Análisis de documentos maliciosos – Parte 01 – Introducción y Máquinas Virtuales (VM)

Español

Introducción

Este (mini) curso está pensado para aquellos entusiastas y practicantes de la seguridad digital (soporte técnico, facilitadores, primeros en responder, etc.) que quieran aprender más sobre los documentos maliciosos y cómo identificarlos. Estos documentos pueden ser archivos adjuntos a correos electrónicos, archivos en memorias USB o descargas de páginas web específicas. Los objetivos principales son:

- Aprender los fundamentos de cómo funcionan los formatos de documentos comunes y cómo pueden convertirse en armas, con especial énfasis en los archivos de formato de documento portátil (PDF) y los documentos de Microsoft Office (al menos de MS Word, Excel y PowerPoint).
- Presentar algunas herramientas que pueden ayudar a identificar señales de documentos peligrosos o confirmar que son seguros de abrir.
- Proporcionar algunos consejos de seguridad y aclarar dudas comunes sobre el manejo de archivos sospechosos.

Este curso utiliza el formato de lecturas cortas y cuestionarios en la mayor parte del contenido abordado, en el que dependiendo del material, será necesario ejecutar algunas herramientas. Esto se abordará en la sección Entorno de trabajo, un poco después de la introducción. Los requisitos generales son:

Para completar los ejercicios propuestos:

- Capacidad de ejecutar 1) una máquina virtual en la computadora utilizando Virtualbox o un software similar, o 2) scripts de python (sólo para analizar los archivos del curso, no para analizar muestras reales).
- El tiempo destinado a cubrir el material (aproximadamente 2 horas).

Este curso toma materiales disponibles en otras referencias y sólo utiliza herramientas de libre acceso. La mayor parte del contenido está inspirado en el trabajo que Didier Stevens ha desarrollado a lo largo del tiempo, especialmente para SANS, así como en otras referencias, una lista corta podría ser:

- https://blog.didierstevens.com/2011/05/25/malicious-pdf-analysis-workshop-screencasts/
- https://github.com/filipi86/MalwareAnalysis-in-PDF
- https://www.sentinelone.com/blog/malicious-pdfs-revealing-techniques-behind-attacks/
- https://www.youtube.com/watch?v=opdVFQEBCNU

Estructura

1. Descargos de responsabilidad
2. Algunas consideraciones sobre el modelado de amenazas
3. Para cada tipo de formato de archivo (PDF, MS Office)
    1. Cómo están estructurados (en un sentido más técnico)
    2. Cómo pueden convertirse en armas
    3. Cómo podemos hacer un análisis introductorio
    4. Algunas conclusiones/hechos sobre el formato de archivo
4. Algunos consejos generales contra las amenazas relacionadas
5. Y ahora, ¿qué sigue?

A continuación, una serie de descargos de responsabilidad útiles antes de empezar con el material

Descargos de responsabilidad

Ante la naturaleza de la tarea que realizaremos después de dominar el contenido proporcionado (analizar archivos maliciosos y peligrosos), y la complejidad del tema (que vemos como una introducción al análisis de malware), recomendamos enfáticamente leer esta sección y estar de acuerdo con todos los puntos antes de seguir adelante.

1. **Este curso es de carácter introductorio:** está diseñado para personas sin ninguna experiencia previa en el análisis de documentos sospechosos. En la sección Siguientes pasos, incluimos una lista de recursos para lecturas y referencias adicionales.
2. **Este curso no trata muchas técnicas avanzadas:** hay muchas amenazas específicas con una complejidad que va más allá del alcance de este material, además, como en todo lo relacionado con la seguridad de la información, puede haber amenazas esperando a ser descubiertas que no serán tratadas en este curso. Recomendamos buscar ayuda en caso de que sospechemos que estamos viendo una amenaza avanzada o desconocida en un archivo o cualquier otro artefacto, más sobre esto en una sección posterior. En cualquier caso, este curso nos ayudará a comprender mejor cómo es habitualmente un archivo benigno, en lugar de cómo están estructurados todos los documentos maliciosos.
3. **Tome sus precauciones al analizar archivos reales:** las muestras utilizadas en este curso son inofensivas, es decir, si repite los flujos de trabajo presentados en muestras reales sin la respectiva seguridad medida, es muy probable que su dispositivo resulte infectado. Por favor, no ejecute ningún archivo sospechoso en su computadora principal, utilice una máquina virtual, un dispositivo dedicado o un entorno en el que no pueda ejecutar el archivo en su máquina sino sólo analizar sus propiedades.

Cuestionario de descargo de responsabilidad

Pregunta 1: Entiendo los riesgos de analizar archivos sospechosos, así como las posibles consecuencias de ejecutar malware a propósito o accidentalmente, he leído el contenido de esta página/sección y comprendo las estrategias más comunes para hacer frente a estas posibles amenazas.

Pregunta 2: ¿Cuál de las siguientes opciones describe mejor lo que necesitamos hacer cuando analizamos un archivo sospechoso real?

- 1. Deberíamos iniciar una máquina virtual (VM) o computadora dedicada para analizar el archivo y darle el menor acceso posible a nuestra máquina host y al resto de la red

- 2. Podemos analizar el archivo en nuestra propia computadora/entorno pero sin tener acceso a Internet

- 3. Debemos analizar el archivo en una computadora o máquina virtual (VM) con Sistemas Operativos menos comunes como Linux o macOS

Acerca de los modelos de amenazas

Cuando buscamos consejos sobre cómo actuar ante archivos sospechosos, por lo general el enfoque propuesto es evitar cualquier interacción con los archivos, por ejemplo:

- No abra archivos desconocidos.
- No interactúe con archivos sospechosos.
- No haga contacto visual con ningún archivo sospechoso.

O bien, podemos encontrar otro tipo de consejos que, aunque son suficientes para la mayoría de la gente, podrían ser engañosos para usuarios sensibles como activistas de Derechos Humanos o periodistas que trabajan en entornos peligrosos, o sencillamente contraproducentes, por ejemplo:

- Utilizar un antivirus es suficiente para protegerse de los archivos maliciosos.
- Sólo los documentos de Microsoft Office con macros son peligrosos, por lo que puede tratar otro tipo de archivos sin necesidad de preocuparse demasiado.
- Eliminar cualquier correo electrónico con archivos adjuntos sospechosos. Este punto es bastante inquietante en algunos escenarios, ya que si eliminamos los correos electrónicos y los archivos adjuntos de nuestra bandeja de entrada, perderemos evidencia clave que puede ayudarnos a evaluar si los artefactos son realmente maliciosos o están dirigidos a un objetivo, lo que podría ser información invaluable.

