



BIG DATA APLICADO AL ANÁLISIS DE APLICACIONES ANDROID

Autor: Luis Gutiérrez Díaz

Director/es: Javier Vales Alonso

I. INTRODUCCIÓN Y CONTEXTO

El universo de la telefonía móvil ha cambiado radicalmente y a un ritmo vertiginoso. Los nuevos dispositivos conectados, con el *IoT* al frente, han abierto todo un nuevo abanico de posibilidades y han cambiado los hábitos de conexión de los usuarios. Desde el año 2018, el teléfono móvil inteligente se ha convertido en la herramienta principal de acceso a Internet, y ha permitido alcanzar las cifras de dispositivos y usuarios conectados, tasas de consumo de datos o descargas de aplicaciones móviles a niveles inimaginables hace unos años.

Por ello, nos encontramos ante un negocio que mueve miles de millones de euros. Cerca del 75% del mercado está dominado por dispositivos que utilizan el sistema operativo *Android*, un 23% de dispositivos *iOS* y un 2% para otros S.O. testimoniales.

Las aplicaciones móviles que corren en estos sistemas operativos son las interfaces de entrada y salida de los datos que utilizamos y generamos diariamente. Es aquí donde las debilidades, en forma de vulnerabilidades o defectos de desarrollo, pueden ser explotadas por terceros malintencionados. Son de dominio público casos donde los atacantes han hecho (o intentado hacer) negocio con el uso de los datos o incluso con la negación de acceso a los mismos.

Dado el volumen de aplicaciones del que hablamos hay que estudiar con cuidado cómo afrontar su análisis. Una línea respuesta viene dada por la propia evolución de las tecnologías de la computación. A grandes volúmenes de datos, y complejos problemas analíticos, grandes soluciones y nuevos métodos computacionales: el *Big Data* y el *Machine Learning*.

Este trabajo de fin de máster ha indagado en cómo operan las modernas herramientas de análisis de aplicaciones móviles, concretamente en el contexto de *Android*, que ocupa la posición predominante en el mercado. En concreto, el proceso está basado en un sistema de *Big Data* utilizado para recopilar, buscar interrelaciones y analizar las aplicaciones móviles de múltiples *markets*.

Los objetivos básicos del proyecto son:

- Mostrar el estado del arte de tecnologías similares que se utilizan para análisis de aplicaciones *Android*.
- Describir de forma básica el sistema *Big Data* que almacena y analiza aplicaciones *Android* y la tecnología sobre la que se asienta.
- Discutir con mayor detalle algunos casos de estudio.



II. DESARROLLO Y RESULTADOS

A lo largo del trabajo se explica someramente y con las restricciones naturales de confidencialidad que afectan al sistema la tecnología (TACYT, *The Tool for Cyber Intelligence*) que ElevenPaths, la unidad de ciberseguridad de Telefónica, utiliza para descargar, analizar, supervisar, buscar relaciones e indexar las aplicaciones para sistemas operativos *Android* e *iOS*.

TACYT, es en realidad una herramienta de *Big Data* que utiliza una base de datos distribuida MongoDB para almacenar sus datos. Es sobre esta base de datos donde se trabaja de forma masiva copiando las aplicaciones que los *crawlers* encuentran en los *markets* que tienen bajo supervisión. Tras la descarga de una aplicación, TACYT las desempaqueta y obtiene datos característicos de las mismas, como su archivo *AndroidManifest.xml* que contiene los permisos solicitados, las funciones y métodos que utiliza, las imágenes y otros datos adicionales. Este conjunto de características puede ser comparado con los más de 15.000.000 de versiones de aplicaciones que TACYT almacena, para identificar semejanzas y posibles amenazas. En particular, en TACYT se buscan relaciones espaciales y temporales:

- Espaciales, porque se pueden buscar características que están en las aplicaciones a lo largo de diferentes *markets* y plataformas.
- Temporales porque se tienen almacenadas todas las aplicaciones que se han eliminado, así que si en el pasado se publicó una aplicación y ésta se descargó a TACYT, se pueden realizar análisis forenses que amplíen la información de las investigaciones.

