Detecting Abnormal DNS Events and Attacks in Real-Time

Martín Panza
martin@niclabs.cl

NIC Chile Research Labs
University of Chile

Abstract

Along with the huge growth of Internet during the last decade, DNS has become a critical part of Internet itself that has equally grown and became more complex. Partly because of the attacks and failures that it has undergone, which have been studied and resolved. Due to its importance, it is vital to be prepared for those attacks that we already know as they are performed by more powerful sources as technology advances. That is why detecting these events properly on time would help combat them. However, the volume of today's DNS data puts another barrier to this goal, making more difficult to make detections. Moreover, new attacks and events are always possible, making the detection systems based on supervised training useless in these cases. This work takes in account all of these issues to propose a real-time attack and abnormal events detector and its components; made and evaluated from big volume real DNS data, from chilean '.cl' top-level domain.

1 Introducción

Internet ha tenido un crecimiento considerable en la última década innegablemente, y junto a él el sistema DNS, que se ha vuelto una parte vital en el funcionamiento de Internet mismo. Junto a esto, DNS se ha vuelto cada vez más complejo debido al desarrollo tecnológico, surgimiento de necesidades, y debido a la ocurrencia de fallas y diseño de variados ataques contra las vulnerabilidades de la infraestructura. Resulta imposible asegurar que se tiene completa seguridad contra ataques o fallas de las que aún no se tiene conocimiento. Es más, en general la protección contra una amenaza se desarrolla luego de que se ha detectado una instancia de esta, se ha estudiado y se ha desarrollado una forma de combatirla; permitiendo por mientras un periodo de vulnerabilidad. Por otro lado, aunque aumente la capacidad del sistema, el mismo desarrollo tecnológico permite la ejecución de nuevos ataques, o mejoras de los ya conocidos, con nuevas magnitudes u objetivos.

Es por esto que el tema de detección automático de ataques en las redes ha tomado mucha relevancia en estudios recientes. En particular para DNS, donde se centra este estudio, presenta especial importancia para los operadores del sistema, quienes deben responder ante cualquier falla que generará sin dudas una gran repercusión en los usuarios del sistema. Sin embargo, cuando hablamos de DNS en el mundo real, nos referimos a considerables volúmenes de datos, dada la enorme y creciente cantidad de usuarios que realizan consultas constantemente. Esto dificulta la detección, a la vez que exige automaticidad y eficiencia en tiempo real.

En este trabajo se propone una forma de abordar dicho sistema capaz de detectar anomalías de forma no supervisada y clasificarlas, comparando diferentes métodos dentro de sus componentes tanto en detecciones, como en eficiencia.

2 Trabajo Relacionado

Existen importantes trabajos realizados acerca de la detección automática de anomalías y ataques en tráfico de red. Este tema está estrechamente relacionado con el tráfico de DNS pues presenta varias similitudes, y varios tipos de ataques se encuentran en ambos tipos de tráfico. A pesar de que sí existen diferencias, resulta

In: Proceedings of the IV School of Systems and Networks (SSN 2018), Valdivia, Chile, October 29-31, 2018. Published at http://ceur-ws.org

relevante observar las reglas y métodos que se evalúan en estos estudios.

La mayoría de los estudios realiza detecciones utilizando aprendizaje supervisado, donde se entrega una instancia específica de la amenaza al sistema y aprende a reconocerla en instancias posteriores [ASB18].

Existen también sistemas implementados que establecen ciertas reglas para encontrar diferentes tipos explícitos de eventos anómalos, como lo son SNORT [Mar94] y BRO [Pax99]. Por otro lado, se encuentra la detección de anomalías mediante el aprendizaje no supervisado, capaz de reconocer anomalías en amenazadas no conocidas, principalmente mediante la detección de outliers [CMO12].

Respecto a DNS, existen estudios donde se analiza y propone la forma de detectar ciertos ataques a DNS de manera particular, como DoS [Fei+03], DDoS [MR04] [DM03], Domain Fluxing [Yad+12], Botnet Domains y Malware [MM14], o Kaminsky Cache Poisoning [Mus+11]. Sin embargo, no existen estudios en profundidad respecto a aprendizaje no supervisado en DNS. Tampoco considerando grandes volúmenes de datos de consultas DNS, que corresponde en su mayoría al protocolo UDP y utiliza reglas y registros particulares; en comparación a otros tipos de tráficos de redes.

