DETECCIÓN DE INYECCIÓN DE DATOS EN SISTEMAS ROBÓTICOS ROS 2 MEDIANTE MÉTRICAS DE ENTROPÍA Y UN PIPELINE AUTOENCODER+LSTM

Jorge N. Gutiérrez[1-2], Bruno Lopes Dalmazo[1], Paulo Lilles Jorge Drews Junior[1]*
[1] Universidad Federal de Río Grande (FURG), Facultad de Ciencias de la computacion , Río Grande, RS, Brasil, jorge.gutierrez@utec.edu.uy
[2] Universidad Tecnológica del Uruguay (UTEC), Ingeniería en Mecatrónica, Fray Bentos, Uruguay.

Temática: Inteligencia artificial aplicada a la ciberseguridad en robótica autónoma.

PALABRAS CLAVE: ROS 2; detección de anomalías; métricas de entropía; Autoencoder+LSTM; series temporales.

1. INTRODUCCIÓN

Los sistemas robóticos distribuidos basados en ROS 2 requieren mecanismos de detección temprana de comportamientos anómalos que sean livianos y operables en tiempo real (Blázquez-García et al., 2021). La inyección de datos o las irregularidades en la publicación de tópicos comprometen seguridad, robustez y continuidad operativa. Aunque ROS 2 incorpora medidas de seguridad (p. ej., DDS Security), su sobrecarga puede ser no trivial; por eso se valoran estrategias complementarias de vigilancia que funcionen sobre telemetría interna, sin instrumentación de red ni criptografía pesada (Zhang et al., 2022; Fernández et al., 2018).

Este trabajo presenta un enfoque híbrido que combina métricas de entropía calculadas en ventanas deslizantes con un pipeline Autoencoder (AE) + LSTM entrenado exclusivamente con datos nominales del robot. Se emplean tres medidas entrópicas complementarias: Shannon (Shannon, 1948), transiciones (basada en diferencias sucesivas; Nardone, 2014) y KDE (aproximación no paramétrica de la densidad; Myers et al., 2025). Todas se calculan sobre ventanas de W = 100 muestras para capturar el comportamiento reciente. Además, se verifica la estacionariedad mediante ADF en segmentos nominales, lo que habilita calibración estadística de umbrales (Dickey y Fuller, 1979; Wang et al., 2023). El uso de AE y LSTM se apoya en representación robusta y modelado secuencial (Vincent et al., 2008; Hochreiter y Schmidhuber, 1997; Malhotra et al., 2016; Hundman et al., 2018).

El sistema se apoya en telemetría interna (p. ej., /cmd_vel, /odom, /laser_scan, /imu/data; y etiquetas /movement_label, /attack_type), y produce una decisión binaria con persistencia K y alertas (/ads/alert). En evaluación *offline*, con selección explícita del punto de operación bajo restricción de FPR, se obtienen altas tasas de detección y baja latencia, lo que indica viabilidad para despliegue en línea sobre ROS 2 (Abokhdair y Baig,

2025).

2. METODOLOGÍA

2.1. Entorno experimental y datos

Se utiliza el simulador Gazebo con un robot TurtleBot3 en un mundo vacío para construir línea base de normalidad y generar secuencias mixtas (normales y anómalas). La adquisición corre en Docker/ROS 2 con nodos de recolección, cálculo de entropías y detección. La tasa nominal es 10 Hz; para modelado se remuestrea a 20 Hz (interpolación breve en huecos), facilitando ventanas uniformes W=100 (~5 s).
El nodo recolector se suscribe a /cmd_vel, /laser_scan, /imu/data, /odom, /movement_label, /attack_type, deriva magnitudes cinemáticas (velocidades y aceleraciones lineal/angulares, magnitud de velocidad, jerk) y registra CSV con timestamp. El generador de anomalías introduce patrones (alta/baja frecuencia de comandos, *stop and go*, ráfagas), con fracción anómala global ~15–20 %.