Además de TACYT, las aplicaciones son especialmente escrutadas por otra herramienta llamada mASAPP, que, analizando en profundidad su código, consigue proporcionar un indicador del nivel de vulnerabilidad que tienen las aplicaciones a sufrir un ataque, uso malintencionado o fugas de información privada. Es decir, mASAPP permite medir el nivel de exposición de una aplicación.

En mASAPP los resultados se presentan de dos formas distintas:

- Por un lado las vulnerabilidades, es decir, fallos en el código de la aplicación que tienen un CWE (*Common Weakness Enumeration*) reconocido, se evalúan y valoran de acuerdo al estándar *Common Vulnerability Scoring System* (CVSS).
- Por otro lado, los comportamientos, esto es los permisos y procedimientos que, en función de las circunstancias, estén activados o solicitados. Por ejemplo, sería absurdo que una aplicación que se utiliza para manejar remotamente por infrarrojos una televisión, solicite permisos de acceso a todos los contactos del teléfono y no solicite el control del sensor IR.

Además de la actividad documental y de formación teórica asociada al trabajo, se ha perseguido también un cariz práctico a través de la exposición de ejemplos reales: campañas de *adware*, uso inapropiado de permisos, falta de control de los desarrolladores, aplicaciones que mutan, o uso de apps para obtener inteligencia en conflictos armados, (por citar algunos). En dichos ejemplos se aportan explicaciones en conjunción de capturas de TACYT.



III. CONCLUSIONES

La tecnología móvil ha modificado los hábitos de consumo digital de los usuarios. Existe un inmenso y cambiante universo de aplicaciones *Android* e *iOS* en múltiples *markets*, y una gran falta de control y supervisión de las mismas. En esta situación, surgen organizaciones que se benefician de nuestros datos, con consentimiento explícito por parte del usuario mediante la autorización de permisos en las aplicaciones o sin él, tras saltar restricciones, al obtener permisos de forma ilícita, o incluso instalando aplicaciones no deseadas de modo oculto y subvertido.

Las organizaciones cibercriminales han aprovechado sus conocimientos en el campo de la informática tradicional para trasladarlos al ámbito móvil, ya sea con troyanos de cualquier variedad, *adware*, aplicaciones de *cliqueo* de anuncios o para realizar posicionamientos de aplicaciones ilícitas en los *markets*. Los controles de seguridad previos que aplican los *markets* oficiales han demostrado no ser infalibles y en ocasiones incluso muy débiles.

Como principal conclusión extraída de este estudio, los usuarios, apoyados por los operadores y proveedores de servicios y equipos, deben extremar el cuidado a la hora de elegir e instalar aplicaciones en los dispositivos. Algunas de las pautas que resultan de aplicación son:

- Descargar las aplicaciones de *markets* oficiales.
- Comprobar los permisos que solicitan las aplicaciones.
- No activar la actualización por defecto de las aplicaciones.
- Actualizar el sistema operativo conforme se liberen los parches.
- Utilizar herramientas para analizar aplicaciones sobre las que existan dudas.
- Utilizar un antivirus y/o una aplicación de seguridad en el teléfono.
- Activar el auto-bloqueo del terminal cuando no se utiliza y usar activación con algún método seguro: Patrón, PIN, huella dactilar, reconocimiento facial, etc.
- Utilizar claves fuertes y diferentes para cada aplicación.
- Conectar sólo a WIFIs seguras y caso de no ser posible usar aplicaciones VPN.
- No *rootear/jailbreakear* el teléfono.
- Encriptar los datos cuando sea posible y hacer *back up* en la nube.
- Activar el borrado remoto del terminal para casos de pérdida o robo.
- Evitar seguir enlaces sospechosos pese a haber sido recibidos de fuentes conocidas.

Y, sobretodo, utilizar el sentido común a la hora de manejar el terminal.