UNIVERSIDAD DE LAS FUERZAS ARMADAS ESPE

Departamento de Ciencias de la Computación

Lectura y Escritura de Textos Académicos

Tema:

Propuesta de Desarrollo y Solución

Autor:

Mateo Medranda
Bryan Quispe
Moisés Benalcázar

NRC: 29765

Ecuador 2025-12-01

1. Sistemas de Detección de Intrusiones (IDS)

Un sistema de detección de intrusiones o "IDS" es un instrumento fundamental para la seguridad y protección de dispositivos IoT (Internet of Things o internet de las cosas), el cual está diseñado para identificar ataques externos e internos a redes o sistemas informáticos, incluso superado el rendimiento de un firewall tradicional [1][2]. El funcionamiento principal de un IDS es recolectar y procesar datos para luego producir alarmas mediante un mecanismo, que serán enviadas a un analista de redes para luego ser investigada a fondo [3].

Los IDS se clasifican en cuatro métodos: IDS basado en firmas (SIDS), IDS basado en anomalías (AIDS), IDS basado en especificaciones (SpIDS) y Sistemas de IDS Híbridos (HyIDS), pero el IDS aplica dos tipos de métodos de detección: métodos basados en firmas y métodos basados en anomalías [1].

IDS basado en Firmas (SIDS) se caracteriza como detección de abuso, la cual comprueba las firmas de intrusión con patrones conocidos previamente almacenados en una base de datos, su base de conocimiento va previo a los ataques mediante reglas codificadas con una detección de ataques conocidos muy eficaz y preciso ya que tiene tasas bajas de falsos positivos (FP) y ante ataques desconocidos muy ineficaz siendo incapaz de detectar así como una baja tarifa de FP como punto a favor; por otro lado el IDS basado en Anomalías (AIDS) denominada como detección basada en comportamiento, monitorea el registro de actividades de un sistema y reporta desviaciones del comportamiento normal, por otro lado tiene un aprendizaje automático, creando un modelo de tráfico normal el cual es capaz de detectar desviaciones (ataques conocidos) y ante los ataques desconocidos puede detectar ataques internos y ataques de día cero, estos últimos son nuevas vulnerabilidades que aún no tienen parches, lo que lo lleva a tener un FP típicamente alto ya que el tráfico anormal no siempre es un ataque [1].

Los AIDS son cruciales porque los vectores de ataques conocidos y principalmente los desconocidos representan una alta amenaza en el tema cibernético en la actualidad, los AIDS funcionan construyendo una distribución estadística de patrones de intrusión mediante un análisis profundo de los datos transmitidos para identificar el malware, por eso el AIDS es una metodología el cual nos permite: Identificar la evolución del ataque con la detección de nuevos ataques y malware, la detección de amenazas internas y tener una detección basada en el comportamiento como un tráfico que se desvía significativamente del comportamiento normal esperado [1].

1.2.Taxonomía de Ataques Comunes en Datasets

Los IDS se suelen evaluar y entrenar utilizando datasets que contienen ejemplos de tráfico benigno y trafico maligno, aunque la mayoría de veces están desequilibrados con un 98% de datos de trafico normal y el otro 2% clasificado como ataques. A continuación, se mostrará los tipos de ataques que frecuentemente se encuentran en los datasets enfocados al tema de ciberseguridad [1]:

