Detección de Ataques DoS con Herramientas de Minería de Datos

Klenzi, Raúl; López, Marcelo
Instituto de Informática / Departamento Informática / Facultad de Ciencias Exactas Físicas y Naturales / Universidad Nacional de San Juan
Domicilio: Av. Ignacio de la Roza 590 (O), Complejo Universitario "Islas Malvinas", Rivadavia, San Juan, CPA: J5402DCS, 0264-260353 0264-4260355
{rauloscarklenzi; marcelo.sanjuan.ar;}@gmail.com

Resumen

En el marco de proyectos “La ciencia de los datos en grandes colecciones de datos” y “Evaluación de arquitecturas distribuidas de commodity basadas en software libre” contenidos en el “Laboratorio de Sistemas Inteligentes para la búsqueda de Conocimiento en Datos Masivos”, integrado por docentes investigadores del Departamento e Instituto de Informática (DI-IdeI) de la Facultad de Ciencias Exactas Físicas y Naturales FCEFN, se trabaja enmodelary desde allí mitigar, ataques a un servidor de red por denegación de Servicios(Denial of Services)-DoS- mediante el análisis offline de un flujo de datos simulados y la utilización de algoritmos y herramientas correspondientes a Data StreamMining (Minería de datos -MD- en flujos de datos continuos). La aplicación utiliza módulos y algoritmos específicos delas herramientas de software libre RapidMiner (RM) 5.3.015 y KNIME 3.3
.

Palabras clave: Extracción de Conocimiento, Análisis off line, DNS, software libre

Contexto

En el ámbito del IdeI, por ordenanza 02/2015-CD-FCEFN, se conformó el “**Laboratorio de Sistemas Inteligentes para la extracción de Conocimiento en Datos Masivos”**, aquí se cubrendiferentes áreas que abarcan conocimiento del hardware, software y fundamentalmente datos que por su magnitud, necesitan de herramientas específicas.

Enlos proyectos insertos en el laboratorio, se desarrollan aplicaciones del área temática: Descubrimiento de Conocimiento en Bases de Datos (Knowledge Discovery Data -KDD-) que es un análisis automático exploratorio y modelado de grandes depósitos de datos e involucra inteligencia artificial, aprendizaje automático (Machine Learning -ML-), estadística, sistemas de gestión de base de datos, técnicas de visualización de datos ymedios que apoyan toma de decisiones.

Data StreamMining es el proceso de extraer conocimiento enestructuras de datos continuas y con rápidas transiciones. Un data stream es una secuencia ordenada de instancias que herramientas de extracción de conocimiento en datos pueden leer utilizando capacidades de cómputo y almacenamiento limitadas. Ejemplos de data streams incluye análisis de tráfico en redes de computadoras, comunicaciones telefónicas, transacciones ATM, búsquedas web y/o datos relevados desde sensores. El Data streammining se considera como un subcampo del ML, KDD y MD.

Desde la medición del flujo de bits que ingresan al servidor se propone una primera instancia de modelación de trazas de redque caracterizan a un ataque DoS o DDoS (Distributed DoS), el cual es un ataque a un sistema de computadoras que causa que un servicio o recurso sea inaccesible a usuarios legítimos. Aquí seanaliza un flujo de datos simulado, utilizando herramientas de software libre de ML Rapidminer 5.3.15 y KNIME 3.3.

Introducción

En el ámbito de redes de comunicaciones, un ataque o intrusión se define como un evento en la red que aprovecha cualquier tipo de vulnerabilidad de un sistema informático para causar daño sin consentimiento del usuario de dicha red, afectando la confidencialidad, integridad, disponibilidad o no repudio y pueden presentar los siguientes signos verificables: *interrupción* (el recurso se vuelve no disponible), *intercepción* ("alguien" no autorizado consigue acceso a un recurso) y *modificación* (además de la intercepción es capaz de manipular los datos) [1].

Según cómo afecte el tipo de ataque a la red, se clasifica en ataque pasivo o activo. En un *ataque pasivo*, el intruso monitoriza el tráfico en la redy hace uso de la información capturada; se centra en la intercepción de datos y análisis de tráfico, para obtener información de la propia red, siendomuy difícil de detectar por no alterar los datos interceptados [2].

En un *ataque activo* el intruso interfiere con el tráfico que fluye por la red, explotando sus vulnerabilidades o de una víctima particular. Según su objetivo, estos ataques se clasifican en cuatro categorías: *Suplantación de identidad, Reactuación, Modificación de mensajes y Degradación fraudulenta del servicio*. Este último, conocido como DoS provoca una denegación de los recursos informáticos y de comunicación de un elemento de la red[3].

