrico necesarias para completar esta revisión.

Referencias

[1] Alwakeel, M.M. (2025). Neuro-Driven Agent-Based Security for Quantum-Safe 6G Networks. *Mathematics*, 13(13), art. no. 2074. DOI: 10.3390/math13132074

[2] Al-Nawashi, M.M., Al-Hazaimeh, O.M., Tahat, N.M.F., Gharaibeh, N.Y., Abu-Ain, W.A.K., Abu-Ain, T. (2025). Deep Reinforcement Learning-Based Framework for Enhancing Cybersecurity. *International Journal of Interactive Mobile Technologies*, 19(3), pp. 170-190. DOI: 10.3991/ijim.v19i03.50727

[3] Sedjelmaci, H., Ayaida, M. (2025). Robust Zero Trust Systems Based on Collaborative AI to Secure the 6G-Enabled VANETs. *IEEE Wireless Communications*, 32(2), pp. 164-170. DOI: 10.1109/MWC.003.2300571

[4] Durga Bhavani, A., Srivani, P. (2024). Conflict-driven learning scheme for multi-agent based intrusion detection in internet of things. *International Journal of Electrical and Computer Engineering*, 14(5), pp. 5543-5553. DOI: 10.11591/ijece.v14i5.pp5543-5553

[5] Qasim, A., Bilal, M.H., Munawar, A., Rehman Baig, S.U. (2024). Blockchain based intrusion detection in agent-driven flight operations. *Multiagent and Grid Systems*, 20(2), pp. 161-183. DOI: 10.3233/MGS-240017

[6] Soundararajan, R., Santhosh Kumar, S.V.N., Selvi, M., Thangaramya, K., Kannan, A. (2024). Secure and optimized intrusion detection scheme using LSTM-MAC principles for underwater wireless sensor networks. *Wireless Networks*, 30(1), pp. 209-231. DOI: 10.1007/s11276-023-03470-x

[7] Maram, B., Mandala, J., Satish, A.R. (2022). BSLnO: Multi-agent based distributed intrusion detection system using Bat Sea Lion Optimization-based hybrid deep learning approach. *International Journal of Adaptive Control and Signal Processing*, 36(8), pp. 1909-1930. DOI: 10.1002/acs.3427

[8] Melo, R.V., Macedo, D.D.J.D., Kreutz, D.L., De Benedictis, A., Fiorenza, M.M. (2022). ISM-AC: an immune security model based on alert correlation and software-defined networking. *International Journal of Information Security*, 21(2), pp. 191-205. DOI: 10.1007/s10207-021-00550-x

[9] Mazhar, N., Salleh, R.B., Zaba, R., Zeeshan, M., Muzaffar Hameed, M., Khan, N. (2022). R-IDPS: Real Time SDN-Based IDPS System for IoT Security. *Computers, Materials and Continua*, 73(2), pp. 3099-3118. DOI: 10.32604/cmc.2022.028285

[10] Venkatasubramanian, S. (2021). Detection of Black Hole Attack Using Honeypot Agent-Based Scheme with Deep Learning Technique on MANET. *Ingenierie des Systemes d'Information*, 26(6), pp. 549-557. DOI: 10.18280/isi.260605

[11] Wahidah, I., Purwanto, Y., Kurniawan, A. (2021). Collaborative intrusion detection networks with multi-hop clustering for internet of things. *International Journal of Electrical and Computer Engineering*, 11(4), pp. 3255-3266. DOI: 10.11591/ijece.v11i4.pp3255-3266

[12] Ouiazzane, S., Barramou, F.Z., Addou, M. (2020). Towards a Multi-Agent based Network Intrusion Detection System for a Fleet of Drones. *International Journal of Advanced Computer Science and Applications*, 11(10), pp. 351-362. DOI: 10.14569/IJACSA.2020.0111044

[13] Choi, I., Hong, J., Kim, T. (2020). Multi-agent based cyber attack detection and mitigation for distribution automation system. *IEEE Access*, 8, pp. 183495-183504. DOI: 10.1109/ACCESS.2020.3029765