- Los ataques de fuerza bruta o Brute Force son un tipo de ataque de contraseña. Su objetivo es obtener acceso no autorizado al sistema mediante el uso de técnicas de adivinación para robar contraseñas.
- Los ataques Red de bots o Botnet es un conjunto de dispositivos comprometidos controlados de forma remota, utilizado a menudo como vector para ataques de Denegación de Servicio DoS y DDoS (Denegación de Servicio y Denegación de Servicio Distribuido)
- Los ataques DoS son ataques que se basan en bloquear temporalmente el uso normal de las utilidades de red inundando la red con tráfico y suelen exhibir patrones secuenciales e implican un gran número de conexiones al mismo host.
- Los ataques DDoS estos se basan en inundar el servidor y hacer que no pueda responder sobrecargándolo con solicitudes de servicio, pero, a diferencia de los ataques DoS, la inundación se realiza a través de múltiples fuentes y tradicionalmente son ataques de alta frecuencia que inundan el ancho de banda; sin embargo, los ataques DDoS en la capa de aplicación son ataques de baja frecuencia que inundan el servidor en su lugar.
- El ataque de infiltración implica obtener acceso a la red, a menudo desde el interior
- Los ataques de sondeo consisten en identificar puntos débiles dentro de la red mediante el envío de paquetes destinados al escaneo y la recopilación de datos sobre los sistemas. Entre los tipos más comunes se encuentran los ataques de Satanás, el barrido de direcciones IP y el barrido de puertos.
- Por otro lado, los ataques de inyección emplean scripts que insertan comandos o consultas maliciosas con el objetivo de lograr acceso no autorizado y extraer información. Entre sus variantes más conocidas están la inyección SQL y los ataques de secuencias de comandos entre sitios (XSS).

1.3. Caracterización del Dataset CIC-IDS-2018

El data set CSE-CICIDS2018 es un conjunto de datos que justifica ampliamente a otros conjuntos de datos mas antiguos como el KDD99 o el NSL-KDD, esto debido a la evolución constante de los ciber ataques [3] [4].

Las características de red se extraen del tráfico de red, ya sea a nivel de paquete en el que llega a usar metodologías como Full Packet CAPture – PCAP o o a nivel de flujo utilizando protocolos como NetFlow [3]. Mediante las fuentes podemos identificar las siguientes características relevantes:

- Los puertos, los números de puertos tanto de origen como destino son datos detallados, también los números de puertos son utilizados como "claves de flujo" o Flujo Key para así poder definir un conjunto de paquetes que comparten propiedades comunes.
- Las banderas se extraen de los paquetes de red completos a través de PCAP. Son características dispersas utilizadas en la detección de ataques de red. El análisis de banderas ayuda a identificar ataques de Denegación de Servicio (DoS). Además, son esenciales para detectar ataques de sondeo o escaneo, ya que proporcionan información sobre el comportamiento de la red.
- La duración de conexiones es clave en la detección de ciertos tipos de ataques, como los ataques de usuario a root (U2R) o de remoto a local (R2L). En conjuntos de datos como el KDD99, la duración y las características de servicio son especialmente útiles para identificar estos ataques.
- El desbalance de clases es un desafío fundamental en la ciberseguridad y los Sistemas de Detección de Intrusiones (IDS), el problema surge de la naturaleza inherente de los conjuntos de datos de seguridad, donde el tráfico normal o benigno constituye la clase mayoritaria, y los ataques o tráfico malicioso representan la clase minoritaria.

2.1 Machine Learning: Random Forest (RF)

El algoritmo Random Forest (RF) se selecciona como el modelo base (*benchmark*) para este estudio debido a su alta eficacia en la clasificación de intrusiones en datasets complejos como CICIDS2017 [2], [7].

- Fundamentos y Ensemble Learning (Bagging): RF es un modelo de aprendizaje supervisado que emplea la técnica de *Bagging* (Bootstrap Aggregating). El algoritmo construye múltiples árboles de decisión entrenados independientemente con subconjuntos

aleatorios de datos. La clasificación final se obtiene mediante votación mayoritaria, optimizando la división de nodos a través de métricas de impureza como Gini o Entropía [1], [7].