Un ataque de DoS tiene como objetivo atacar una infraestructura de red, causando que susservicios sean inaccesibles a usuarios que acceden de una forma legítima, normalmente ocasiona pérdida total de conectividad a la red debido al excesivo consumo del ancho de banda de la víctima y se implementa a través de la saturación intencional de los puertos del host atacado con un flujo constante de información, sobrecargando los recursos de los servidores y la capacidad de responder a las peticiones realizadas por los usuarios originales.

Existen varios métodos mediante los cuales se pueden realizar DoS a los servicios[4]: *Spoofed* (paquetes con una dirección de origen falsificada), *Malformed* (paquetes con bits o flags encendidos en forma anormal), *Floods* (paquetes conformados de manera legítima en gran cantidad), *Null* (paquetes sin contenido), *Protocol* (paquetes con protocolos ilegítimos), *Fragmented* (paquetes fragmentados los cuales nunca se completarán) y *BruteForce* (paquetes que exceden el umbral definido de „flowrates") [5].

En ataques de tipo (flooding)se centra en consumir recursos disponibles para el servicio o recursos de todo tipo que existan en el camino como routers, firewalls, etc., mediante la inyección de grandes volúmenes de tráfico.

Las estrategias de inundación, se clasifican como de baja o alta tasa. El flooding de tasa baja explota vulnerabilidades de los protocolos de red y permite que el tráfico inyectado adopte patrones periódicos de volumen fluctuante en el tiempo, en tanto el flooding de tasa alta consiste en la emisión de grandes cantidades de tráfico de manera constante y uniforme[6].

Los métodos más populares de ataque siguen siendo SYN-DDoS, TCP-DDoS y HTTP-DDoS y UDP-DDoS, y pueden detectarse analizando la variación temporal de señales del protocolo de comunicación. En el SYN Flood, un usuario realiza un número especialmente alto de inicios de conexión que nunca son finalizados evidenciándose por permanecer activa la señal syn del protocolo un tiempo prolongado.

Sin herramientas que permitan un adecuado monitoreo del tráfico de red en busca de anomalías, puede presentarse una paulatina degradación de los servicios ofrecidos.

El éxito de un método para la detección y mitigación de ataques DoS, depende en parte de factores como el consumo de recursos, tiempo de respuesta, complejidad, además de garantizar que su implementación no ocasionará interrupciones en los servicios provistos.

Lasactuales herramientas para combatir este problema supone contar con grandes recursos de hardware y software, ya que la gran mayoría de estos sistemas afectan recursos valiosos como procesador, memoria física y ancho de banda, requiriéndose en casi todos ellos, horas, semanas o incluso meses de análisis previos antes de mitigar realmente un ataque. Algunas herramientas para combatir ataques DoS, sólo realizan investigación forense que generan estrategias de prevención en el futuro, es decir, no son reactivas [8].

El aporte del trabajo a la mitigación de ataques DoS, se orienta al estudio de la efectividad de su detección mediante el Análisis de Datos (AD) utilizando técnicas de MD.

Las plataformas de AD para Ciberseguridad se agrupan en tres categorías (Figura 1): (1) de propósito experimental, (2) de propósito específico y (3) de propósito general. La primera agrupa aquellas plataformas dirigidas al desarrollo y prueba de algoritmos que posteriormente se aplican al escenario real de la Ciberseguridad. El segundo grupo contiene plataformas implementadas en escenarios reales de Ciberseguridad aunque realizan un número específico y limitado de tareas. Por último, las plataformas de propósito general tienen en común con las de propósito específico que pueden insertarse en escenarios reales, sólo que estas últimas, se pueden adecuar a cualquier tarea de Ciberseguridad.

Fig. 1. Taxonomía de las plataformas de AD orientadas a la Ciberseguridad.

Ejemplos de plataformas experimentales de AD son Clementine, Weka, RapidMiner, KNIME y Orange. Existe una gran cantidad de trabajos vinculados con la aplicación del AD en Ciberseguridad que solo se centran en el mejoramiento de los algoritmos de MD[9].

Aquí se propone un análisis de tráfico de red off_line basado en estrategias derivadas del ML, de MD y particularizando en el uso de herramientas de software libre que permitendesde la generación de WF con módulos específicos, la aplicación comparativa de diferentes algoritmos y formas de visualización de extracción de conocimiento en datos.

