Uso de IA en el Análisis DNS: Detección de Dominios Generados Algorítmicamente (DGA)

Resumen de la investigación

Detectar dominios generados por algoritmos (DGA) desde la resolución DNS es clave para bloquear la comunicación C&C de las *botnets*, mitigar la propagación de *malware* y reducir campañas de *phishing*. Se presenta DeepD2V, un enfoque que utiliza solamente el nombre de dominio, preservando la privacidad y permitiendo un despliegue a gran escala. En DeepD2V cada dominio se segmenta en palabras/*tokens* probables y se generan embeddings utilizando Word2Vec (*skip-gram*) previamente entrenado. Sucesivamente, estos *embeddings* se combinan mediante distintas técnicas de agregación para formar una representación de 500 dimensiones. Sobre esta representación, se utiliza una arquitectura CNN unidimensional con múltiples capas para la clasificación binaria (DGA o benigno). Se evalúa el modelo en un conjunto balanceado, compuesto por dominios benignos extraídos de Alexa y 25 familias DGA. DeepD2V introduce una representación novedosa basada en *embeddings* y CNNs, alcanzando precisión y *recall* >97% y TPR ≈90% con FPR = 1%, generando mejoras notables en la detección de DGAs basados en diccionarios.

PhD (c) Lucas Torrealba

PhD Pedro Casas

PhD Javier Bustos-Jiménez

PhD Ivana Bachmann

PhD Mislav Findrik

PhD Germán Capdehourat

Antecedentes, problemática y relevancia del estudio

Los Dominios Generados por Algoritmos (DGA) son ampliamente utilizados por *botnets* y *malware* para establecer comunicación con servidores de Comando y Control (C&C). Una de sus principales aplicaciones es el *domain fluxing*, es decir, la generación masiva y dinámica de dominios que permite a los atacantes mantener la comunicación de la infraestructura maliciosa, evadiendo mecanismos de detección como las listas de bloqueo.

Detectar tempranamente estos dominios —por ejemplo, al momento de su registro—, es crucial para prevenir la propagación de *malware*, detener ataques DDoS y reducir campañas de *phishing*. Los DGA basados en diccionarios representan un desafío particular, ya que generan dominios con apariencia similar a los legítimos, dificultando su identificación con métodos tradicionales como listas de bloqueo, análisis de n-gramas, entre otras técnicas.

Metodología y diseño del estudio

El enfoque propuesto se basa en una representación innovadora de los nombres de dominio denominada Dom2Vec. En primer lugar, cada dominio se segmenta en palabras/*tokens* probables, los cuales se transforman en *embeddings* utilizando un modelo Word2Vec (arquitectura *skip-gram*) previamente entrenado. Los *embeddings* de cada palabra se combinan a través de diferentes técnicas de agregación —mínimo, máximo, promedio, suma y TF-IDF— para generar una representación vectorial de 500 dimensiones. El modelo DeepD2V aplica una red convolucional unidimensional de múltiples capas sobre esta representación de 500 dimensiones para clasificar dominios como benignos o DGA. El modelo fue evaluado en un conjunto balanceado de 675,000 dominios, compuesto en partes iguales por dominios benignos de Alexa y 25 familias DGA (21 aleatorias y 4 basadas en diccionarios).

Gráficos, datos y principales resultados

Los resultados muestran que diferentes enfoques de *Machine Learning* y *Deep Learning* —desde características básicas como la longitud o la entropía, hasta *embeddings* a nivel de caracteres— ofrecen buen rendimiento, especialmente en la detección de DGAs aleatorios. Sin embargo, su exactitud disminuye frente a DGAs basados en diccionarios, cuyos dominios se asemejan a los legítimos en estructura. En contraste, DeepD2V genera representaciones más expresivas mediante *embeddings* de palabras, lo que ha permitido separar de mejor forma los dominios benignos y maliciosos. Con este enfoque, el modelo logra una precisión y *recall* superiores al 97%, y una TPR cercana al 90% con FPR del 1% en el conjunto de datos evaluado.

Potencial uso y aplicación de los resultados

El enfoque propuesto es aplicable en distintos puntos de la infraestructura, como servidores DNS, proveedores de Internet, registradores de dominios, etc. Puede actuar en tiempo real como un sistema de recomendación o alerta temprana capaz de identificar dominios sospechosos antes de su eventual uso malicioso. De este modo, también facilita la detección temprana de intentos de comunicación con servidores de Comando y Control (C&C), sin necesidad de recurrir a fuentes externas de información, convirtiéndolo en muy buena alternativa para despliegues masivos en entornos de alto tráfico.

