Universidad del Valle de Guatemala
Facultad de Ingeniería
Departamento de Ciencias de la Computación
CC3094 - Security Data Science - Sección 10
Catedrático: JORGE ANDRES YASS COY
Ciclo I de 2023

FASE 1
DDoS Detection

Integrantes:
JULIO ROBERTO HERRERA SABAN 19402
OLIVER JOSUE DE LEON MILIAN 19270
RANDY SAMUEL VENEGAS LORENTI 18341

Guatemala, 28 de febrero de 2023

MOTIVACIÓN

El objetivo principal de este proyecto de detección DDoS es desarrollar un modelo que pueda detectar con precisión los ataques DDoS que se dirigen a un servidor y alertar a los administradores de red en tiempo real. Los ataques DDoS pueden ser muy perjudiciales para los servicios en línea, ya que pueden causar interrupciones en la conectividad, pérdida de datos y costos significativos de tiempo de inactividad.

El propósito principal de este proyecto es ayudar a proteger a las organizaciones y empresas de los ataques DDoS, permitiéndoles tomar medidas preventivas para minimizar el impacto de estos ataques. Al tener la capacidad de detectar un ataque DDoS antes de que cause daño significativo, los administradores de red pueden tomar medidas para mitigar el ataque y reducir el tiempo de inactividad en el servicio.

Este proyecto también tiene como objetivo mejorar la seguridad en línea en general, al ofrecer un modelo de detección de DDoS que puede ser utilizado por diferentes organizaciones que necesiten proteger sus servicios en línea contra este tipo de ataques.

PREGUNTAS CLAVES

- **¿Cómo identificar patrones de tráfico anormales que podrían indicar un ataque DDoS en un servidor?** Esta pregunta es importante porque para detectar un ataque DDoS, es necesario poder identificar patrones de tráfico que sean distintos a los patrones de tráfico normales. Si se pueden identificar estos patrones de tráfico anormales, se puede activar una alerta para que los administradores de red tomen medidas preventivas.
- **¿Qué tipo de medidas preventivas se pueden tomar para mitigar el impacto de un ataque DDoS?** Esta pregunta es importante porque una vez que se detecta un ataque DDoS, es necesario tomar medidas para minimizar su impacto en el servicio. Es importante tener un plan de acción que permita mitigar el ataque y reducir el tiempo de inactividad del servicio.
- **¿Cuál es la capacidad de carga del servidor y cuál es el umbral de tráfico normal para ese servidor?** Esta pregunta es importante porque cada servidor tiene una capacidad de carga máxima y un umbral de tráfico normal. Si se puede determinar cuál es ese umbral de tráfico normal, se puede detectar con más precisión cuándo se está presenciando un ataque DDoS.
- **¿Cuáles son las fuentes de tráfico malintencionado que están siendo utilizadas en un ataque DDoS?** Esta pregunta es importante porque conocer las fuentes de tráfico malintencionado permite a los administradores de red tomar medidas preventivas específicas para bloquear esas fuentes de tráfico.

REVISIÓN DE LA LITERATURA

DDoS por sus siglas en inglés Distributed Denegation of Services es una evolución de los ataques DoS (Denegation of Services) el cuál es un método malicioso donde los atacantes saturan un servidor desde un solo computador por medio de envío de una cantidad masiva de solicitudes de paquetes TCP o UDP a un mismo servidor, la diferencia de un ataque DDoS es que utilizan múltiples computadoras para generar este tráfico masivo malicioso contra un servidor. El DDoS es uno de los 10 ataques más comunes, donde los atacantes no obtienen un beneficio directo o acceso al sistema sino que simplemente afectan a la efectividad del recurso atacado lo que puede hacerlo vulnerable a otro tipo de ataques. (Baker, K., 2023)

En una arquitectura de red SDN (Software Defined Networking) los ataques DDoS en la cual se pueden clasificar en dos tipos (Reflection and Exploitation) cada una con subtipos (como application level, protocol o volumetric). En cuanto a los mecanismos de detección se pueden clasificar en 5 tipos incluyendo los basados en machine learning, en estadística, en umbrales o combinación de varios de estos, luego cada uno de estos también se puede clasificar en subtipos como el de machine learning pueden estar basados en redes neuronales, clasificación, agrupación, aprendizaje profundo y de conjuntos. Los métodos de aprendizaje automático y de umbrales (threshold) son los más utilizados. (Yunhe Cui, et al., 2021)

Actualmente ya existen productos que integran métodos de ML para la detección de ataques DDoS, una de las desventajas de estos métodos es que requieren del ingreso de datos clasificados manualmente y de un analista humano al momento de la detección para evitar bloqueos de sistema ante un falso positivo. M Devendra Prasad propone un método usando potenciación por gradiente estocástico sobre un dataset no balanceado para simular tráfico real demostró que no hay clasificaciones erróneas, teniendo así mejores métricas que los algoritmos de K-NN, Decision trees, Random Forest y Naive Bayes. (M Devendra Prasad, et al., 2019)