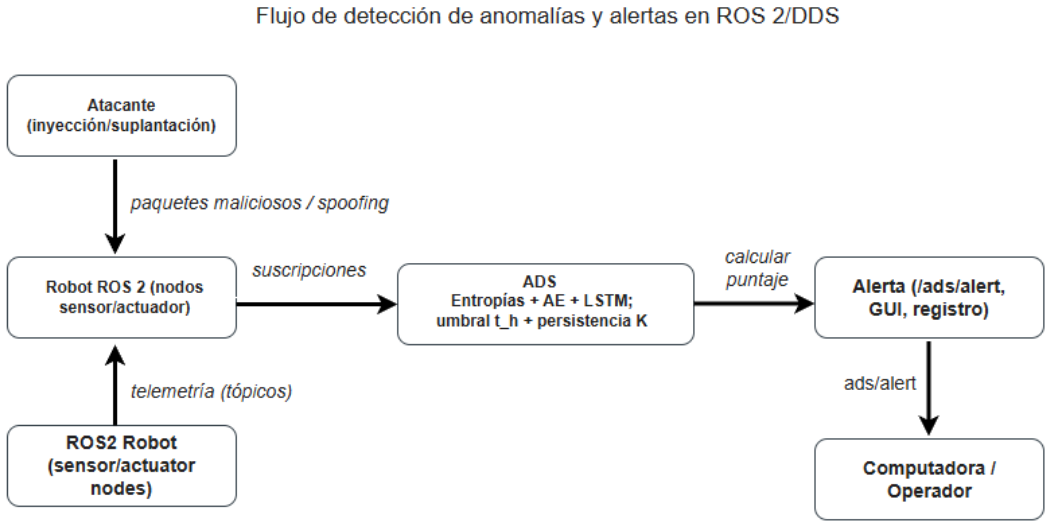

Figura 1. Flujo del ADS en ROS 2/DDS. El ADS suscribe telemetría, combina entropías+AE+LSTM y decide con umbral t_h y persistencia K; ante anomalía publica /ads/alert. Superficie de ataque indicada a la izquierda. (Fuente: elaboración propia).

2.2. Métricas de entropía y ventana deslizante

En cada ventana W=100 se calculan tres medidas. Shannon (H) cuantifica la incertidumbre global de la distribución de Δt (intervalos de llegada) u otras señales; la implementación usa conteos/probabilidades con corrección numérica para evitar log(0). La entropía de transiciones mide la incertidumbre de $\Delta x = x_{t+1} - x_t$, sensible a jitter, ráfagas e irregularidades breves. La entropía por KDE (H_{KDE}) estima densidad no paramétrica y usa un plug in de entropía diferencial, capturando cambios sutiles de forma (colas, multimodalidad)

(Myers et al., 2025). La actualización deslizante añade el dato reciente y descarta el más antiguo, manteniendo W constante y reflejando condiciones presentes.

2.3. Pipeline AE+LSTM y regla de decisión

El Autoencoder (sobrecompleto; regularización L2, dropout, early stopping; estandarización aprendida solo en entrenamiento) produce (i) error de reconstrucción y (ii) embedding latente para un LSTM que predice un paso. El puntaje de anomalía combina, con estandarización robusta (mediana/MAD) y pesos (α, β), el error del AE y el del LSTM (Vincent et al., 2008; Hochreiter y Schmidhuber, 1997; Malhotra et al., 2016). La decisión aplica umbral t_h con persistencia *K* para reducir falsos disparos. La calibración offline recorre una rejilla de hiperparámetros bajo restricción de FPR . (Tatbul et al., 2018).

