Curso: Deep Learning
Semestre: 2022-01
Estudiante: Omar Augusto Roa Romero
Correo electrónico: omar.roa@udea.edu.co
Programa: Maestría en Ingeniería de Telecomunicaciones
Entrega: 1

CONTEXTO

Las Infraestructuras Críticas (IC) representan sistemas de alto impacto como servicios públicos, transporte y comunicaciones, entre otros. Dichos sistemas tienen una interoperabilidad tan estrecha que llegan a convertirse en un gran “sistema de sistemas” donde cualquier falla en uno de ellos, impacta ampliamente a los demás. La IC Eléctrica tiene un alto impacto en las demás por la dependencia intrínseca de los sistemas eléctricos en los procesos básicos.

Los sistemas eléctricos han ido implementando flujos de electricidad e información evolucionando en Smart Grids, las redes eléctricas del siglo XX. Esto ha implicado la adopción de tecnologías tradicionales de las redes de datos para transmitir información y así mejorar las fases de generación, transmisión, distribución y consumo.

Dicha integración ha heredado los problemas de ciberseguridad propios de las redes de datos que se han convertido en foco de interés en los últimos años. Los ataques a los sistemas eléctricos de acuerdo con su impacto en otros sistemas aparecen clasificados en primer lugar en tres de cuatro categorías dentro de 16 ICs en estudio. De acuerdo con el Foro Económico Mundial, los ciberataques en las ICs se han ubicado como el quinto riesgo económico más alto en 2020 y ha definido este hecho como “la nueva normalidad en sectores como la energía, la salud y el transporte”.

Las Smart Grids como IC presentan diferentes ataques tales como denegación de servicio, reprogramación no autorizada de equipos, alteración de registros, modificación de políticas de seguridad, entre otros. Además, las soluciones tradicionales como firewalls, IPS y otras implementaciones basadas en reglas no se ajustan a estas ya que presentan características propias y conllevan a buscar nuevas estrategias.

Las Redes Definidas por Software (SDN) presenta una solución para alcanzar los diferentes requerimientos de las Smart Grids en relación con conectividad, retardos, sincronización, ancho de banda y seguridad.

Dentro de mi trabajo de maestría se debe definir si generar un dataset desde un testbed propios o recurrir a alguno ya publicado. La búsqueda aún continúa para encontrar uno que se ajuste a los requerimientos del trabajo propuesto en maestría, que son las subestaciones digitales bajo el estándar IEC 61850.

OBJETIVO

Identificar tráfico malicioso basado en las características propias de los flujos y detalles incluidos en los encabezados dentro de una Smart Grid.

DATASET

Debido a la escasez de datasets para subestaciones eléctricas digitales con el estándar IEC 61850, he decidido comenzar mi trabajo en el curso con un dataset público similar, identificando las estrategias para en un futuro cercano poder aplicarlas al dataset seleccionado para mi trabajo de maestría.

El dataset seleccionado está publicado en **kaggle** con los siguientes detalles:

- **Nombre**: SDN Intrusion Detection
- **Descripción**: Intrusion Detection in Software Defined Network with 4 Attack Categories
- **URL**: https://www.kaggle.com/datasets/subhajournal/sdn-intrusion-detection
- **Tamaño en disco**: 426.85 MB
- **Observaciones**: 1.188.333
- **Características**: 79 cuantitativas y 1 cualitativa
- **Clases**:
    - Tráfico BENIGN con 798322 observaciones
    - Tráfico DDoS con 383439 observaciones
    - Tráfico Web Attack Brute Force con 4550 observaciones
    - Tráfico Web Attack XSS con 1962 observaciones
    - Tráfico Web Attack Sql Injection con 60 observaciones

MÉTRICAS

Como es ampliamente usado en sistemas de detección de intrusiones, utilizaré como métrica algunas de las más populares:

Precision expresa qué proporción de muestras que se clasifican como comportamiento malicioso, de hecho presentan un comportamiento malicioso.

$$Precision = \frac{TP}{TP\ +\ FP}$$

Accuracy es el número de casos detectados correctamente en la muestra de flujo total.

$$Accuracy = \frac{TP\ +\ TN}{TN + TP + FP + F\ N}$$

Existen otras métricas como Recall, Selectivity y F1_Score que pueden estar en posibilidad de utilizarse como métrica para este trabajo.

Donde:

- Verdaderos positivos (TP) se consideran como el número de clasificaciones correctas que detectaron los ciberataques como comportamiento anormal.
- Verdaderos negativos (TN) se identifican como el número de clasificaciones correctas que reconocieron actividades no maliciosas como comportamiento normal.
- Falsos positivos (FP) se consideran como el número de clasificaciones incorrectas que identificaron actividades no maliciosas como comportamiento anormal
- Falsos negativos (FN) se consideran como el número de clasificaciones incorrectas que reconocieron los ciberataques como un comportamiento normal

REFERENCIAS

A. Dawoud, S. Shahristani, y C. Raun, “Deep learning and software-defined networks: Towards secure IoT architecture”, Internet of Things, vol. 3–4, pp. 82–89, oct. 2018.

G. Efstathopoulos et al., “Operational data based intrusion detection system for smart grid”, IEEE Int. Work. Comput. Aided Model. Des. Commun. Links Networks, CAMAD, vol. 2019-Septe, 2019.

S. Gómez Macías, L. P. Gaspary, y J. F. Botero, “ORACLE: Collaboration of Data and Control Planes to Detect DDoS Attacks”, Proc. IM 2021 - 2021 IFIP/IEEE Int. Symp. Integr. Netw. Manag., pp. 962–967, sep. 2020.

P. I. Radoglou-Grammatikis y P. G. Sarigiannidis, “Securing the Smart Grid: A Comprehensive Compilation of Intrusion Detection and Prevention Systems”, IEEE Access, vol. 7, pp. 46595–46620, 2019.

S. Sengan, S. V, I. V, P. Velayutham, y L. Ravi, “Detection of false data cyber-attacks for the assessment of security in smart grid using deep learning”, Comput. Electr. Eng., vol. 93, núm. May, p. 107211, jul. 2021.