- Robustez y Mitigación de Overfitting: A diferencia de los árboles de decisión individuales, RF reduce la varianza y evita el sobreajuste (*overfitting*) al promediar los resultados de múltiples árboles descorrelacionados y al seleccionar aleatoriamente un subconjunto de características en cada división, lo cual es crucial para manejar el ruido en el tráfico de red [1], [2].
- Hiperparámetros: Para controlar la complejidad del modelo, se ajustan principalmente el número de estimadores ($n\_estimators$), que define la cantidad de árboles, y la profundidad máxima (max_depth), que limita el crecimiento del árbol para mantener la generalización [7].

2.2 Deep Learning: Convolutional Neural Networks (CNN)

Las Redes Neuronales Convolucionales (CNN) constituyen la propuesta de modelo complejo, justificadas por su capacidad para extraer automáticamente características latentes y jerárquicas en los datos, superando la ingeniería manual de atributos [3], [9].

- Arquitectura y Adaptación 1D: Aunque típicas en visión por computador, este estudio implementa una 1D-CNN para procesar el tráfico de red. El modelo interpreta los vectores de características como secuencias espaciales (matrices 1D), permitiendo que los filtros convolucionales detecten correlaciones locales y dependencias entre atributos adyacentes del flujo de red [5], [9].
    - Capas: Se compone de capas *Convolucionales* (extracción de patrones), *Pooling* (reducción de dimensionalidad y costo computacional), *Flatten* (vectorización) y capas *Dense* (clasificación final) [4], [5].
- Funciones de Activación: Se emplea ReLU en las capas ocultas para acelerar la convergencia y mitigar el desvanecimiento del gradiente, y Softmax en la capa de salida para generar una distribución de probabilidad sobre las clases de ataques (multiclase) [6], [8].

3. Técnicas de Preprocesamiento e Ingeniería de Características

El preprocesamiento es la aplicación de una serie de operaciones sobre un conjunto de datos, con el objetivo de disminuir el tiempo de entrenamiento de un modelo y mantener la objetividad en la evaluación [5]. Dado que la validez de un algoritmo de aprendizaje automático depende directamente de la calidad de sus datos, es un proceso obligatorio para realizar. Dentro de datasets como el CICIDS2018, los datos crudos pueden contener información inconsistente, lo que requiere de una fase de normalización y codificación para minimizar cualquier impacto o discrepancias en un modelo desarrollado [6].

La alta dimensionalidad de los datos representa un problema para el entrenamiento de un modelo de detección de intrusiones, dado que suelen haber entre 79 y 80 características por muestra [7], además este puede ser un problema al momento de generar datos sintéticos ya que se produce ruido además de resaltar los atributos irrelevantes impactando en la capacidad de generalización y la estabilidad del modelo [6].

3.1. Limpieza y Codificación de los Datos

En conjuntos de datos masivos con información sobre el tráfico de red se puede albergar ruido inherente, lo que permite la aparición de datos nulos, infinitos e incluso valores negativos respecto al tiempo de flujo del tráfico de red, por lo que se presentan características que complican el procesamiento y la clasificación [7].

La limpieza de datos se enfoca en la eliminación de valores NaN, redundantes, duplicados, valores infinitos e inconsistencias. De esta forma se elimina el dato no deseado y se mejora el rendimiento [8].

Por otro lado, la codificación de los datos se realiza sobre cada categoría, preservando la eficiencia de la memoria y el procesamiento de un algoritmo matemático, ya que se requiere de entradas numéricas. Ciertamente una herramienta de codificación también se refiere a One-Hot, la cual genera vectores ortogonales binarios o matrices para la optimización del modelo [9].

3.2. Estandarización y Normalización

Si bien Random Forest no presenta problemas con la escala de ciertas características, como la duración del flujo o el número de paquetes dentro del tráfico de red, otros modelos como CNN son sensibles a estos datos de entrada con rangos dispares, lo cual puede perjudicar el aprendizaje de patrones pequeños [8], [10]. La solución para la normalización corresponde al uso de Min-Max (ecuación 1) mientras que para estandarización se puede utilizar Z-score.