En la práctica, al trabajar con comunidades objetivo (especialmente periodistas), dejar de interactuar con los archivos no es una opción. Muchas organizaciones, grupos e individuos necesitan abrir archivos potencialmente peligrosos como parte de su trabajo, y lo harán incluso siendo conscientes de los riesgos, algunos ejemplos:

- Los periodistas reciben una invitación para una conferencia de prensa.
- Los activistas reciben un documento de respaldo como prueba en un caso de violación de derechos humanos o como filtración.
- Una institución adversaria envía un documento que debe ser revisado y abordado.

Un factor extra a considerar es que los actores de la sociedad civil están expuestos a amenazas selectivas desconocidas por los motores antivirus. Otro es que, según el tipo de ataque, otros formatos de archivo también pueden convertirse en armas. Estos factores deben ser considerados por las personas que ayudan a los grupos vulnerables para comprender mejor cómo los documentos y otros formatos de archivo comunes pueden ser utilizados como armas, para dar consejos útiles, pero también para ayudarles a analizar archivos específicos para saber y evaluar si están siendo víctimas de ataques específicos.

Con todo esto en cuenta, nos vamos a centrar en comprender cómo están estructurados los formatos de archivo estándar, cómo identificar los ataques más comunes que los utilizan y algunas medidas defensivas actualizadas para evitar ser víctimas de este tipo de amenazas.

Cuestionario sobre el modelo de amenaza

Pregunta 1: Para una organización fuertemente atacada que recibe muchos documentos de Microsoft Office por correo electrónico, ¿cuál de las siguientes opciones es cierta? (Sólo una es correcta)

- 1. Aunque el software antivirus (AV) diga que el archivo es seguro, podría contener malware.

- 2. Deben eliminar inmediatamente cualquier archivo adjunto sospechoso porque podría ser peligroso tenerlo en la bandeja de entrada.

- 3. No deben abrir ningún archivo adjunto que provenga de fuentes desconocidas.

Entorno: consideraciones generales

Para ejecutar la mayoría de las tareas de este curso utilizaremos herramientas básicas escritas en el lenguaje de programación Python, debido a la amplia compatibilidad de Python con todos los sistemas operativos. Hay incontables formas en las que podemos configurar un entorno, nosotros proponemos una específicamente, pero si está familiarizado con Python, el análisis de malware y/o la virtualización puede configurar una versión diferente que funcione para usted. La única recomendación enfática sería tener un entorno aislado para manipular artefactos peligrosos (en este caso, archivos), hay otras consideraciones, pero probablemente esta sea la más importante.

Entorno aislado y otras buenas prácticas

Las muestras utilizadas en este curso son inofensivas, sólo sirven para demostrar cómo están estructurados los archivos y cómo identificar las banderas rojas, sin embargo, si su intención es analizar archivos reales, lo más probable es que encuentre uno infectado que podría causarle todo tipo de problemas, como infectar la computadora que está utilizando, comprometer su información o inutilizar su dispositivo, entre otros. Dicho esto, es una práctica común contar con un entorno exclusivo para analizar y ejecutar muestras sospechosas de forma controlada, de modo que si algo sale mal mientras está manipulando la muestra, esto no afectará a su dispositivo ni a la información que contenga.

Otra ventaja de contar con un entorno dedicado es que después de manipular las muestras de malware, se puede borrar todo y empezar nuevamente sin temor a perder archivos no relacionados. Esto nos permite planificar formas prácticas de "restablecer" nuestro entorno a un estado listo para usar antes de cada análisis.

Una de las estrategias más utilizadas para garantizar un entorno aislado es el uso de máquinas virtuales (VM), las cuales básicamente emulan una computadora completa dentro de otra computadora, incluyendo el sistema operativo (SO), los discos duros, la pantalla, etc. Las herramientas más comunes para configurar y utilizar las VM son Virtualbox y VMware Workstation Player, entre otras. Utilizar hardware dedicado también es una opción, siempre y cuando esté protegido en caso de infección.

Un posible inconveniente podría ser que algunos malware incluyen código para verificar si son ejecutados en entornos aislados y no se ejecutan, lo que dificulta su análisis. Sin embargo, el peligro inherente de ejecutar malware en nuestros entornos cotidianos no vale la pena ni siquiera intentarlo, por lo que recomendamos buscar ayuda, centrarse en técnicas que no dependan de ejecutar los archivos sospechosos o conseguir información sobre cómo configurar un entorno que se parezca a una máquina real para una muestra de malware. Para este recurso, esto no debería ser un problema, ya que no vamos a ejecutar ningún código de los documentos, pero si usted quiere aprender y efectuar análisis dinámicos de archivos sospechosos, esto le será útil.

Otras consideraciones

Además de la buena práctica de tener un entorno aislado, otras prácticas comunes son:

- **Asegúrese de que la computadora que se está utilizando no está conectada a Internet o a la red local:** especialmente si se está abriendo archivos sospechosos. La razón más común para hacerlo es evitar la generación de señales que alerten a los operadores de malware de que el código está siendo ejecutado o probado según otros datos como la dirección IP, o el tipo de dispositivo que ejecuta el malware. Además, algunos malware intentarán propagarse a la red local, tratando de infectar otros dispositivos no previstos, por lo que es una práctica común aislar los dispositivos de prueba en diferentes redes físicas o virtuales (o VLAN). Hay que recordar que, en caso de analizar una muestra ejecutándola, es posible que el malware detecte que no tiene acceso a Internet y no se ejecute.