Los datasets que se usan para entrenar y testear modelos no son siempre los mismos y dependiendo los servidores, regiones y dispositivos que se atacan, estos pueden variar en sus características. Por eso también se investigan y se generan distintos tipos de datasets, algunos recolectados de tráfico real y otros a partir de simulaciones. Para simulaciones se usan algunas herramientas como NS2, LOIC, XOIC, HUKL, PyLoris y DAVOSET. Para los dataset recolectados de tráfico real se obtienen con ayuda de instituciones como el conjunto de datos de IoT de Latam que fue recolectado por una colaboración entre Aligo, la Universidad de Antioquia y el Tecnológico de Monterrey a partir de un entorno para ataques DoS utilizando componentes físicos. También se encuentran los datasets de la Universidad de New Brunswick que utiliza la generación de tráfico malicioso y benigno por medio de su herramienta CICFlowMeter-V3 generando ataques como PortMap, NetBIOS, LDAP,

MSSQL, UDP, UDP-Lag, entre otros. (Sabah Alzahrani, Loang Hong, 2018) (Almaraz, J., et al. 2022) (UNB, sin fecha)

RECOLECCIÓN DE DATOS

La data se obtendrá por medio de la página Kaggle, más específicamente en el Dataset DDoS, cabe aclarar que no hay datasets exclusivos recientes para DDoS públicos, entonces los datos con los que se trabajarán serán del año 2016, Para introducir más variación, los datos DDoS se extraen de diferentes conjuntos de datos IDS que se produjeron en diferentes años y diferentes herramientas experimentales de generación de tráfico DDoS. Los flujos DDOS extraídos se combinan con flujos "benignos" que se extraen por separado del mismo conjunto de datos base y se convierten en un solo conjunto de datos más grande.

Esta misma Dataset podría incluir información sobre el tráfico de red que se está monitoreando. Los datos recolectados podrían ser los siguientes:

- Direcciones IP de origen: Las direcciones IP de origen de los paquetes que llegan al sistema pueden ser capturadas y utilizadas para identificar patrones de tráfico sospechosos. Los ataques DDoS generalmente involucran un gran número de direcciones IP de origen.

- Tiempo de llegada: Es útil registrar la hora exacta en que se recibió cada paquete, ya que esto puede ayudar a identificar patrones de tráfico sospechosos.

- Tamaño del paquete: Es importante registrar el tamaño de los paquetes, ya que los ataques DDoS a menudo involucran un gran número de paquetes pequeños que pueden abrumar al sistema.

- Tipo de paquete: También es importante registrar el tipo de paquete, ya que algunos ataques DDoS pueden involucrar paquetes específicos, como paquetes ICMP o UDP.

- Puerto de origen: Registrar el puerto de origen del paquete también puede ser útil, ya que algunos ataques DDoS pueden involucrar un gran número de conexiones a un puerto específico.

- Ancho de banda: Registrar la cantidad de ancho de banda utilizado por el tráfico entrante puede ser útil para identificar ataques DDoS que intentan abrumar el ancho de banda disponible.

- Protocolo: Registrar el protocolo utilizado en los paquetes puede ser útil para identificar ataques DDoS específicos que utilizan un protocolo particular.

- Comportamiento de conexión: Observar el comportamiento de la conexión entre las direcciones IP de origen y los puertos de destino puede ayudar a identificar patrones sospechosos, como un gran número de conexiones desde una sola dirección IP.

- Duración del ataque: Registrar la duración del ataque puede ser útil para identificar ataques DDoS que se extienden durante un período prolongado de tiempo.

Referencias

Baker, Kurt. (2023). 10 MOST COMMON TYPES OF CYBER ATTACKS. Cybersecurity 101, Crowdstrike. Extraído de: https://www.crowdstrike.com/cybersecurity-101/cyberattacks/most-common-types-of-cyberattacks/

J. G. Almaraz-Rivera, J. A. Perez-Diaz, J. A. Cantoral-Ceballos, J. F. Botero and L. A. Trejo, "Toward the Protection of IoT Networks: Introducing the LATAM-DDoS-IoT Dataset," in IEEE Access, vol. 10, pp. 106909-106920, 2022, doi: 10.1109/ACCESS.2022.3211513. Extraído de: https://ieeexplore.ieee.org/document/9908531

M Devendra Prasad, Prasanta Babu V, C Amarnath. (2019). Machine Learning DDoS Detection Using Stochastic Gradient Boosting. IJCSE. Extraído de: https://www.ijcseonline.org/pdf_paper_view.php?paper_id=4011&28-IJCSE-06600.pdf

Sabah Alzahrani, Liang Hong. (2018). Generation of DDoS Attack Dataset for Effective IDS Development and Evaluation. Scientific Research Publishing. Extraído de: https://www.researchgate.net/publication/327036867_Generation_of_DDoS_Attack_Dataset_for_Effective_IDS_Development_and_Evaluation

University New Brunswick. (Sin fecha). DDoS Evaluation Dataset (CIC-DDoS2019). Extraído de: https://www.unb.ca/cic/datasets/ddos-2019.html

Yunhe Cui, Qing Qian, Chun Guo, Guowei Shen, Youliang Tian, Huanlai Xing, Lianshan Yan. (2021). Towards DDoS detection mechanisms in Software-Defined Networking, Journal of Network and Computer Applications. https://doi.org/10.1016/j.jnca.2021.103156. Extraído de: https://www.sciencedirect.com/science/article/pii/S1084804521001703