Aplicación de redes neuronales profundas para la detección automática de Nombres de Dominio generados de manera algorítmica.

Carlos A. Catania, Jorge Guerra, Martin Marchetta, Gabriel Caffaratti, Lucia Cortez, Alfredo Rezinovsky , Franco Palau y Juan Manuel Romero.

Universidad Nacional de Cuyo, Facultad de Ingeniería, LABSIN
Campus Universitario, Mendoza, Argentina.

{harpo, martin.marchetta, gabriel.caffaratti, jorge.guerra,franco.palau, alfredo.rezinovsky, lucia.cortez,juan.romero}@ingenieria.uncuyo.edu.ar

RESUMEN

En el contexto de la seguridad de redes de datos, un nombre de dominio generado de manera algorítmica (DGA, de sus siglas en inglés) es utilizado por el software malicioso (malware) para generar de manera dinámica un gran número de nombres de dominios de manera pseudo aleatoria, y luego utilizar un subconjunto de estos como parte del canal de Comando y Control (C&C). El presente proyecto se enfoca en el desarrollo de algoritmos de detección de DGA mediante la utilización de algoritmos de aprendizaje de máquinas en general y las redes neuronales profundas en particular. Durante el último periodo del proyecto, se ha puesto especial énfasis en la puesta a punto de los modelos obtenidos con vista a su despliegue en ambientes de producción. En particular lo referido a la evaluación de los distintos aspectos necesarios para la estimación del error de generalización, más allá de la división aleatoria entre conjuntos de entrenamiento y prueba.

Palabras Claves: Seguridad de Redes, Detección de Anomalías, Aprendizaje Automático

CONTEXTO

El presente proyecto se desarrolla en el marco de Facultad de Ingeniería dentro Laboratorio de sistemas inteligentes (LABSIN). Este trabajo es parte del proyecto de investigación que dio inicio en septiembre de 2019 en el marco de los proyectos bienales de secretaria de Investigación, Internacionales y Posgrados (SIIP) de la Universidad Nacional de Cuyo.

1 INTRODUCCIÓN

La detección temprana de secuencias DGA y su posterior bloqueo por parte de los administradores de sistemas resulta fundamental a fin de evitar o al menos mitigar la propagación de las acciones maliciosas dentro de la red. Es por ello que es de vital importancia desarrollar herramientas de detección, no solo con una baja tasa de falsos positivos sino también con la capacidad de operar en tiempo real. La utilización de técnicas de aprendizaje de máquinas surge como la alternativa más interesante para la construcción de una herramienta de detección de DGA. La metodología más común para la construcción de modelos de detección basados en técnicas de aprendizaje de máquinas consiste en utilizar un conjunto de datos conteniendo información sobre nombres de dominios normales y DGA. Este conjunto de datos es utilizado para entrenar un algoritmo que da como resultado un modelo de detección capaz de discriminar entre dominios normales y DGA. La principal ventaja de las técnicas de aprendizaje de máquinas es su capacidad de generalizar a casos similares nunca antes vistos. Lo que facilita considerablemente la tarea del administrador de redes, ya que este no tiene

que mantener actualizada la base de datos con nombres de dominios maliciosos. Sin embargo estas presentan algunos inconvenientes para su aplicación en escenarios reales. Entre los inconvenientes principales se destacan: (a) la tasa de falsos positivos, (b) el diseño del grupo correcto de atributos de entrada y (c) la necesidad de reentrenar periódicamente.

La capacidad de reconocer casos nunca antes vistos trae como consecuencia que algunos nombres de dominios normales, puedan ser detectados como maliciosos. Esto se conoce como falsos positivos. En algunos dominios de aplicación, el número de falsos positivos puede llegar a ser considerablemente alto sin perjuicio de la viabilidad en la aplicación de la técnica de aprendizaje de máquina elegida. Sin embargo, en el caso de la detección de DGA, un nombre de dominio bloqueado de manera incorrecta puede perjudicar seriamente la usabilidad del servicio de nombres de dominios (DNS). Basta que un usuario que no pueda acceder a un sitio con un nombre de dominio erróneamente clasificado como DGA para que genere inconvenientes en las tareas cotidianas de los administradores de sistemas. Es por esto último que mantener un número de falsos positivos muy bajo, resulta fundamental para realizar la detección en escenarios reales.

