

11. Recolección de información

Friday, October 9, 2020 4:23 PM

La recolección de información es uno de los pasos esenciales de prácticamente cualquier proyecto de un profesional de la seguridad informática.

Al ser una parte tan importante, es necesario mejorar los métodos que usamos para realizarla y, en la medida que la tecnología lo permita, automatizarla.

La recolección de información incluye la búsqueda pasiva de información de la red, tal como búsqueda de IP, escaneo de puertos, análisis de información de servidores, entre otras.

Dentro de la recolección de información también podemos incluir la extracción de información a través del análisis de metadatos de documentos y archivos multimedia.

Además, es posible obtener información valiosa a través del análisis del contenido de una página web, sobre todo cuando esta no ofrece ninguna API para acceder sus datos. En este tenor, existen varias técnicas:

- **Web scraping:** Es la actividad de tratar de obtener información de un recurso como una página web y procesar esa información para extraer datos relevantes.
- **Report mining:** Extraer información directamente de un archivo (HTML, PDF, CSV). La diferencia es que aquí no nos tenemos que conectar, ya que ya contamos con el archivo en nuestra posesión.
- **Spiders:** Siguen unas reglas para moverse por un sitio web y recabar información imitando la interacción que realizaría un usuario con el sitio web.

Para estos procesos exploraremos diversas librerías que Python que pueden ayudarnos en este proceso.