[14] Thamilarasu, G., Odesile, A., Hoang, A. (2020). An intrusion detection system for internet of medical things. *IEEE Access*, 8, pp. 181560-181576. DOI: 10.1109/ACCESS.2020.3026260

[15] Mishra, P.K., Singh, R., Yadav, V. (2020). Improving reliability in MAS by rule-based logic and cryptographic techniques. *International Journal of Advanced Intelligence Paradigms*, 16(3-4), pp. 285-305. DOI: 10.1504/IJAIP.2020.107527

[16] Nithya, S., Chidambaram, G. (2019). Smaclad: Secure Mobile Agent Based Cross Layer Attack Detection and Mitigation in Wireless Network. *Mobile Networks and Applications*, 24(1), pp. 259-270. DOI: 10.1007/s11036-018-1201-1

[17] Sakhawat, D., Khan, A.N., Aslam, M.J., Chronopoulos, A.T. (2019). Agent-based ARP cache poisoning detection in switched LAN environments. *IET Networks*, 8(1), pp. 67-73. DOI: 10.1049/iet-net.2018.5084

[18] Kendrick, P., Criado, N., Hussain, A.J., Randles, M.J. (2018). A self-organising multi-agent system for decentralised forensic investigations. *Expert Systems with Applications*, 102, pp. 12-26. DOI: 10.1016/j.eswa.2018.02.023

[19] Aranganathan, A., Suriyakala, C.D. (2018). Agent based secure intrusion detection and prevention for rushing attacks in clustering MANETs. *International Journal of Engineering and Technology (UAE)*, 7(2), pp. 22-25. DOI: 10.14419/ijet.v7i2.20.11736

[20] Bajtos, T., Gajdoš, A., Kleinová, L., Lučivjanská, K., Sokol, P. (2018). Network Intrusion Detection with Threat Agent Profiling. *Security and Communication Networks*, 2018, art. no. 3614093. DOI: 10.1155/2018/3614093

[21] Nezarat, A. (2018). Distributed intrusion detection system based on mixed cooperative and non-cooperative game theoretical model. *International Journal of Network Security*, 20(1), pp. 56-64. DOI: 10.6633/IJNS.201801.20(1).07

[22] Rajeshkumar, G., Valluvan, K.R. (2017). An Energy Aware Trust Based Intrusion Detection System with Adaptive Acknowledgement for Wireless Sensor Network. *Wireless Personal Communications*, 94(4), pp. 1993-2007. DOI: 10.1007/s11277-016-3349-y

[23] Rajarajan, G., Ganesan, L. (2017). A Decoy Framework to Protect Server from Wireless Network Worms. *Wireless Personal Communications*, 94(4), pp. 1965-1978. DOI: 10.1007/s11277-016-3298-5

[24] Devanagavi, G.D., Nalini, N., Biradar, R.C. (2016). Secured routing in wireless sensor networks using fault-free and trusted nodes. *International Journal of Communication Systems*, 29(1), pp. 170-193. DOI: 10.1002/dac.2810

[25] Cherian, R.K., Shajin Nargunam, A. (2016). Distributed agent-based detection system using PSO and neural network for MANET. *International Journal of Mobile Network Design and Innovation*, 6(4), pp. 185-195. DOI: 10.1504/IJMNDI.2016.081659

[26] Newton, D.P. (2010). Quality and peer review of research: An adjudicating role for editors. *Accountability in Research*, 17, pp. 130-145. DOI: 10.1080/08989621003791945

[27] Haddaway, N.R., Bayliss, H.R. (2015). Shades of grey: Two forms of grey literature important for reviews in conservation. *Biological Conservation*, 191, pp. 827-829. DOI: 10.1016/j.biocon.2015.08.018

[28] van Eck, N.J., Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, 84, pp. 523-538. DOI: 10.1007/s11192-009-0146-3