- **Si va a conectarse a Internet, utilice una VPN o similar:** el objetivo es ocultar su ubicación real en caso de que el malware que estamos analizando se ejecute y envíe señales a sus operadores. Una vez más, no se suele recomendar ejecutar malware sin medidas para evitar cualquier posible comunicación con los operadores, sin embargo, utilizar una VPN podría ser una buena medida en caso de ejecución accidental o si otras configuraciones fallan en un determinado momento.
- **Planifique un proceso para restablecer su entorno a un estado "limpio"**: Dependiendo de si se está utilizando una máquina virtual o hardware dedicado, hay algunas herramientas y funciones que son útiles para restablecer el entorno de modo que siempre que se analice una muestra, la máquina estará limpia. Para las máquinas virtuales el uso de instantáneas son un buen ejemplo, y hay software para revertir una computadora física a un estado anterior.
- **Limítese al análisis estático:** En general, podemos dividir el análisis de malware en función de si estamos ejecutando las muestras o no. El análisis estático trata de analizar minuciosamente los archivos y otros artefactos para recopilar toda la información posible sin ejecutarlos, al tiempo que el análisis dinámico ejecuta las muestras para ver qué cambia en el entorno de prueba. En función del tipo de malware, es posible que un tipo de análisis sea más útil que el otro, pero en general, el análisis dinámico requerirá más medidas para proteger el entorno de pruebas y la red para poder soportar la ejecución de malware real. Este curso sólo muestra técnicas de análisis estático.
- **Tenga cuidado cuando publique muestras u otra información de las muestras analizadas:** En general, esto podría alertar a los operadores de malware de que estamos analizando la campaña de malware, lo que podría hacerles cerrar la infraestructura, eliminar cualquier rastro para dificultar la atribución, entre otras acciones. Esto se aplica a cualquier plataforma pública, como redes sociales y sitios web, incluidas algunas plataformas públicas donde podemos enviar archivos para analizarlos en la nube buscando alertas de motores antivirus y de la comunidad de seguridad informática. Para el último escenario, compartiremos algunos ejemplos y técnicas para verificar la información que necesitamos sin alertar a nadie.

Prueba de entorno

Pregunta 1: ¿Cuál de las siguientes afirmaciones es verdadera?

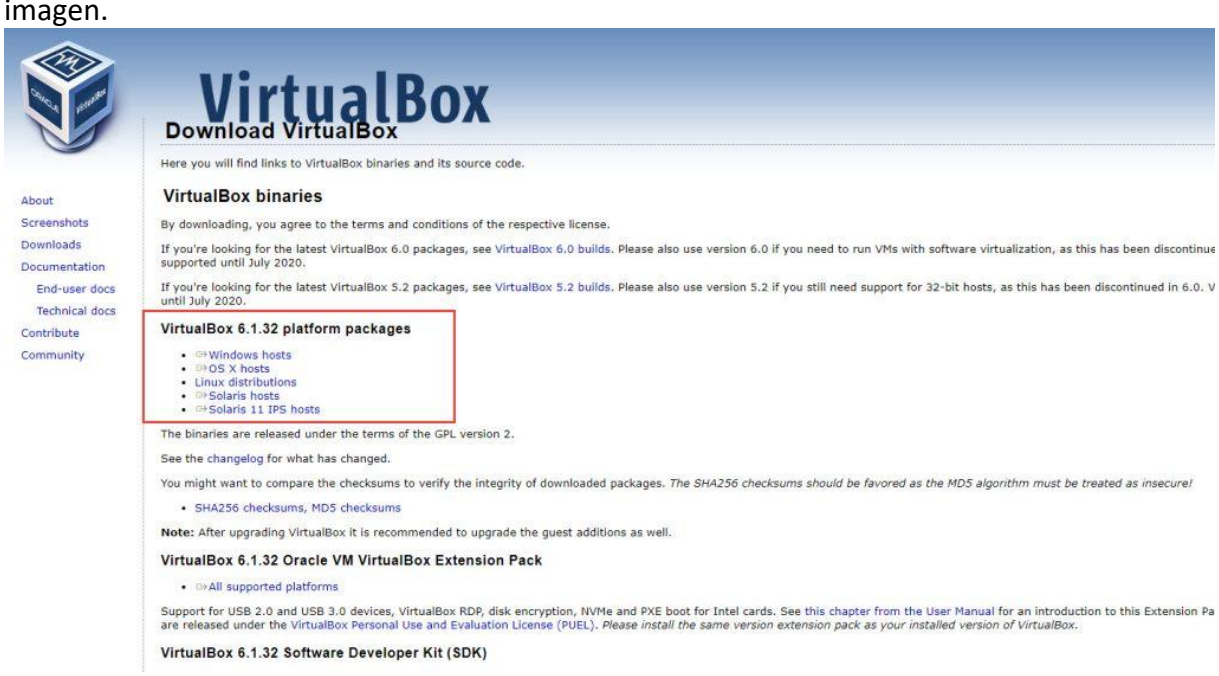
- 1. La ejecución de muestras requerirá menos medidas de seguridad que intentar analizar detenidamente los artefactos en busca de información útil.
- 2. Cortar el acceso a Internet hará más difícil que una muestra de malware notifique a los creadores que ha sido ejecutada.
- 3. La forma más eficaz de analizar el malware es utilizando máquinas virtuales porque si la máquina se infecta podemos volver a crearlas desde cero.