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

3.3. SMOTE para Desbalance de Clases

Dentro del Dataset CICIDS2018 se presenta un desbalance significativo en las clases, esto se debe a que los registros del tráfico benigno superan de forma masiva a los registros de intrusiones, teniendo valores de cerca de 10 millones de registros contra aproximadamente 2 millones de registros respectivamente [9][6],[1]. Este problema puede generar sesgos inclinados a la clase dominante, y para contrarrestar este efecto, aparece la técnica SMOTE (*Synthetic Minority Over-sampling Technique*), donde se sintetizan nuevas instancias de las clases minoritarias basado en el funcionamiento del algoritmo KNN donde se buscan los k vecinos más cercanos y se realiza una interpolación para obtener nuevos valores, pero sin perder la varianza intrínseca de los datos.

Bibliografía

[1] M. S. Ahsan, S. Islam, and S. Shatabda, "A systematic review of metaheuristics-based and machine learning-driven intrusion detection systems in IoT," *Swarm Evol Comput*, vol. 96, Jul. 2025, doi: 10.1016/j.swevo.2025.101984.

[2] Z. K. Maseer, R. Yusof, N. Bahaman, S. A. Mostafa, and C. F. M. Foozy, "Benchmarking of Machine Learning for Anomaly Based Intrusion Detection Systems in the CICIDS2017 Dataset," *IEEE Access*, vol. 9, pp. 22351–22370, 2021, doi: 10.1109/ACCESS.2021.3056614.

[3] D. Gümüşbaş, T. Yıldırım, A. Genovese, and F. Scotti, "A Comprehensive Survey of Databases and Deep Learning Methods for Cybersecurity and Intrusion Detection Systems," *IEEE Syst J*, vol. 15, no. 2, pp. 1717–1731, 2021, doi: 10.1109/JSYST.2020.2992966.

[4] W. H. Aljuaid and S. S. Alshamrani, "A Deep Learning Approach for Intrusion Detection Systems in Cloud Computing Environments," *Applied Sciences (Switzerland)*, vol. 14, no. 13, 2024, doi: 10.3390/app14135381.

[5] V. Hnamte, A. A. Najar, H. Nhung-Nguyen, J. Hussain, and M. N. Sugali, "DDoS attack detection and mitigation using deep neural network in SDN environment," *Comput Secur*, vol. 138, Mar. 2024, doi: 10.1016/j.cose.2023.103661.

[6] K. Zhang *et al.*, "SE-DWNet: An Advanced ResNet-Based Model for Intrusion Detection with Symmetric Data Distribution," *Symmetry (Basel)*, vol. 17, no. 4, Apr. 2025, doi: 10.3390/sym17040526.

[7] M. Deng, C. Sun, Y. Kan, H. Xu, X. Zhou, and S. Fan, "Network Intrusion Detection Based on Deep Belief Network Broad Equalization Learning System," *Electronics (Switzerland)*, vol. 13, no. 15, Aug. 2024, doi: 10.3390/electronics13153014.

[8] R. Bingu, S. Adinarayana, J. S. Dhatterwal, S. Kavitha, E. Patnala, and H. R. Sangaraju, "Performance comparison analysis of classification methodologies for effective detection of intrusions," *Comput Secur*, vol. 143, 2024, doi: 10.1016/j.cose.2024.103893.

[9] M. Deng, C. Sun, Y. Kan, H. Xu, X. Zhou, and S. Fan, "Network Intrusion Detection Based on Deep Belief Network Broad Equalization Learning System," *Electronics (Switzerland)*, vol. 13, no. 15, Aug. 2024, doi: 10.3390/electronics13153014.

[10] S. V. Pingale and S. R. Sutar, "Remora whale optimization-based hybrid deep learning for network intrusion detection using CNN features," *Expert Syst Appl*, vol. 210, p. 118476, Dec. 2022, doi: 10.1016/J.ESWA.2022.118476.