La Figura 2 presenta la encuesta del sitio KDnuggets.com respecto a las herramientas más utilizadas enML, KDD y MD.

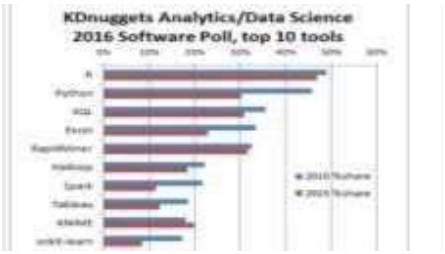


Fig. 2: Herramientas de software de ML más utilizadas en 2016 según kdnuggets.com

Desde la Fig. 2 y por ser herramientas de software libre, se seleccionan RapidMiner, y KNIME. Estas herramientas disponen módulos específicos que procesan flujos de datos y gran riqueza de formas de visualización.

- **RapidMiner**: Herramienta de software libre licencia AGPL RapidMiner 5.3.015, desarrollado en java en Dortmund Alemania.
- **KNIME** - KonstanzInformationMiner- en versión KNIME ANALYTICS 3.3 bajo licencia GNU 3, es una plataforma de ML construida en Eclipse y programada en Java, desarrollada en el departamento de MD y bioinformática de la Universidad de Constanza, Alemania.

Por la disponibilidad de grandes cantidades de datos surge el área de estudio del KDD, definida como “un proceso no trivial de identificación de patrones válidos, novedosos, potencialmente útiles y comprensibles a partir de los datos”[10].

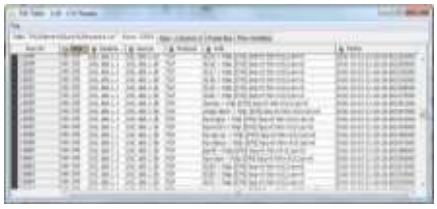
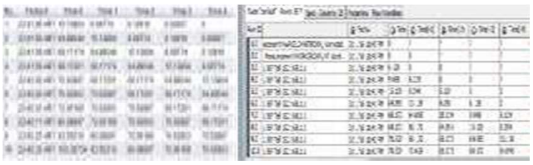


Figura 3: Tabla con fracción de datos a analizar

Desde el análisis del data streamming que ingresa al servidor, simulado en el laboratorio y presentado en la Fig. 3, se propone una primera instancia de modelación de tramas que caracterizan a un ataque DoS reconociendo en ciertos lapsos de tiempo el estado de un conjunto de señales como SYN, LEN, etc, que forman parte del atributo Info y permiten reconocer la existencia de un posible DoS.

Dado que el análisis de datos en toda estructura de tablas se realiza por filas, es necesario a los efectos de procesar información temporal (cada registro asociado a un evento temporal expreso), realizar una transformación mediante el uso del operador windowing (sliding window technique) [12] que permite establecer una ventana de análisis temporal ajustada por el usuario y así tratar de detectar patrones correspondientes a un ataque DoS, como se aprecia en la Figura 4.



RM KNIME

Figura 4: Sliding Windows, para el análisis de flujo de datos en las herramientas utilizadas

Para la tarea nos valemos de las potencialidades visuales de cada herramienta que permitirá de manera rápida y amigable reconocer, en ese análisis fuera de línea, las características asociadas a un ataque DoS y así tratar de mitigar posteriores intromisiones.

Líneas de Investigación,Desarrollo e Innovación

En el marco del laboratorio se llevan adelante diferentes trabajos de investigación aplicada, caracterizada por el tipo de datos observados y analizados. En este trabajo se ha realizado una aplicación de Data streammining siendo el objetivo extender la aplicación a otros tipos de datos anómalos que puedan reconocerse en diferentes trazas de red y generar la mejor forma de presentar conclusiones desde las potencialidades de visualización que poseen las herramientas.

Así mismo según la variedad y tipología de datos se está trabajando en análisis de series temporales, aplicaciones georeferenciadas, y reconocimiento de perfiles de usuarios, intentando llevarlas a plataformas paralelas en cluster de computadoras.