Ejemplo de entorno: Remnux + Virtualbox

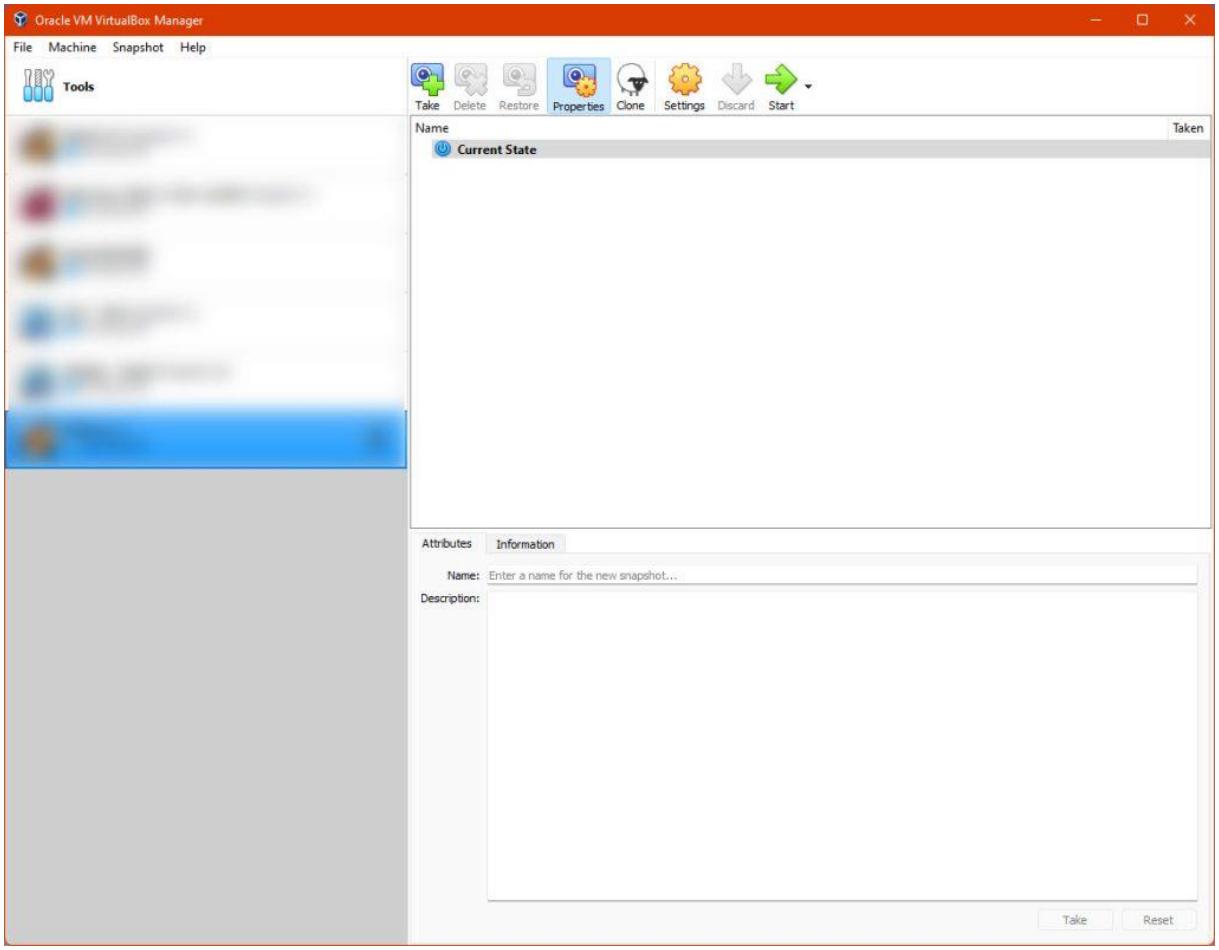
En caso de que usted quiera un entorno funcional listo para funcionar, le recomendamos que utilice Remnux, una máquina virtual (VM) descargable preconfigurada con algunas herramientas útiles para el análisis de malware. Aquí utilizaremos Virtualbox para virtualizar la máquina Remnux, si ya está familiarizado con este proceso, no dude en pasar a la siguiente sección del curso.

Instalar Virtualbox

Para empezar, necesitaremos un programa para gestionar nuestras máquinas virtuales. Nosotros elegimos Virtualbox porque es la solución más utilizada, es compatible con las tres plataformas más importantes (Windows, macOS y Linux) y es gratis. Para descargar el instalador respectivo visite https://www.virtualbox.org/ y busque el gran botón azul. A continuación, busque la sección con los paquetes por plataforma como se muestra en la imagen.



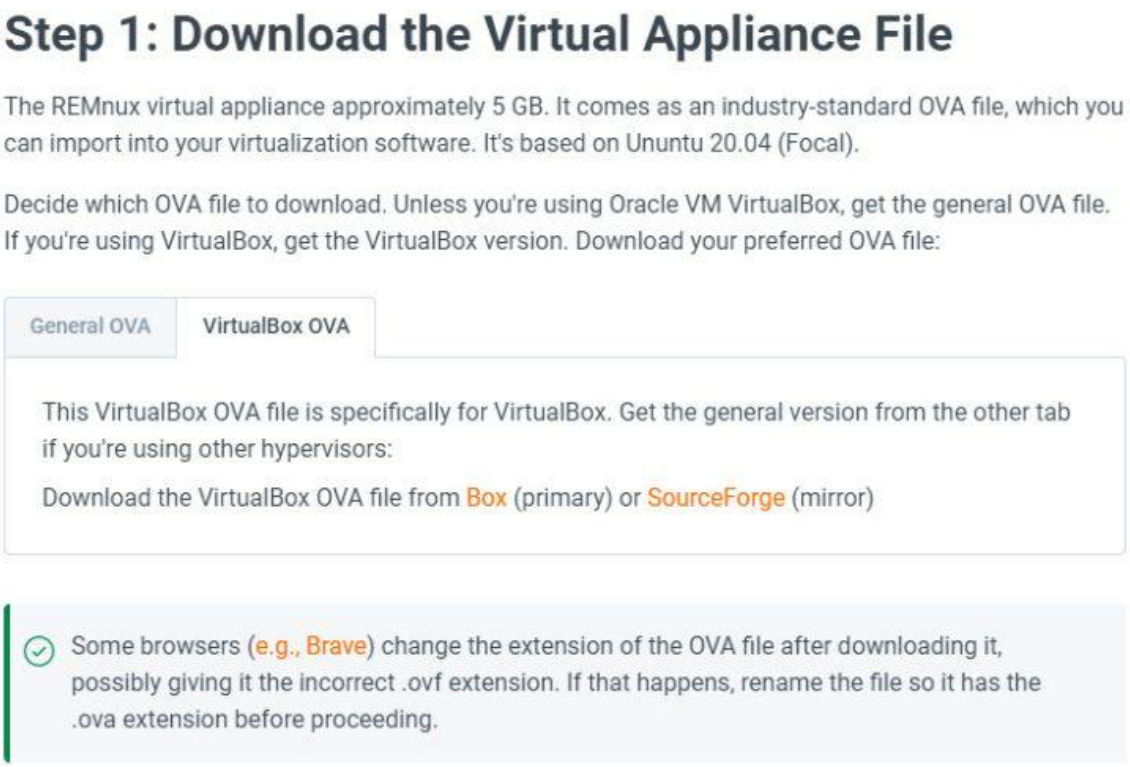
Aquí, haga clic en su plataforma y luego siga las instrucciones. Después de esto, usted puede ejecutar Virtualbox y ver una ventana como esta



Todavía no tendrá nada en la zona borrosa, a partir de aquí estamos listos para descargar e instalar Remnux

Instalar Remnux

Ahora, usted puede ir a https://remnux.org/ y hacer clic en "Descargar" en la sección correspondiente. Es posible que sea redireccionado a otra página pidiéndole que seleccione si desea descargar un OVA General o un OVA de Virtualbox, en nuestro caso, la última será la correcta.