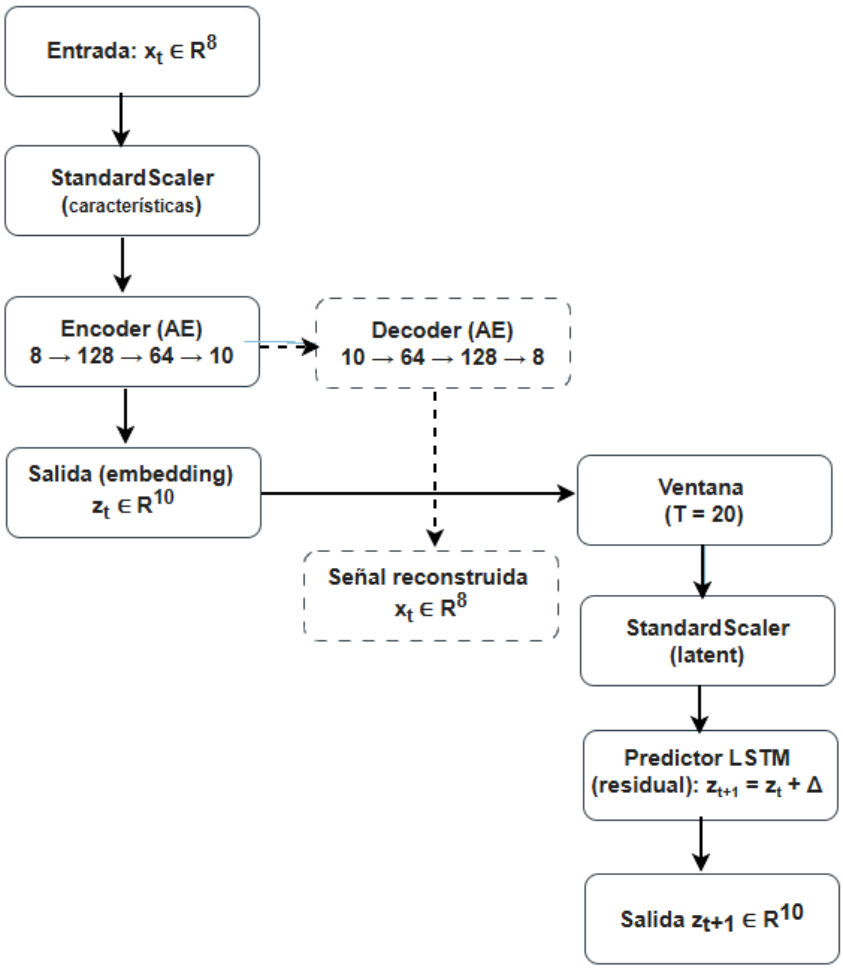

Figura 2. Flujo del pipeline AE+LSTM: el encoder transforma x_t (R^8) en z_t (R^{10}); el decoder permite reconstrucción; ventanas de tamaño T=20 alimentan el LSTM que estima z_{t+1}; el puntaje combina errores de reconstrucción y predicción. (Fuente: elaboración propia).

2.4. Preprocesamiento y estabilidad

El conjunto comprende ~15,8 k registros. Se filtran segmentos nominales, se aplica burn in (~600 muestras) y se validan integridad y orden temporal. Se analizan variabilidad media y desviación estándar de las entropías: KDE es más sensible a fluctuaciones (útil para cambios sutiles), mientras Shannon y transiciones aportan estabilidad y sensibilidad complementarias (Myers et al., 2025; Shannon, 1948; Nardone, 2014). La ADF sobre nominal respalda calibración de umbrales bajo supuestos de estacionariedad (Dickey y Fuller, 1979; Wang et al., 2023).