El número de falsos positivos está directamente relacionado no solo con el tipo de algoritmo de aprendizaje de máquina elegido, sino también con los atributos (variables predictoras) utilizados como entrada del algoritmo. Dentro del área de aprendizaje de máquinas, la construcción de los atributos de entrada adecuados al problema se denomina normalmente ingeniería de atributos. La ingeniería de atributos es una de las tareas que muchas veces demanda no solo los mayores recursos computacionales sino también humanos.

Por otro lado, los modelos de detección construidos a partir de técnicas de aprendizaje de máquinas requieren en muchos casos ajustes periódicos a fin de lidiar con casos significativamente diferentes respecto a los vistos durante la construcción del modelo de detección. Este proceso normalmente implica la incorporación al conjunto de entrenamiento de los nuevos casos observados y la posterior re-ejecución del algoritmo de aprendizaje de máquinas. Lamentablemente, el tiempo de entrenamiento de algunas de las técnicas de aprendizaje puede resultar excesivamente alto para los requerimientos de una aplicación en un escenario real. La realidad es que si se quiere entrenar un modelo de detección en un tiempo adecuado se debe considerar una alta demanda de recursos computacionales. Sobre todo cuando el volumen de los datos del conjunto de entrenamiento comienza a ser significativo. Es por ello que hay que considerar aquellas técnicas que sean capaces de minimizar el tiempo requerido para el ajuste de los modelos.

En los últimos 10 años han sido desarrolladas técnicas de aprendizaje de máquinas conocidas bajo el nombre de Aprendizaje Profundo (DL). La utilización de estas teécnicas ha sido la causa detrás de los mayores avances en el reconocimiento automático de imágenes, audio, video y análisis de texto. En particular las redes neuronales profundas ofrecen ventajas que permiten lidiar con los inconvenientes (a), (b) y (c) mencionados anteriormente. La principal ventaja radica en no tener que lidiar con el diseño del grupo correcto de atributos de entrada. A través de sus múltiples capas, las redes neuronales profundas son capaces de ir aprendiendo los atributos de entrada más adecuados para el problema. Además, esta selección automática de los atributos de entrada puede mejorar significativamente la eficiencia en términos de su tasa de detección de casos tanto positivos como negativos. Por último, recientes avances en la tecnología de procesamiento de los algoritmos de redes neuronales profundas han permitido disminuir considerablemente los tiempos de entrenamiento con grandes conjuntos de datos. En particular, la utilización de la capacidad de cómputo provista por las placas gráficas (GPU, del inglés Graphics Processing

Unit) ofrece una ventaja significativa frente al procesamiento en computadoras con CPU (del inglés Central Processing Unit).

El presente proyecto propone analizar la aplicación de redes neuronales profundas para el aprendizaje de los patrones comunes a los DGA de tal manera que permita desarrollar herramientas de detección no solo con una baja tasa de falsos positivos sino también con la capacidad de operar en tiempo real. Esto último resulta fundamental para lidiar con las amenazas de seguridad de hoy.

2 LÍNEAS DE INVESTIGACIÓN Y DESARROLLO.

El presente proyecto se enmarca en una línea de investigación que se viene desarrollando en el LABSIN desde 2017, la cual se centra en la aplicación de técnicas de aprendizaje automático a la seguridad informática.

Para el desarrollo del presente proyecto pueden diferenciarse 3 etapas principales:

A) Análisis preliminar del problema. Las tareas asociadas a esta etapa tienen por objetivo conocimiento de los problemas asociados a la detección de DGA como así también los modelos probabilísticos actualmente implementados. Una tarea fundamental consiste en realizar una breve revisión de la literatura sobre la aplicación de las técnicas de aprendizaje profundo a problemas de detección de DGA. Dicha tarea tiene por objetivo tratar de determinar cuáles han sido los beneficios e inconvenientes al aplicar este tipo de algoritmos.