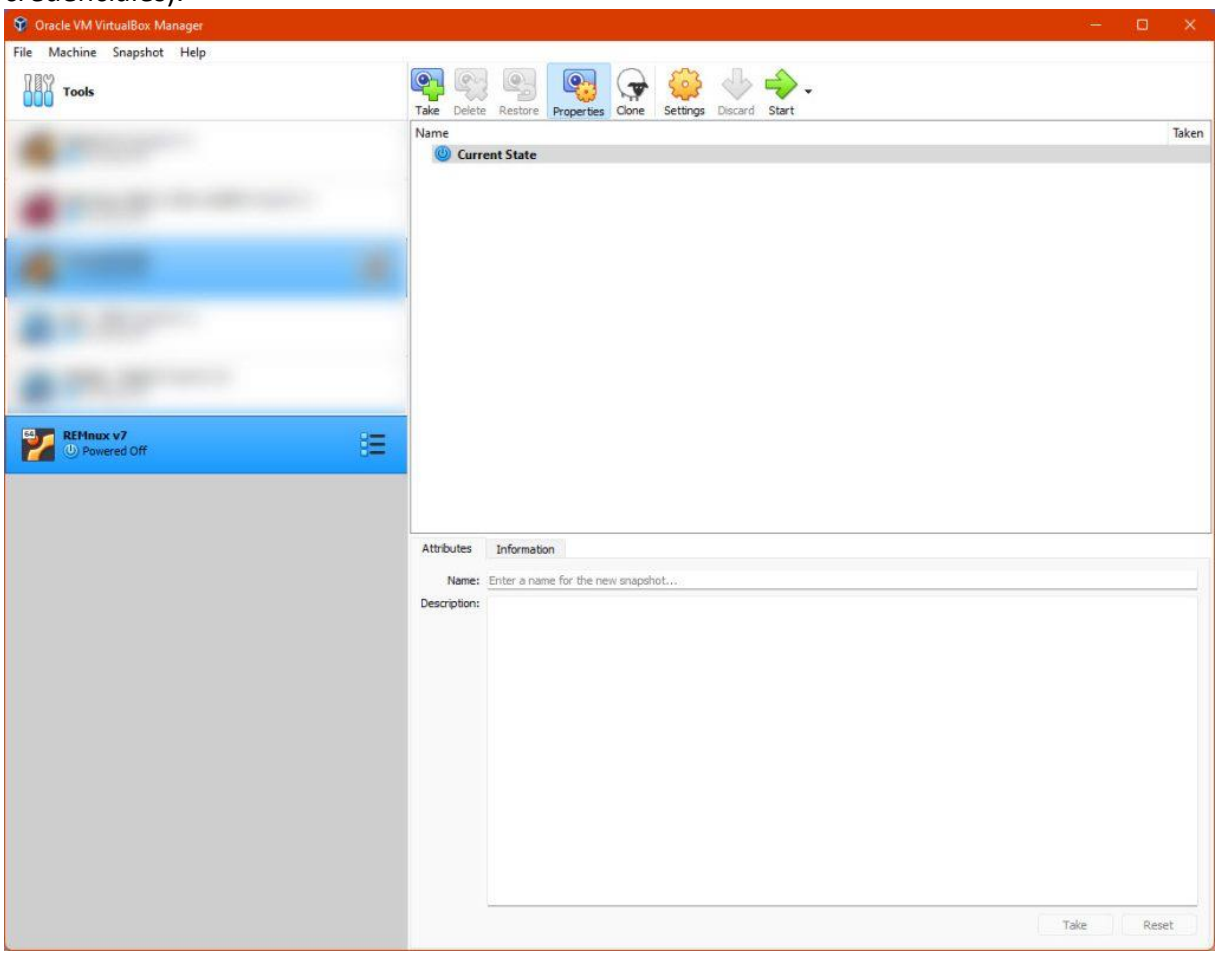


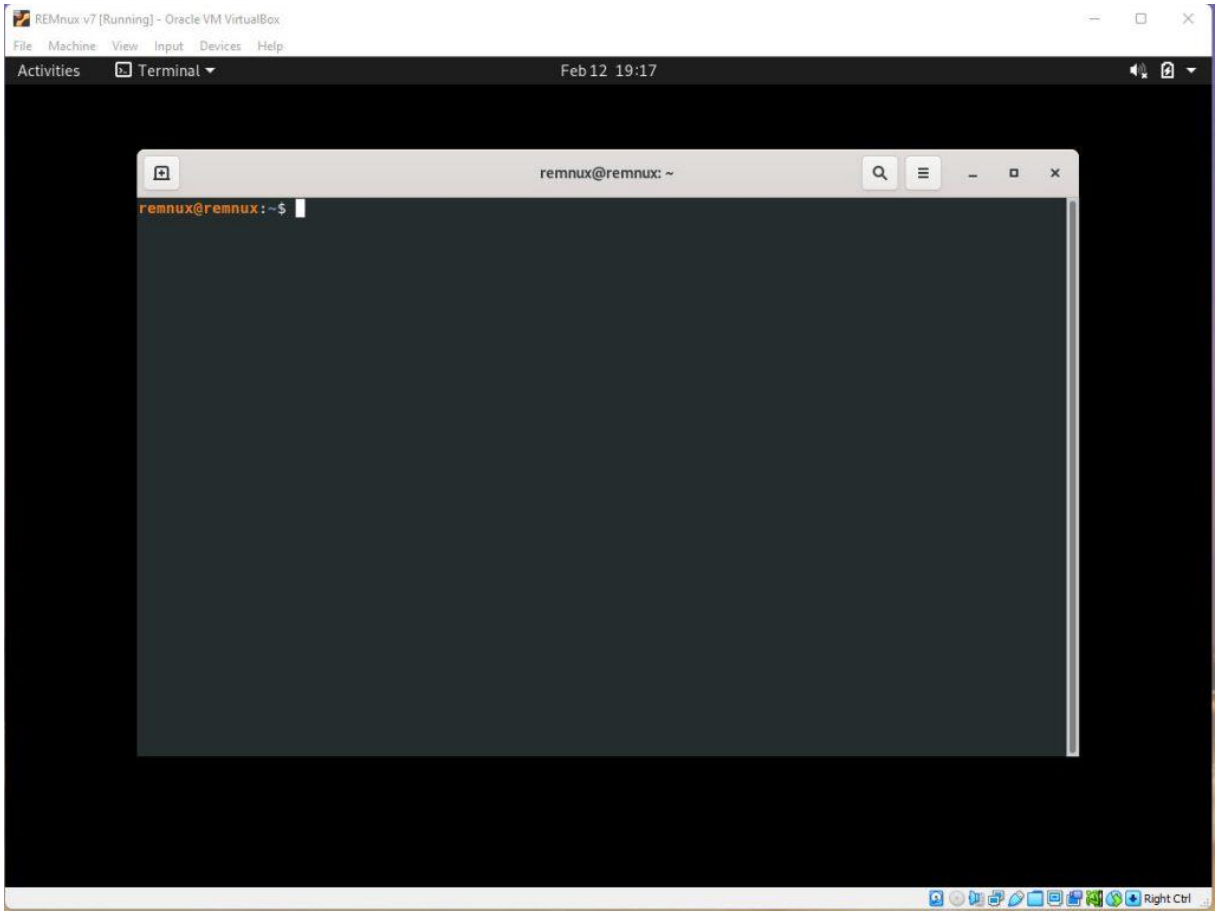
Después de descargar el archivo, se recomienda comprobar si el archivo se ha descargado correctamente, para ello, necesitamos comprobar el hash asociado del archivo. El hashing es un tema complejo que recomendamos aprender y aplicar (además, es muy utilizado en el análisis de malware). Sin embargo, por ahora, podemos resumirlo como un proceso matemático que transforma un conjunto de datos (como un texto o un archivo), en un código alfanumérico. Este código debe ser único para los datos que están analizando, y aún con pequeños cambios, el hash cambiará significativamente, por lo que verificar que nuestro archivo descargado tiene el mismo hash publicado en la página web de Remnux, nos dirá que el archivo se descargó sin problemas. Si el hash es diferente, podría ser una señal de que el archivo se corrompió debido a un proceso de descarga defectuoso o de alguna manera no es el archivo correcto (tal vez un error por nuestra parte al seleccionar la versión correcta, o como escenario remoto, alguien cambió el archivo por una versión maliciosa, así que hay que estar alerta). En https://technastic.com/check-md5-checksum-hash/ encontrará una referencia rápida sobre cómo verificar los hashes.