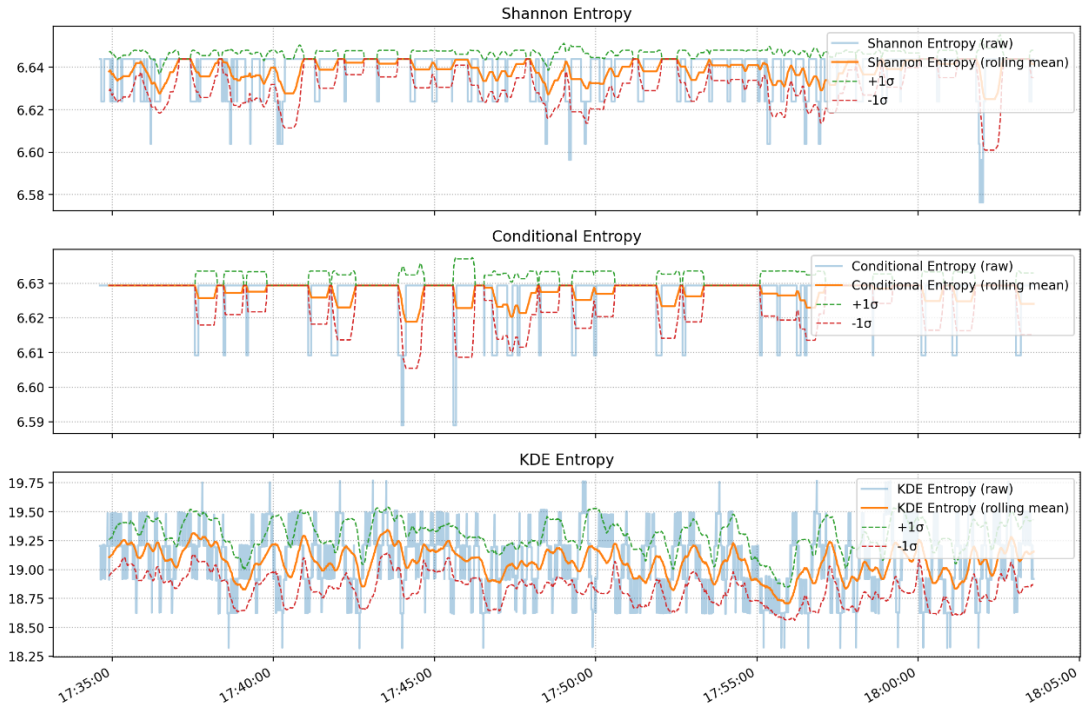

Figura 3. Evolución temporal de las entropías (Shannon, transiciones y KDE) en datos nominales: señal cruda, media móvil y bandas ±1σ. (Fuente: elaboración propia).

3. RESULTADOS Y DISCUSIÓN

3.1. Selección del punto de operación offline

Bajo una búsqueda en rejilla y con restricción explícita de FPR, la relación FPR-recall observada en el barrido de parámetros (Tatbul et al., 2018) se muestra en la Figura 3.