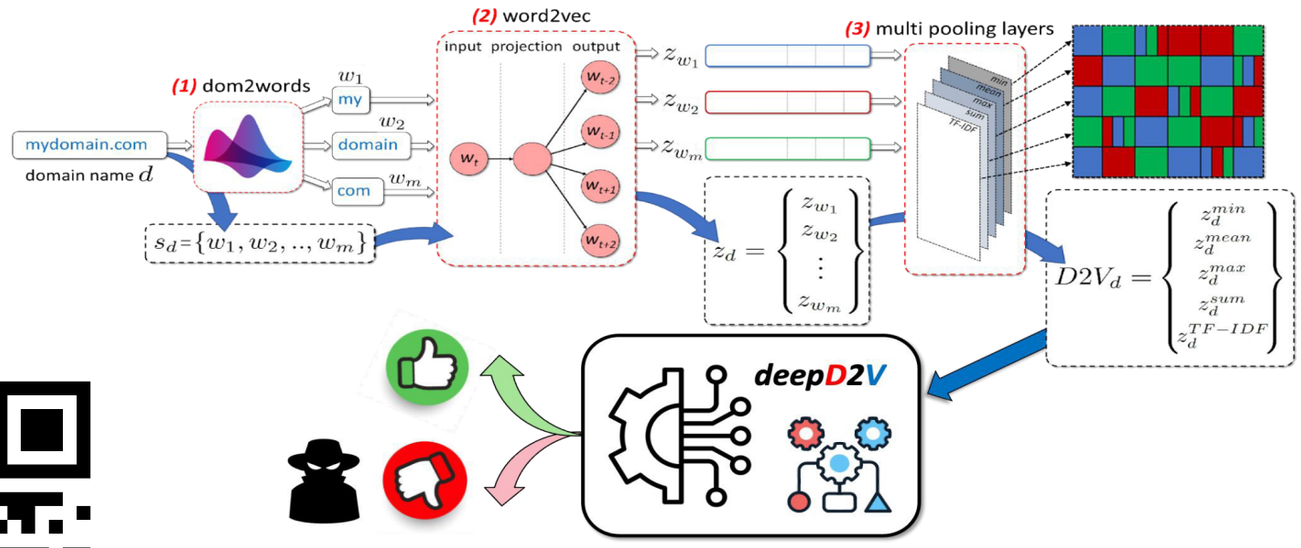
Debido a que utiliza solamente el nombre de dominio, DeepD2V preserva la privacidad de los usuarios al no requerir acceso al contenido de la página ni a información sensible asociada, como direcciones IP o servicios ofrecidos. Junto con esto, la utilización del modelo Word2Vec pre-entrenado no aumenta sustancialmente la complejidad computacional. Debido a esto, su arquitectura es altamente escalable, lo que facilita procesar grandes volúmenes de tráfico DNS sin afectar el rendimiento. Esta solución resulta especialmente útil para proveedores de Internet, entidades gubernamentales, registradores y organizaciones que buscan reforzar su protección frente a *phishing*, *botnets* y otras amenazas basadas en dominios maliciosos.

Potenciales líneas de desarrollo futuro y conclusiones

Una línea de trabajo futuro es extender la metodología hacia la detección de otros tipos de dominios maliciosos, más allá de los generados por algoritmos. Esto incluye dominios utilizados en ataques de *phishing*, *typosquatting* o la suplantación de marcas conocidas, que representan una amenaza creciente en Internet y requieren mecanismos de detección igualmente eficientes.

Una limitación del enfoque actual es que utiliza un diccionario mayoritariamente en inglés y caracteres latinos para la segmentación de dominios, lo que podría restringir su desempeño frente a alfabetos distintos, como el chino, árabe o cirílico, y también generar diferencias entre idiomas. Se propone como línea futura la incorporación de alfabetos y lenguas adicionales, con el fin de ampliar el alcance del modelo para fortalecer su aplicabilidad en distintos escenarios.

En conclusión, los resultados demuestran que DeepD2V constituye una solución práctica, escalable y robusta para la detección de DGAs. Su efectividad frente a casos complejos confirma el potencial de esta herramienta, aunque aún se requiere validación con tráfico DNS en tiempo real para consolidar su uso en entornos de producción.



Afiliación

NIC Chile Research Labs

Austrian Institute of Technology

Ceibal & Universidad de la República

Cyan Security Group

Contactos

lucas@niclabs.cl

investigacionaplicada@lacnic.net

Reconocimientos

Esta investigación es parte de la colaboración desarrollada por LACNIC con grupos de investigación e instituciones académicas regionales, en el marco del **Proyecto de Colaboración Efectiva para Investigación Aplicada (LACNIC).**

Citación de esta publicación

L. Torrealba, P. Casas, J. Bustos-Jiménez, I. Bachmann, M. Findrik, G. Capdehourat. “Uso de IA en el Análisis DNS: Detección de Dominios Generados Algorítmicamente (DGA)” [Presentación de póster]. Presentado en: LACNIC 44-LACNOG 2025, 6-10 de octubre de 2025, San Salvador, El Salvador.