Resultados y Objetivos

Desde al análisis en diferentes ventanas temporales de las señales que conforman los protocolos de comunicación se han logrado detectar las diferentes variantes de ataques DoS. La Figura 5 muestra el reconcimiento de dos tipos de ataques DoS. El objetivo próximo es extender la aplicación al reconocimiento de otros Data Stream que caracterizan datos anómalos existentes en bases de datos del sitio www.Kaggle.com

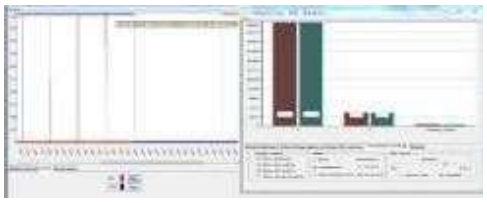


Figura 5: Histograma de Visualización de ataques SYN y UDP Flood

Al utilizar las herramientas de ML se ha observado que KNIME genera archivos de simulación de mayor tamaño que los que genera

RM, guardando las visualizaciones de salida, conjuntamente con la estructura del WF. Así, al momento de cargar un WF, se pueden observar las salidas de los diferentes módulos, en tanto en RM cada vez que se carga un WF se debe ejecutar y esperar las visualizaciones de salida hasta que la ejecución concluya.

Formación de Recursos Humanos

En el último año desde la actividad de ambos proyectos de investigación y del área de contención generada por el laboratorio se ha generado una gran actividad que tuvo su prólogo inicial en el trabajo publicado en el WICC 2016 realizado en Concordia con el trabajo "Confluencia de áreas de trabajo en un laboratorio de sistemas inteligentes". Allí se planteaba la posibilidad de llevar adelante dos trabajos finales de grado en Licenciatura en Ciencias de la Computación que destacaba la conjunción de dos áreas de trabajo bien diferenciadas y que hoy están en el proceso de redacción de informe final, como así también la posibilidad cierta de incorporar alumnos becarios y adscriptos. Así es como en el marco del proyecto "Ciencia de los Datos en grandes colecciones de datos" se cuenta con dos alumnos de grado adscriptos al proyecto, se están dirigiendo otros tres trabajos finales de grado y dos trabajos de maestría, en tanto integrantes del proyecto son doctorandos del Doctorado en Ciencias de la Informática que se dicta en la FCEFN_UNSJ. De manera similar se cumplen estas tareas en el otro proyecto de investigación mencionado en el trabajo.

Referencias

- P. Deepa Lakshmi, J. S. Praveen, V. Venkatraman, and N. Manoharan Director-Research, "A Review On Data Security In Distributed System," Int. J. Comput. Eng. Technol. N. A Rev. Data Secur.Distrib. Syst. Int. J. Comput. Eng. Technol., vol. 6, no. 610, pp. 13–16, 2015.
- [2] H. Sandberg, S. Amin, and K. H. Johansson, "Cyberphysical Security in Networked Control Systems: An Introduction to the Issue," IEEE Control Syst., vol. 35, no. 1, pp. 20–23, Feb. 2015.
- [3] N. Paulauskas and E. Garsva, "Computer System Attack Classification," Elektron.irElektrotechnika, vol. 66, no. 2, pp. 84–87, 2015.
- [4] G. Kumar, "Denial of service attacks – an updated perspective," Syst. Sci. Control Eng., vol. 4, no. 1, pp. 285–294, Jan. 2016.
- [5] P. Gasti, G. Tsudik, E. Uzun, and L. Zhang, "DoS and DDoS in Named Data Networking," in 2013 22nd International Conference on Computer Communication and Networks (ICCCN), 2013, pp. 1–7.
- [6] N. Sharma, M. Alam, and M. Singh, "Denial of Service: Techniques of Attacks and Mitigation," J. Comput. Sci. Eng. Softw.Test., vol. 1, no. 2, 2015.
- [7] H. Wang, Q. Jia, D. Fleck, W. Powell, F. Li, and A. Stavrou, "A moving target DDoS defense mechanism," Comput. Commun., vol. 46, pp. 10–21, 2014.
- [8] K. Singh, N. Kaur, and D. Nehra, "A Comparative Analysis of Various Deployment Based DDoS Defense Schemes," Springer Berlin Heidelberg, 2013, pp. 606–616.
- [9] A. F. Rivas and O. A. P. García, "Estado del Arte de las Plataformas de Análisis de Datos en la Ciberseguridad," 2016.
- [10] D. T. Larose and C. D. Larose, "Data Mining and Predictive Analytics-Wiley," 2015.
- [11] J.-P. Mens, Alternative DNS Servers. 2009.
- [12] H. Ryang and U. Yun, "High utility pattern mining over data streams with sliding window technique."