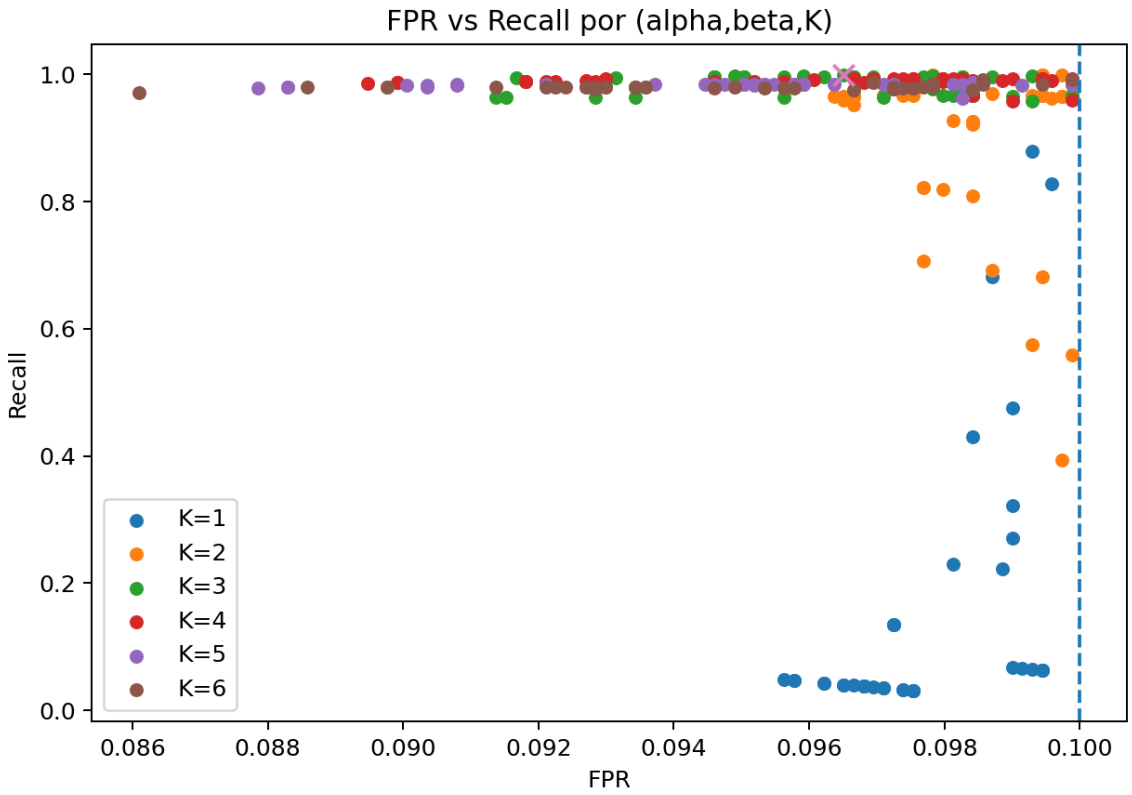

Figura 3 – Barrido de parámetros FPR vs. Recall

Con base en ese barrido se seleccionó el punto (α = 1.1, β = 1.4, K = 3) (véase Tabla 1)

Parámetro	**α^***	**β^***	**K^***	**t_h^***	**Prevalencia π**	**Ventana T**
Valor	1.1	1.4	3	15.4413	0.098	20

Tabla 1 – Punto de operación seleccionado Fuente: elaboración propia.

La comparación entre valores de persistencia K bajo FPR ≤ 0.10 se resume en la Tabla 2.

K	**α**	**β**	**t_h**	**Precisión**	**Recall**	**F1**	**FPR**
1	0.500	2.000	25.385444	0.491	0.880	0.630	0.099
2	0.500	2.000	19.219507	0.526	0.999	0.689	0.098
3†	1.100	1.400	15.441346	**0.529**	**0.999**	**0.692**	**0.097**
4	1.100	1.400	15.441346	0.537	0.993	0.697	0.093
5	1.500	1.000	12.477342	0.519	0.992	0.682	0.100
6	1.500	0.800	10.423824	0.522	0.992	0.684	0.099

Tabla 2 – Mejor combinación por K. Fuente: elaboración propia

En la evaluación punto-a-punto (π ≈ 0.098; N = 7 559), se obtiene recall ≈ 0.999, FPR ≈ 0.097 y F1 ≈ 0.692. A nivel de evento, el *hit rate* es 100 %, con latencia media ≈ 0.27

ventanas (mediana 0) y cobertura media ≈ 0.908. La persistencia K atenúa saltos espurios y privilegia desviaciones sostenidas, adecuado para despliegues en línea.

	Pred. Pos	**Pred. Neg**
Real Pos	740 (TP)	1 (FN)
Real Neg	658 (FP)	6160 (TN)

Tabla 3. Matriz de confusión y métricas (mejor global). Parámetros: α=1.1, β=1.4, K=3, th=15.44; N=7 559, π ≈ 0.098; P=0.529, Recall=0.999, F1=0.692; FPR=0.097 (TNR=0.903); Accuracy=0.913. (Fuente: elaboración propia).

3.2. Análisis de sensibilidad de métricas

Las tres métricas aportan señales complementarias: Shannon enfatiza cambios en dispersión/uniformidad de Δt y otras señales; transiciones captura dinámica local (*jitter*, ráfagas, desorden breve); KDE resalta cambios de forma (colas, multimodalidad) (Shannon, 1948; Nardone, 2014; Myers et al., 2025). Su combinación amplía la cobertura ante anomalías heterogéneas (inyección de comandos, desincronizaciones, cambios de régimen) y reduce la dependencia de un único indicador (Blázquez-García et al., 2021).

3.3. Ventajas prácticas

El enfoque es ligero y agnóstico de red: opera sobre tópicos y realiza inferencia local, viable en recursos modestos. La ventana deslizante W=100 a 20 Hz (~5 s) equilibra sensibilidad/estabilidad temporal. La calibración se apoya en ADF (segmentos nominales) y en rejilla con FPR constreñido, lo que facilita un ajuste reproducible con criterio operativo (Dickey y Fuller, 1979; Tatbul et al., 2018). La integración con ROS 2/DDS es directa: el ADS consume telemetría, decide y publica /ads/alert, integrándose con GUI y registros (Zhang et al., 2022; Fernández et al., 2018).

