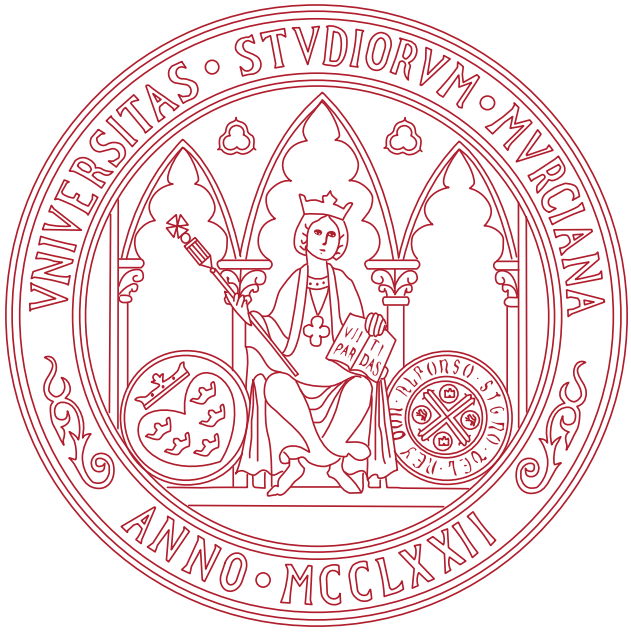

Universidad de Murcia

Trabajo Final de Grado

Identificación de dispositivos IoT a través de huellas hardware

Autor:
Sergio Marín Sánchez
sergio.marins@um.es

Tutores:
Gregorio Martínez Pérez
gregorio@um.es
Pedro Miguel Sánchez Sánchez
pedromiguel.sanchez@um.es

4 de julio de 2022

Me gustaría expresar mi agradecimiento a todos los amigos, profesores y familia que han hecho posible que llegue a este punto.

Índice general

Índice de figuras

Índice de tablas

CAPÍTULO 1

Introducción

1.1 Contexto

En los últimos años el número de dispositivos conectados a internet se ha incrementado en gran medida [1]. Esto se debe al uso de smartphones, tablets y demás dispositivos que requieren de conexión a internet para llevar a cabo la mayoría (o la totalidad) de tareas para las que han sido diseñados.

Cada dispositivo conectado a internet tiene asociados varios identificadores, como la dirección IP y la dirección MAC. Estos identificadores deberían servir para identificar unívocamente a un dispositivo, pero en la práctica no se da esta situación. Las direcciones IP pueden cambiar automáticamente debido al direccionamiento IP dinámico (mediante servidores DHCP), pero también pueden ser modificadas por las propias personas.

Estas modificaciones pueden ser por temas únicamente de privacidad, pero en muchas ocasiones están relacionadas con la ciberdelincuencia. Los delicuentes pueden intentar falsificar sus identificadores con el objetivo de que las personas, buscando conectarse a un servicio legítimo, acaben conectándose a sus equipos.

Los fines de esto son, por ejemplo, introducir virus en sus equipos, para realizar ataques DDoS (ataques de denegación de servicio distribuidos) mediante miles de equipos infectados, para cifrar los datos de dicho equipo (ataque de ransomware), para realizar ataques de phishing, enviando páginas visualmente idénticas a las que consulte el equipo, pero los datos sensibles de los formularios (contraseñas) pasen a disposición del atacante. Otro fin posible es el mero espionaje de los datos.

1.2 Motivación

Es en este punto en el que se requieren técnicas de comunicación seguras entre los dispositivos. Desde el punto de vista de las redes existen protocolos de comunicación segura como

SSL, TLS, IPsec, etc. Pero estos dispositivos son inútiles si el equipo se quiere conectar voluntariamente al equipo atacante (esto debido al engaño que se ha comentado anteriormente).

Por estos motivos, existe la pregunte sobre cómo podemos saber a qué equipos nos estamos conectando, o qué diferencia a un dispositivo de otro en internet si ambos presentan los mismos identificadores.

Una respuesta a estas preguntas es que los dispositivos aunque presenten el mismo hardware, tengan los mismos identificadores y ejecuten el mismo software, nunca serán exactamente iguales. Esto es debido a que en el proceso de fabricación de los dispositivos siempre habrá diferencias (por pequeñas que sean) que harán que los dispositivos sean distinguibles entre sí, por ejemplo, un dispositivo ejecuta una función en 1.2 ns y otro en 1.4 ns. Las diferencias son mínimas pero existen.

En el panorama actual del big data y el machine learning podemos explotar estas diferencias de tal forma que se generen huellas de cada dispositivo y con ello saber si realmente nos estamos conectando con el dispositivo adecuado o no.

En este marco de trabajo es en el que se centra este proyecto. Se busca crear un sistema que partiendo de un reloj exacto, compare las desviaciones de los relojes de los distintos dispositivos y con ello cree una huella estadística del comportamiento de cada uno. Posteriormente se automatizará el proceso de analizar esos valores estadísticos mediante un modelo de machine learning.

1.3 Objetivos

Para lograr nuestro objetivo final de identificar dispositivos idénticos de forma automática, podemos establecer diversas metas intermedias.

- **Objetivo 1**. Presentar la arquitectura IoT, así como sus diversas aplicaciones.
- **Objetivo 2**. Presentar distintas soluciones dentro del campo del Machine Learning que pueden ser aplicadas a nuestro problema.
- **Objetivo 3**. Analizar las distintas formas de obtener una marca de tiempo, con suficiente precisión, de un dispositivo.
- **Objetivo 4**. Generar un dataset con las distintas desviaciones de reloj de los dispositivos bajo análisis.
- **Objetivo 5**. Analizar estadísticamente las diferencias entre los distintos relojes de los dispositivos, con el fin de ver si son estadísticamente diferenciables.
- **Objetivo 6**. Generar un nuevo dataset con distintas variables estadísticas de las desviaciones previas.
- **Objetivo 7**. Dividir el nuevo dataset en conjuntos de entrenamiento y test para los modelos de Machine Learning, de forma que no se pierdan las características del mismo.

- **Objetivo 8**. Evaluar distintos algoritmos de Machine Learning para la tarea de distinguir entre los dispositivos.
- **Objetivo 9**. Describir las futuras vías de investigación de trabajos similares a este.

1.4 Estructura del documento

Este documento está compuesto en primer lugar por un resumen, tanto en español como en inglés (de forma extendida), seguidos de 6 capítulos.

- **Capítulo 1**. Es el capítulo actual, donde se presentan el contexto, la motivación y los objetivos del trabajo.
- **Capítulo 2**. En este capítulo se realiza una presentación de la arquitectura IoT y sus aplicaciones, así como, una presentación del Machine Learning y algunos de sus algoritmos.
- **Capítulo 3**.
- **Capítulo 4**. En este capítulo se hablará de nuestra propuesta para abordar este problema. Se obtendrán varios dataset y con ellos se entrenarán diversos modelos de Machine Learning.
- **Capítulo 5**. En este capítulo se analizarán los resultados obtenidos, en concreto, se evaluarán los distintos clasificadores usados.
- **Capítulo 6**. En este capítulo se exponen las conclusiones finales del trabajo y se comentan posibles vías futuras para esta línea de investigación.

Bibliografía

[1] R. van der Meulen, *8.4 billion connected things will be in use 2017 | gartner*, https://www.gartner.com/en/newsroom/press-releases/2017-02-07-gartner-says-8-billion-connected-things-will-be-in-use-in-2017-up-31-percent-from-2016, 2017.