De esta manera, el trabajo presentado a continuación pretende centrar el problema de la detección automática de eventos anómalos en tráfico DNS de alto volumen, basándose en los sistemas diseñados para tráfico de red, y las soluciones para ataques particulares de DNS.

3 Datos

Los datos a utilizar, corresponden a un mes de funcionamiento normal de los servidores de NIC Chile. Comenzando desde el 2 de octubre del 2017 hasta el 2 de noviembre del mismo año.

NIC Chile es el actual administrador del registro y servicio de nombres de dominio '.cl', el dominio de nivel superior geográfico de Chile. Se cuenta con 17 servidores anycast, para los cuales se tienen todos los paquetes DNS de consultas y respuestas al servidor. Este set de datos es mayor a 1 TeraByte en volumen.

Dado que los datos son reales en un funcionamiento normal del sistema, toman suma importancia en el análisis de este trabajo y da relevancia a los resultados en caso de que se les dé un uso aplicado. Por otro lado, al ser datos de carácter cotidiano, no aseguran que se tengan registros de los ataques o eventos que se planean detectar. Esto requerirá una búsqueda de instancias de estos eventos en el set de datos. Además, el set de datos requerirá una incorporación de ejemplos de ataques para enriquecer el entrenamiento y validación del sistema. Para esto se hará uso de un simulador de consultas DNS con el que se cuenta: Dnszeppelin-clickhouse. De esta forma, es posible realizar instancias de ataques en conjunto con el tráfico normal de paquetes.

4 Trabajo a Desarrollar

El trabajo a desarrollar considera diferentes etapas independientes en implementación y evaluación, para las cuales se probarán diferentes métodos en busca del mejor desempeño y eficiencia:

4.1 Procesamiento de Datos

Dada la gran cantidad de datos que se obtienen para cada uno de los múltiples servidores, es vital la descripción correcta del sistema mediante agregaciones representativas de los paquetes DNS. De esta manera se pretende procesar la mayor cantidad de información sin perder partes importantes de esta. Para esto, se deberá agregar en base a diferentes llaves críticas para la detección de ataques, como lo son por ejemplo: IP de fuente (DoS, Spam, esparcimiento de Malware) y de destino (DoS, DDoS), consulta (Spam, PRSD), respuesta (Fallas en servidores, dominios no existentes), tipo de registro DNS, o ubicación geográfica. Además, agregación de los paquetes debe permitir la detección de nuevos ataques o eventos desconocidos. Esto resultará en múltiples series de tiempo que caracterizarán a grandes rasgos el sistema, buscando cumplir con el objetivo inicial de realizar este procesamiento en tiempo real.

4.2 Detección de Eventos Anómalos

Utilizando la descripción del sistema mediante múltiples series de tiempo obtenidas a partir de agregación de paquetes, en base a la búsqueda de outliers dentro de la serie temporal, se detectarán eventos anómalos en el sistema. Esto con el objetivo de encontrar también eventos de los que no necesariamente se tenga conocimiento, bajo la premisa de que estos eventos alterarán la descripción del funcionamiento normal del sistema. Esta detección considerará algoritmos cuya complejidad no irrumpa el procesamiento en tiempo real.

4.3 Caracterización del Evento

Al identificar el evento, será posible conseguir un timestamp que permita re-hacer una consulta a la base de datos, obteniendo información más detallada en un intervalo de tiempo que contenga al timestamp. Es importante mencionar que a partir de este paso puede realizarse una ejecución paralela que no interfiera con la detección en tiempo real llevada a cabo hasta ahora.

