Financiado por la Unión Europea NextGenerationEU
Gobierno de España · Vicepresidencia Primera del Gobierno · Ministerio de Asuntos Económicos y Transformación Digital · Secretaría de Estado de Digitalización e Inteligencia Artificial
Plan de Recuperación, Transformación y Resiliencia
España | digital 2026
incibe_ Instituto Nacional de Ciberseguridad

Alicia Mayor Bal
presenta el resumen del trabajo

Una nueva propuesta para la detección y clasificación de ciberataques basada en ensemble learning

Óscar Mogollón Gutiérrez, Javier Alonso Díaz, José Carlos Sancho Núñez, Andrés Caro Lindo

Actas VIII JNIC, atlanTTic, 69-76, (2023).

10 de julio de 2024

SERIE DE RESÚMENES EN ESPAÑOL

Palabras clave: Ciberataques, Deep learning, Ensemble, Machine learning, Modelos binarios, Sistemas de detección de intrusiones.

Introducción

El incremento de ciberataques en los últimos años hace necesario adoptar un **enfoque preventivo en ciberseguridad**. Los sistemas de detección de intrusiones (IDS) ayudan a identificar tanto el uso indebido de sistemas como anomalías en su comportamiento. También la inteligencia artificial ayuda a la detección de intrusiones eficaz, posibilitando el modelado inteligente de sistemas según su comportamiento.

Este trabajo presenta un sistema de clasificación mediante técnicas de aprendizaje automático y profundo, detectanto en primer lugar si el tráfico es anómalo, y, en caso positivo, clasificando el tipo de ataque. Estas dos fases permiten reducir el desequilibrio entre clases, problema visto frecuentemente en la detección de intrusiones.

1. Metodología

En primer lugar, para el **preprocesado de datos**, las características numéricas han sido normalizadas y las características categóricas se han codificado utilizando etiquetas numéricas.

A continuación se genera un **modelo de clasificación binaria** para cada tipo de tráfico, que distingue entre ese tipo y el resto. Para esto se crean conjuntos de datos binarios compuestos por muestras de la categoría objetivo así como del resto, garantizando si fuese necesario mediante sobremuestreo que cada categoría tiene representación suficiente.

Para cada categoría, se **genera el modelo binario** mediante un ajuste de hiperparámetros junto con un preprocesado para cuatro algoritmos de clasificación: algoritmo k-medidas, máquinas vectoriales, árboles de decisión y perceptrón multicapa. Se selecciona el mejor modelo según la métrica de evaluación Macro-F1, que combina la precisión y la recuperación en un único valor.

La clasificación del tráfico se realiza mediante un **modelo ensemble** que primero identifica el comportamiento como Normal o Ataque y a coninuación clasifica la amenaza como uno de los posibles ataques. Cada clasificador obtiene la probabilidad de pertenecer a su clase o al resto y se toma como predicción final la clase con el valor más alto.

Para resolver la problemática del desequilibrio de categorías, es necesario el uso de **conjuntos de datos fiables**. En este caso se han utilizado los conjuntos NSL-KDD y UNSW-NB15.

2. Resultados

2.1. Conjunto NSL-KDD

En este conjunto se evalúan **cinco tipos de tráfico**, para cada uno de los cuales y en base a la obtención de resultados se asignan los siguientes modelos binarios: SVM para DoS, DT para Probe, R2L y U2R, y MLP para tráfico Normal. Tras la aplicación del modelo ensemble, se alcanza una métrica**F1 de 0,7213**, superior a la mejor puntuación obtenida por otro trabajo científico en el campo de detección de intrusiones para este conjunto de datos.

2.2. Conjunto UNSW-NB15

En este conjunto se evaúan **diez tipos de tráfico**: Analysis, Backdoor, DoS, Exploits, Fuzzers, Generic, Normal, Recconnaissance, Shellcode y Worms. El tráfico mejor detectado fue el Genérico con un valor de 0,9835, mientras que los ataques con menor detección fueron DoS, Exploits y Fuzzers con alrededor de

0,8. Independientemente de la categoría, la detección de ataques supera el 0,91. Globalmente, se alcanza un **F1 de 0,7754** para la propuesta en este conjunto.

3. Discusión/Conclusiones

El sistema de clasificación propuesto para ambos conjuntos de datos en los que se ha probado **mejora en términos de F1** otras propuestas en la literatura científica. Además, la aplicación ensemble junto al esquema en dos pasos clasifica eficazmente el tráfico en escenarios donde existe desbalanceo notable.

4. Valoración del documento original

Este paper hace uso de técnicas usadas en otras investigaciones para el mismo campo como son el aprendizaje automático y profundo pero añade el modelo ensemble, con la intención de solucionar la problemática del desbalanceo de clases, muy común en otros estudios del campo. Esto es interesante, ya que no sólo se intenta llegar a una solución, sino añadir una capa propia más para mejorar los resultados. La metodología sigue un esquema ordenado, pasando por todas las partes necesarias para entender el proceso, y los resultados son bastante específicos, comprobando el modelo propuesto en varias ocasiones lo que ofrece posibilidad de comparación.