4. CONCLUSIONES

Se presentó un esquema híbrido para detección de inyección de datos en ROS 2 que integra métricas de entropía (Shannon, transiciones y KDE) con un pipeline AE+LSTM entrenado con normalidad (Shannon, 1948; Vincent et al., 2008; Hochreiter y Schmidhuber, 1997). La ADF y la calibración bajo restricción de FPR permiten fijar un punto de operación con recall casi perfecto, latencia baja y FPR controlado (Dickey y Fuller, 1979; Tatbul et al., 2018). El enfoque es ligero, no intrusivo y agnóstico del tráfico de red, lo que facilita su despliegue en línea. (Abokhdair y Baig, 2025; Zhang et al., 2022).

Limitaciones y líneas de mejora: la evaluación es *offline* y en un mundo vacío con un único robot y anomalías sintéticas; los resultados dependen del remuestreo a 20 Hz y del tamaño de W=100; la umbralización es estática; no se analizó el impacto de jitter de red ni de perfiles QoS heterogéneos. Como trabajo futuro, se propone validar en escenarios reales o simulados más complejos (obstáculos, interacción rica), desplegar online con

medición extremo-a-extremo de latencia/jitter, incorporar umbrales adaptativos y detección de *drift* con aprendizaje continuo, y ampliar variables sensoriales/cinemáticas y estados de misión. En aplicaciones, el ADS puede operar como watchdog que publica /ads/alert hacia GUI y registros, habilita contención (p. ej., *safe stop* o degradación controlada) y actúa como capa de vigilancia complementaria en arquitecturas robóticas distribuidas.

REFERENCIAS

Abokhdair, M., & Baig, Z. (2025). A deep learning–based intrusion detection system for ROS 2. *In Proceedings of the International Joint Conference on Neural Networks (IJCNN).* IEEE. (En prensa)

Ahmed, S., Khan, S., & Mian, A. (2023). Entropy-SLAM: Improving robustness and resilience of SLAM systems using entropy-based anomaly detection. *In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS).* IEEE.

Blázquez-García, A., Conde, A., Mori, U., & Lozano, J. A. (2021). A review on outlier/anomaly detection in time series data. *ACM Computing Surveys, 54*(3), 1–33.

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association, 74*(366), 427–431.

Fernández, G., Rossi, F. A. E., & García, F. J. (2018). Security and performance considerations in ROS 2: A balancing act. *arXiv preprint* arXiv:1809.09566.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation, 9*(8), 1735–1780.

Hundman, K., Constantinou, V., Laporte, C., Colwell, I., & Soderstrom, T. (2018). Detecting spacecraft anomalies using LSTMs and nonparametric dynamic thresholding. *In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 387–395). ACM.

Malhotra, P., Ramakrishnan, A., Anand, G., Vig, L., Agarwal, P., & Shroff, G. (2016). LSTM-based encoder–decoder for multi-sensor anomaly detection. *arXiv preprint* arXiv:1607.00148.

Myers, A., Kay, B., Alvarez, I., Hughes, M., Mackenzie, C., Ortiz Marrero, C., Ellwein, E., & Lentz, E. (2025). Entropic analysis of time series through kernel density estimation. *arXiv preprint* arXiv:2503.18916.

Nardone, P. (2014). Entropy of difference. *arXiv preprint* arXiv:1411.0506.
Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal, 27*, 379–423, 623–656.

Tatbul, N., Lee, T. J., Zdonik, S., Alam, M., & Gottschlich, J. (2018). Precision and recall for time series. *In Advances in Neural Information Processing Systems (NeurIPS).*

Vincent, P., Larochelle, H., Bengio, Y., & Manzagol, P.-A. (2008). Extracting and composing robust features with denoising autoencoders. *In Proceedings of the 25th International Conference on Machine Learning* (pp. 1096–1103). ACM.

Wang, Y., Zhang, M., & Li, R. (2023). Entropy-based anomaly detection in time series using ADF stationarity validation. *IEEE Transactions on Industrial Informatics, 19*(4), 4450–4461.

Zhang, T., Shangguan, L., Su, Q., & Zhou, Y. (2022). On the (in)security of secure ROS2. *In Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security (CCS)* (pp. 1929–1944). ACM.