B) Desarrollo de un algoritmo para la detección de DGA basado en técnicas de aprendizaje profundo. Esta etapa tiene por objetivo el desarrollo, evaluación y puesta a punto de un primer prototipo funcional para la detección de anomalías en el tráfico de red. Este primer prototipo es desarrollado utilizando alguna de las bibliotecas disponibles que permitan la implementación de modelos basados en aprendizaje profundo de manera simple y eficiente. En esta etapa se definen los diferentes aspectos de la red como ser: el tipo de red a utilizar, la topología de la red y la secuencia de entrada entre otras. Luego se evaluan los resultados utilizando un conjunto de datos conteniendo tráfico etiquetado (DGA y normal).

C) Experimentación. Finalmente en la última etapa se centra en la evaluación del algoritmo propuesto sobre distintos conjuntos de datos: En particular, se evalua el algoritmo propuesto con diferentes conjuntos de datos previamente etiquetados. Durante este proceso se consideran las métricas habituales en el área utilizando mecanismos para validar la generalidad del modelo obtenido como validación cruzada. Durante esta etapa se realizan también estudios comparativos con otros modelos de reconocimiento de anomalías de la literatura.

3 RESULTADOS OBTENIDOS

Durante los últimos 6 meses del proyecto se ha puesto el foco en la experimentación y evaluación de los modelos para detección de DGA (Etapa 3 del proyecto). En particular se han explorado distintas técnicas de evaluación con el fin de obtener una estimación de su capacidad de generalización. La estrategia común para evaluar la capacidad de generalización de un modelo consiste en separar el conjunto de datos disponible en un conjunto de entrenamiento y otro de prueba. Si bien esta separación aleatoria del conjunto de datos es una práctica habitual, podría no ser siempre el mejor enfoque para estimar la generalización del rendimiento en algunos escenarios. El hecho es que esta metodología habitual dentro del área de aprendizaje automático puede a veces sobreestimar el error de generalización cuando un conjunto de datos no es representativo o cuando los ejemplos raros y esquivos son un aspecto fundamental del problema de detección. Es por ello que durante los últimos 6 meses del

presente proyecto se puso especial énfasis en el estudio de diferentes estrategias de muestreo a fin de obtener una estimación del error de generalización de los modelos obtenidos.

Durante esta etapa del proyecto se realizaron las siguientes actividades:

a) Recopilación de información bibliográfica sobre el tema poniendo especial énfasis en la aplicación de diferentes estrategias para muestreo de datos sobre la cual evaluar los diferentes modelos de detección.
b) Construcción de diferentes conjuntos de datos de entrada aplicando las diferentes estrategias de muestreo (Ver Figura 1)
c) Evaluación de los resultados de los diferentes modelos implementados sobre los distintos conjuntos de datos.
d) Desarrollo de un informe técnico y presentación de los resultados preliminares en CICCSI 2020 [10].

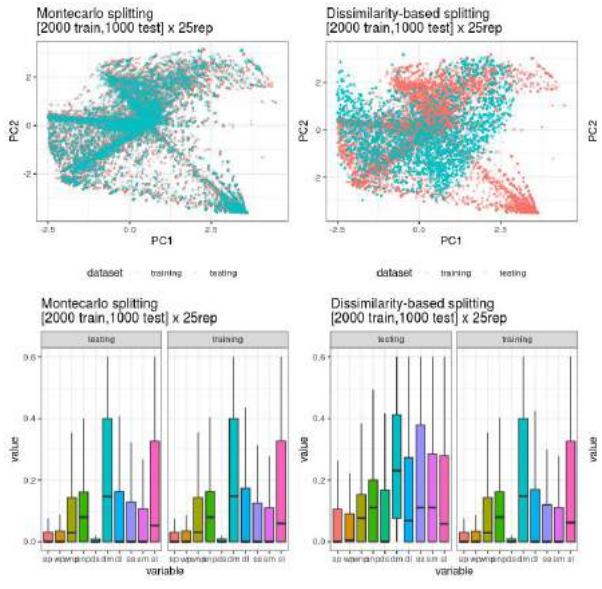

Figura 1: Dos de las técnicas de muestreo evaluadas. Montecarlo y Dissimilarity-based Splitting.

