SUPERCOMPUTACIÓN CASTILLA Y LEÓN

Detección de **muestras** de **malware** pertenecientes a APTs mediante **aprendizaje** basado en **Machine Learning**

Objetivo del proyecto

El objetivo de este estudio es el desarrollo de una herramienta que permita la detección de muestras de malware perteneciente a APTs mediante el análisis estático y dinámico de muestras anteriores utilizando técnicas de Machine Learning.

Periodo de ejecución

15 de marzo de **2018** al 17 de diciembre de **2019**.

Participantes del proyecto

INCIBE, *www.incibe.es*

SCAYLE Supercomputación Castilla y León, *www.scayle.es*

Financiación del proyecto

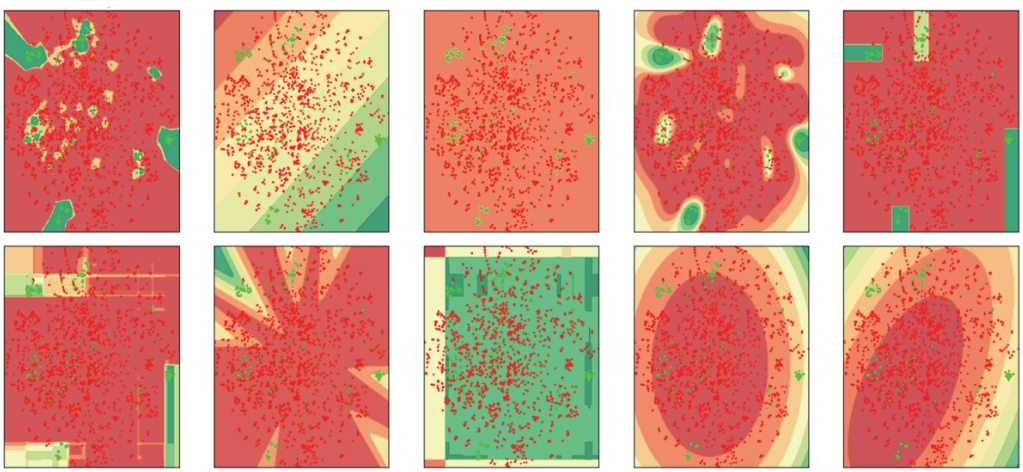
El proyecto se enmarca dentro de la adenda 18 del convenio de colaboración entre la universidad de León e INCIBE ("Desarrollo del equipo de investigación avanzada en ciberseguridad").

Funciones de SCAYLE

Scayle ha aportado el apoyo técnico y logístico para la realización de la investigación en sus instalaciones. Los cálculos en el supercomputador han permitido agilizar el proyecto, gracias a la capacidad de cómputo de Caléndula. SCAYLE también ha prestado soporte a los procesos de evaluación del rendimiento, así como proporcionado el espacio de almacenamiento para los datos de entrenamiento.

En una segunda fase el objetivo del proyecto es, entre otros, poder discernir, dentro del tráfico habitual de una organización si se vislumbra tráfico perteneciente a un ataque APT. Para ello es necesario disponer de una fuente de tráfico estable.

La RedCayle, Red Iris de Castilla y León, que gestiona Scayle es el medio perfecto para simular el tráfico que tendría una organización. El análisis del tráfico generado por el malware de una APT dentro del flujo de tráfico de RedCayle, es analizado en busca de patrones que permitan detectarlo en una organización real.



Detección de grupos de APTs utilizando diferentes técnicas de Machine Learning.

Justificación del proyecto

Una APT (del inglés Advanced Persistent Threat) es un tipo de ataque basado en malware, a menudo orquestado con un objetivo específico, dirigido a penetrar la seguridad informática de una entidad específica, ya sea privada o pública. Una APT generalmente fija sus objetivos en organizaciones o naciones por motivos de negocios o políticos. Este tipo de ataques involucra generalmente técnicas muy sofisticadas que utilizan malware para explotar vulnerabilidades, a menudo de día cero, en los sistemas internos de una organización.

A menudo, cuando las entidades son atacadas de forma exitosa mediante un ataque APT, no son conscientes de dicho ataque hasta pasado un tiempo muy elevado, que puede llegar a ser un plazo de años, tiempo durante el cual han perdido su ventaja corporativa, así como los datos de usuarios y clientes, lo que implica una pérdida de imagen y credibilidad de la que a menudo no pueden recuperarse.

Por otra parte, muy comúnmente las casas antimalware son incapaces de detectar si una muestra de malware pertenece o no a un ataque APT, puesto que se trata de muestras de malware normales que se camuflan entre los millones de muestras que estas casas antimalware reciben a diario.

La detección temprana de un ataque APT ayudaría a una entidad a evitar dichas pérdidas de dinero y credibilidad. Sin embargo, la detección temprana de ataques APT continúa siendo, a día de hoy, una tarea muy compleja.

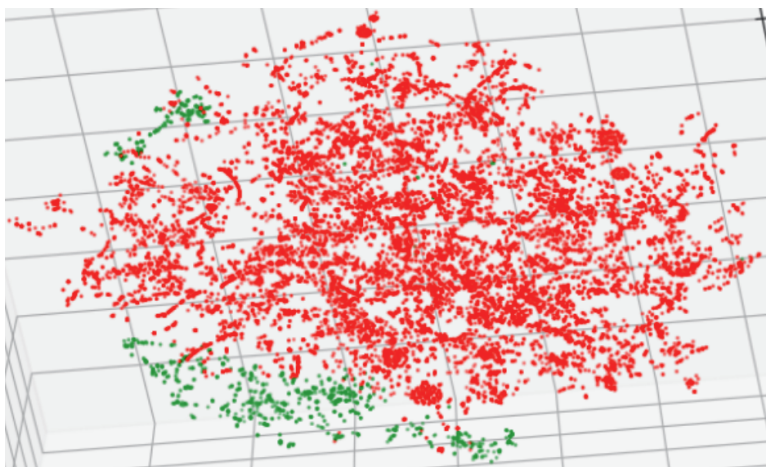
El presente proyecto pretende desarrollar un modelo inteligente, mediante técnicas de Machine Learning que permita detectar si una muestra de malware es susceptible de pertenecer a un ataque APT. Para ello se apoya por una parte en el análisis estático y dinámico de las muestras de malware y por otra en el análisis de tráfico de red generado por las APTs, de manera que la herramienta desarrollada sea capaz de analizar el tráfico de red circulante y permita determinar si una entidad está siendo infectada por un APT de una forma más proactiva.

Líder del proyecto

El proyecto está liderado por el profesor Adolfo Rodríguez de Soto y el grupo de trabajo formado en la actualidad por el investigador Luis Martín Liras.

Adolfo Rodríguez de Soto es Profesor Titular de Ciencias de la Computación e Inteligencia Artificial en el Departamento de Ingenierías Mecánica, Informática y Aeroespacial de la Universidad de León. Ha impartido docencia en la Universidad de León desde 1990 en diversas asignaturas, principalmente en el área de Estadística y de la Programación. Es autor y coautor de 3 libros docentes y ha participado en varios proyectos en este ámbito.

Su labor investigadora se enmarca en el área de la Inteligencia Artificial, siendo autor y coautor de más de 40 publicaciones tanto nacionales como internacionales y en revistas internacionales de alto impacto y ha participado en más de 35 congresos nacionales e internacionales.



Nube de puntos representando las muestras de malware de APT(en verde) y las muestras de malware lanzadas, donde se aprecia que es posible filtrar unas sobre otras