Descarga del archivo

Después de verificar que nuestro archivo se descargó correctamente, ahora podemos importarlo a Virtualbox. En la página de Remnux donde descargamos la VM, hay instrucciones disponibles, sin embargo, es suficiente con hacer doble clic en el archivo .ova, y un asistente nos guiará a través del proceso de importación. Podemos dejar todo como se sugiere en la configuración propuesta. Al final, deberíamos ver la máquina Remnux en nuestra ventana de Virtualbox. Al hacer clic en "Inicio" se encenderá nuestra máquina en una ventana independiente. Esta es una máquina Linux, y para iniciar sesión el usuario es

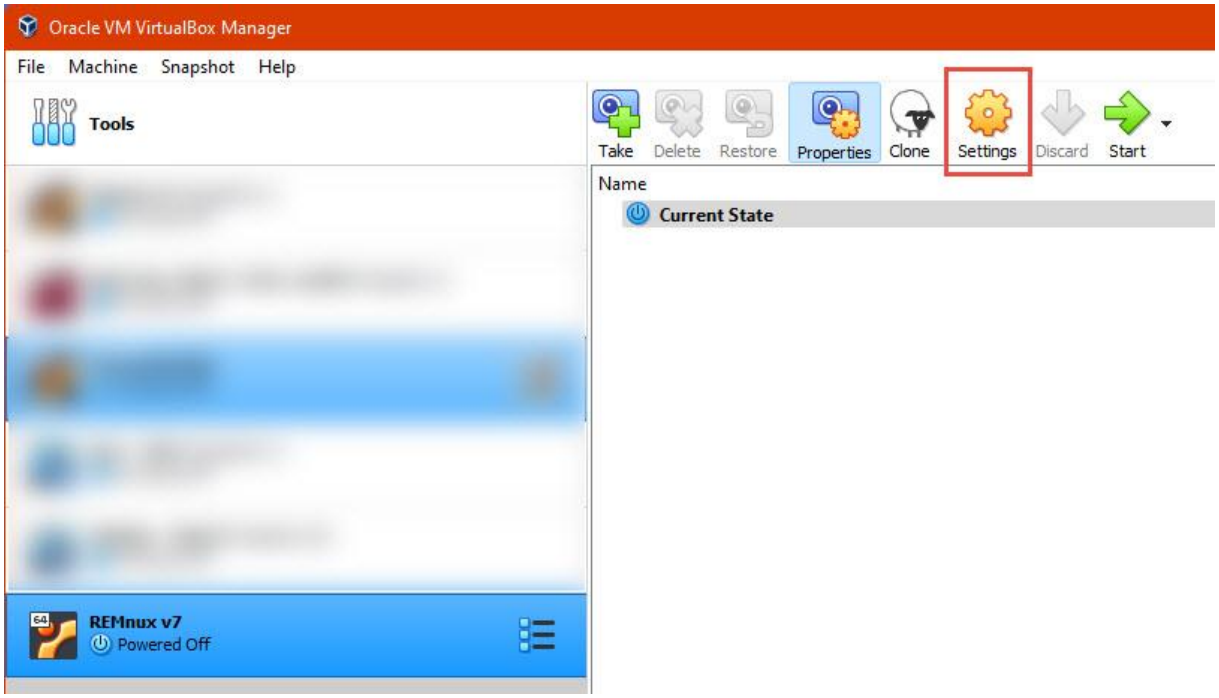
remnux y la contraseña *malware* (sin embargo, es posible que se abra la sesión sin solicitar credenciales).