3 RESULTADOS ESPERADOS

Durante los últimos 6 meses del proyecto se espera:

1. Continuar fortaleciendo la línea de investigación en la aplicación de aprendizaje de máquinas aplicados a la seguridad informática.
2. Obtener una implementación funcional del modelo para la detección de DGA basado en redes neuronales con aprendizaje profundo. Sobre todo conocer los procedimientos necesarios para el despliegue de los modelos obtenidos en el contexto de una infraestructura de red de computadoras
3. Incrementar la experiencia para la posterior aplicación de modelos probabilísticos basados en aprendizaje profundo a nuevas líneas de investigación relacionadas con problemas de ciencia y tecnología.

4 FORMACIÓN DE RECURSOS HUMANOS

El proyecto ha permitido capacitar en el ámbito de la investigación a profesores y alumnos interesados en participar en un entorno académico y tecnológico innovador.

Sobre la temática de este proyecto se está trabajando en:

• La tesis doctoral de Jorge Guerra, en el doctorado en Ciencias Informáticas de la Universidad Nacional del Centro de la provincia de Buenos Aires.

Además el desarrollo del proyecto a permitido desarrollar:

• La capacitación del Sr. Franco Palau, alumno de la carrera de Ingeniería en Mecatrónica de la Facultad de Ingeniería.

• La capacitación del Sr. Juan Manuel Romero, alumno de la carrera de Licenciatura en Ciencias de la Computación de la Facultad de Ingeniería.

5 BIBLIOGRAFÍA

[1] Plohmann, D., Yakdan, K., Klatt, M., Bader, J., Gerhards-Padilla, E. (2016): A comprehensive measurement study of domain generating malware. In: 25th USENIX Security Symposium (USENIX Security 16). pp. 263–278. USENIX Association, Austin, TX.

[2] Kührer, M., Rossow, C., Holz, T. (2014): Paint it black: Evaluating the effectiveness of malware blacklists. In: Stavrou, A., Bos, H., Portokalidis, G. (eds.) Research in Attacks, Intrusions and Defenses. Springer International Publishing.

[3] Antonakakis, M., & Perdisci, R. (2012). From throw-away traffic to bots: detecting the rise of DGA-based malware. Proceedings of the 21st USENIX Security Symposium, 16. Retrieved from https://www.usenix.org/system/files/conference/usenixsecurity12/sec12-final127.pdf

[4] Jason Reed, Adam J Aviv, Daniel Wagner, Andreas Haeberlen, Benjamin C Pierce, and Jonathan M Smith (2010). Differential privacy for collaborative security. In Proceedings of the Third European Workshop on System Security, pages 1–7. ACM.

[5] Grill, M., Nikolaev, I., Valeros, V., & Rehak, M. (2015). Detecting DGA malware using NetFlow.

[6] M. J. Erquiaga, C. Catania and S. García, "Detecting DGA malware traffic through behavioral models," 2016 IEEE Biennial Congress of Argentina (ARGENCON), Buenos Aires, Argentina, 2016, pp. 1-6.

[7] Yadav, S., Reddy, A.K.K., Narasimha Reddy, A.L., Ranjan, S.: Detecting algorithmically generated domain-flux attacks with DNS traffic analysis. IEEE/ACM Transactions on Networking 20(5), 1663–1677 (2012)

[8] Schiavoni, S., Maggi, F., Cavallaro, L., & Zanero, S. (2014). Phoenix: DGA-based botnet tracking and intelligence. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 8550 LNCS, 192–211. https://doi.org/10.1007/978-3-319-08509-8_11.

[9] Ahluwalia, A., Traore, I., Ganame, K., Agarwal, N.: Detecting broad length algorithmically generated domains. In: Traore, I., Woungang, I., Awad, A. (eds.) Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments. pp. 19–34. Springer International Publishing, Cham (2017)

[10] Catania, Guerra, Romero, Caffaratti, Marchetta. Beyond Random Split for Assessing Statistical Model Performance. Anales 3er Congreso Internacional de Ciencias de la Computación y Sistemas de Información CICCSI 2020. (en prensa)