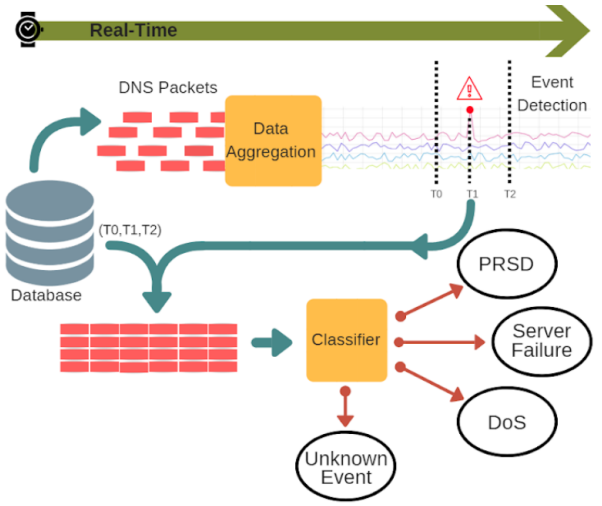
De esta forma se podrá volver a utilizar información no agregada del evento para facilitar su futura clasificación, incorporando nuevos atributos como características de los paquetes y comparación del contenido léxico de las consultas. Por otro lado, se podrá hacer una más fina agregación de IPs y paquetes, describiendo series temporales más precisas.

4.4 Clasificación de Eventos

Para validar los resultados obtenidos y evaluar el desempeño del sistema, es necesario reconocer los eventos de los que ya se tiene conocimiento; para verificar que efectivamente se esté detectando lo que se espera. Es por esto que, dadas las características obtenidas en el paso anterior, se utilizarán técnicas de clasificación para determinar el evento detectado. Sin embargo, para la detección de eventos desconocidos, será necesario mantener una clase extra para marcar ciertos eventos como "anomalía no identificada". Además, la resolución debe dar un grado de certeza para comprobar que se identifique una instancia correcta respecto de la que se tiene conocimiento, y no una tan solo similar. Es decir, obtener una cifra que indique la seguridad con la que se da el resultado de la clasificación.

5 Esquema del Sistema

La figura a continuación representa a los componentes del sistema mencionados previamente junto a su flujo de datos e interacción, de modo explicativo.



References

[ASB18] Amjad Alsirhani, Srinivas Sampalli, and Peter Bodorik. "DDoS Attack Detection System: Utilizing Classification Algorithms with Apache Spark". In: *9th IFIP International Conference on New Technologies, Mobility and Security (NTMS)* (Feb. 2018).

[CMO12] Pedro Casas, Johan Mazel, and Philippe Owezarski. "Unsupervised Network Intrusion Detection Systems: Detecting the Unknown without Knowledge". In: *Computer Communications* 35 (Apr. 2012).

[DM03] Christos Douligeris and Aikaterini Mitrokotsa. "DDoS attacks and defense mechanisms: classification and state-of-the-art". In: *Computer Networks: The International Journal of Computer and Telecommunications Networking* (Oct. 2003).

[Fei+03] Laura Feinstein et al. "Statistical Approaches to DDoS Attack Detection and Response". In: *Proceedings of the DARPA Information Survivability Conference and Exposition* (2003).

[Mar94] Roesch Martin. "SNORT - Lightweight Intrusion Detection For Networks". In: *Proceedings of LISA '99: 13th Systems Administration Conference* 2 (Nov. 1994).

[MM14] Thomas Matthew and Aziz Mohaisen. "Kindred Domains: Detecting and Clustering Botnet Domains Using DNS Traffic". In: *WWW '14 Companion Proceedings of the 23rd International Conference on World Wide Web* (Apr. 2014).

[MR04] Jelena Mirkovic and Peter Reiher. "A taxonomy of DDoS attack and DDoS Defense mechanisms". In: *ACM SIGCOMM Computer Communication Review* (Apr. 2004).

[Mus+11] Yasuo Musashi et al. "Detection of Kaminsky DNS Cache Poisoning Attack". In: *Fourth International Conference on Intelligent Networks and Intelligent Systems* (2011).

[Pax99] Vern Paxson. "Bro: a system for detecting network intruders in real-time". In: *Computer Networks* 31 (Dec. 1999).

[Yad+12] Sandeep Yadav et al. "Detecting Algorithmically Generated Domain-Flux Attacks With DNS Traffic Analysis". In: *IEEE/ACM TRANSACTIONS ON NETWORKING* 20 (Oct. 2012).