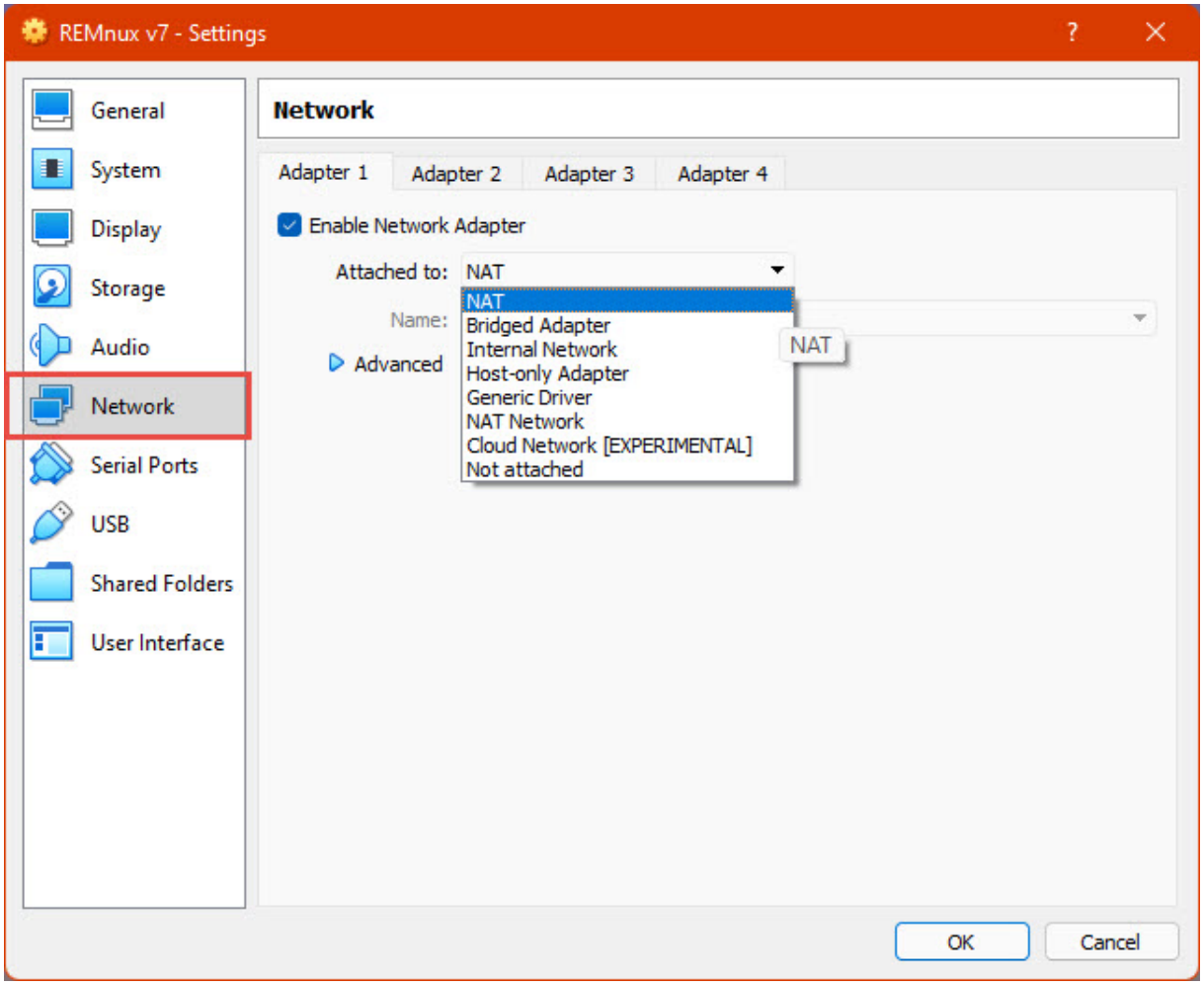


Configuraciones adicionales en Virtualbox – Red

Puesto que estaremos analizando archivos potencialmente dañinos, no es recomendable ejecutar la máquina de forma que pueda comunicarse con el resto de nuestra red. La estrategia específica puede variar dependiendo del estilo del analista, sin embargo, la configuración se realiza principalmente en la pantalla de interfaces de nuestra VM. Con nuestra máquina Remnux apagada, hacemos clic en el botón "Configuración" de la barra de herramientas.



A continuación, en la sección "Red", usted dispondrá de una serie de opciones, las más importantes son:

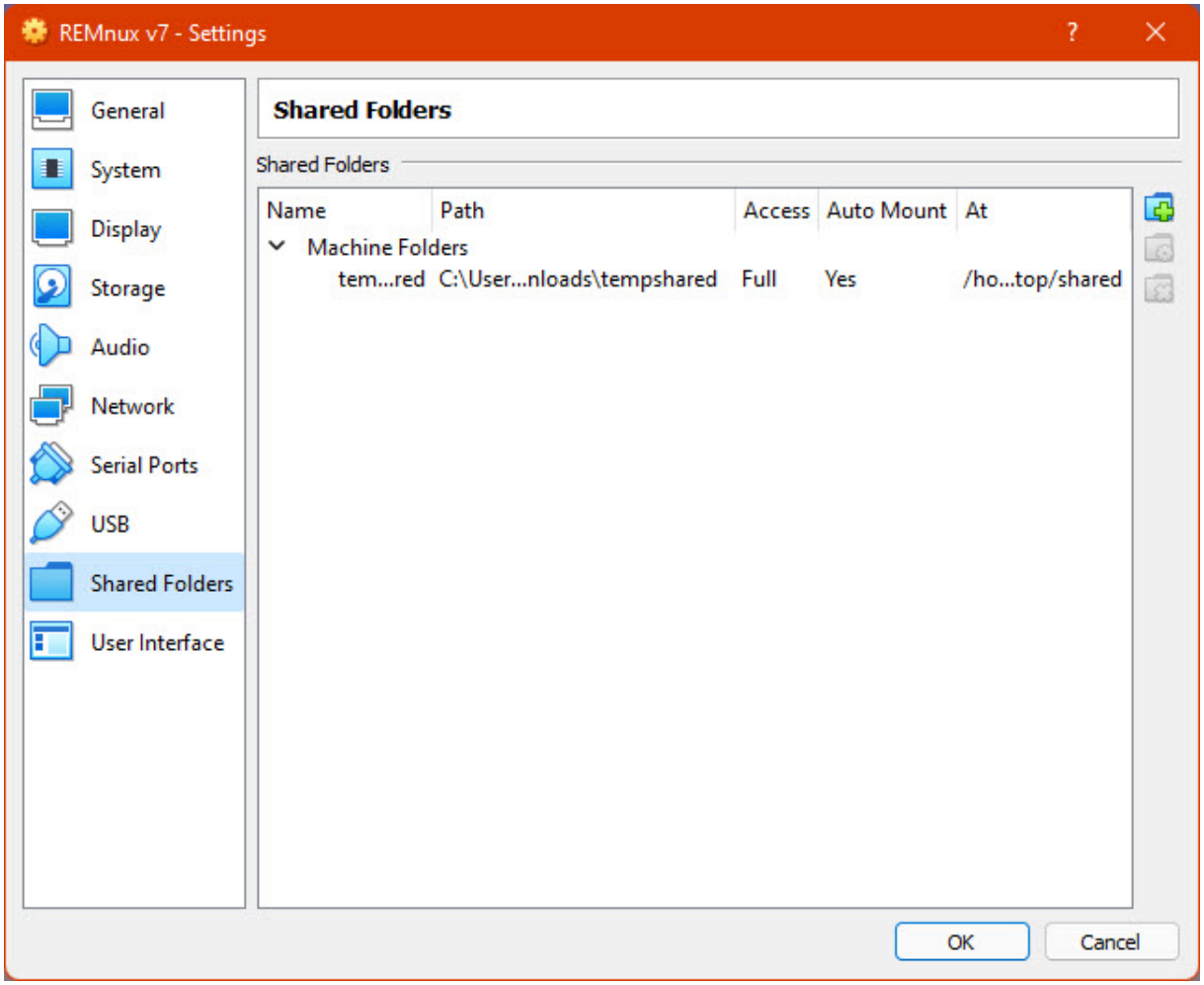


- **Habilitar adaptador de red:** deshabilitar esta opción eliminará cualquier conectividad entre nuestra VM y otros dispositivos a través de la red (incluyendo el nuestro, lo gestionaremos mediante la interfaz gráfica). Esta configuración emulará la ausencia de hardware para conectarse a cualquier red en la VM.
- **Conectado a – NAT:** la configuración predeterminada emulará una nueva red para la VM. Esto le permite acceder a internet, pero también a otros dispositivos de nuestra red. No se recomienda para el tipo de uso que le daremos a nuestra VM.
- **Conectado a – Adaptador puente:** esta opción compartirá el adaptador de red de nuestra computadora host física a la VM, poniéndola como cualquier otro dispositivo de nuestra red. Esta configuración tampoco es recomendable para nuestro caso de uso.
- **Conectado a – Adaptador solo host:** esta configuración conecta la VM a una red que únicamente se comunica con nuestra máquina host y otras VM con la misma configuración. En algunos casos, esto puede resultar útil; sin embargo, esto también puede exponer nuestra máquina a actividades maliciosas.
- **Conectado a – Red interna:** similar a la configuración anterior pero en este caso nuestra máquina host no es accesible. Esto es útil cuando queremos ver cómo dos o más máquinas interactúan entre sí.
- **Conectado a – No conectado:** esta configuración emula un adaptador de red sin un cable conectado a él.

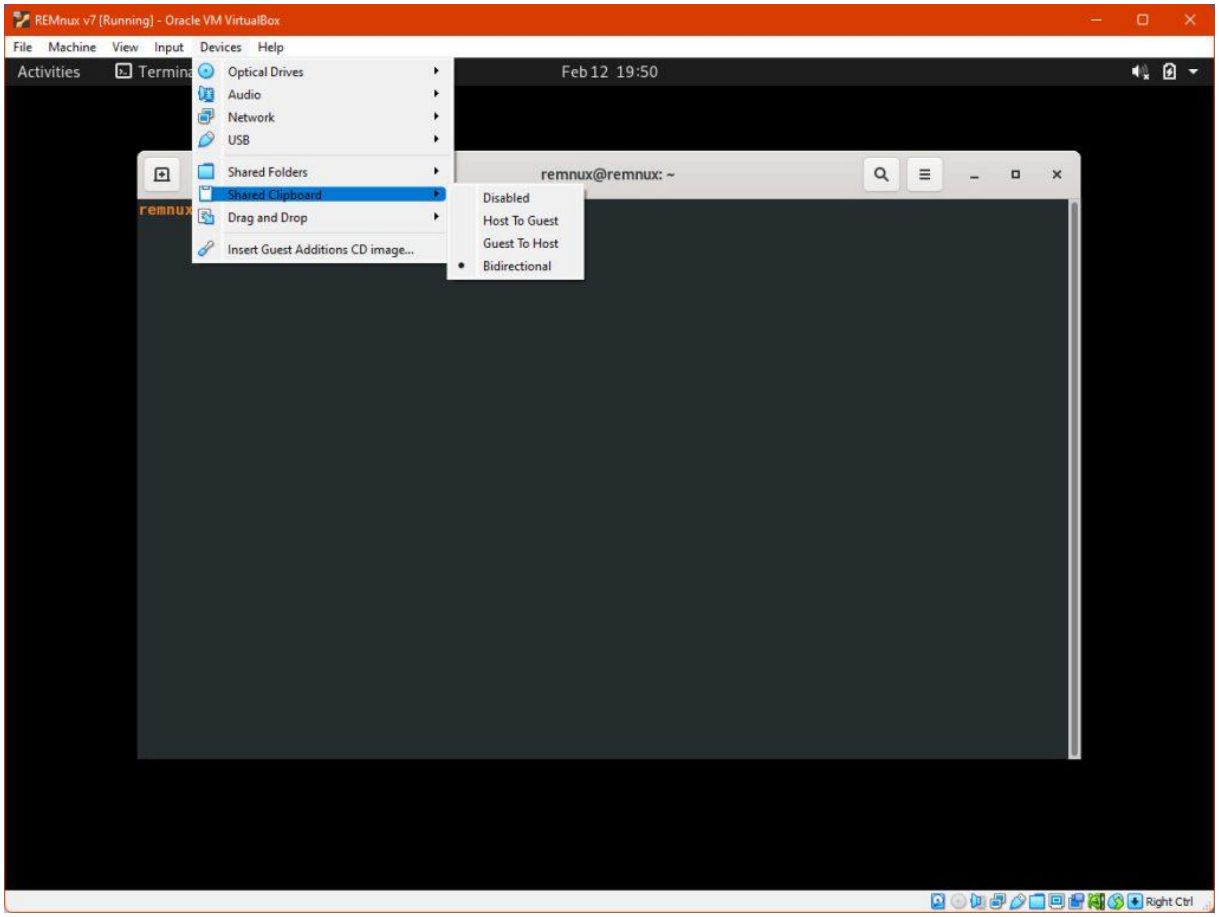
Dependiendo del uso que le demos a nuestra máquina, para la configuración inicial podemos mantener habilitado NAT para acceder a Internet y descargar herramientas, etc. Antes de comenzar nuestro análisis, podemos cambiarlo a No conectado, Red interna o deshabilitar el adaptador.

Configuraciones adicionales en Virtualbox – Compartir información con la máquina host

Es muy común compartir archivos y otros datos entre nuestra computadora y la VM, de nuevo, hay diferentes enfoques que podemos adoptar: **Carpetas compartidas:** similar a una carpeta compartida en red, podemos sincronizar una carpeta entre nuestro sistema host y nuestro sistema guest (la máquina virtual). Sin embargo, no siempre se recomienda compartir muestras de malware, ya que se abriría un espacio en nuestra computadora que es controlado por nuestra VM, la cual puede ser infectada durante nuestro análisis. Para configurar las carpetas compartidas, existe una sección dedicada en la configuración.



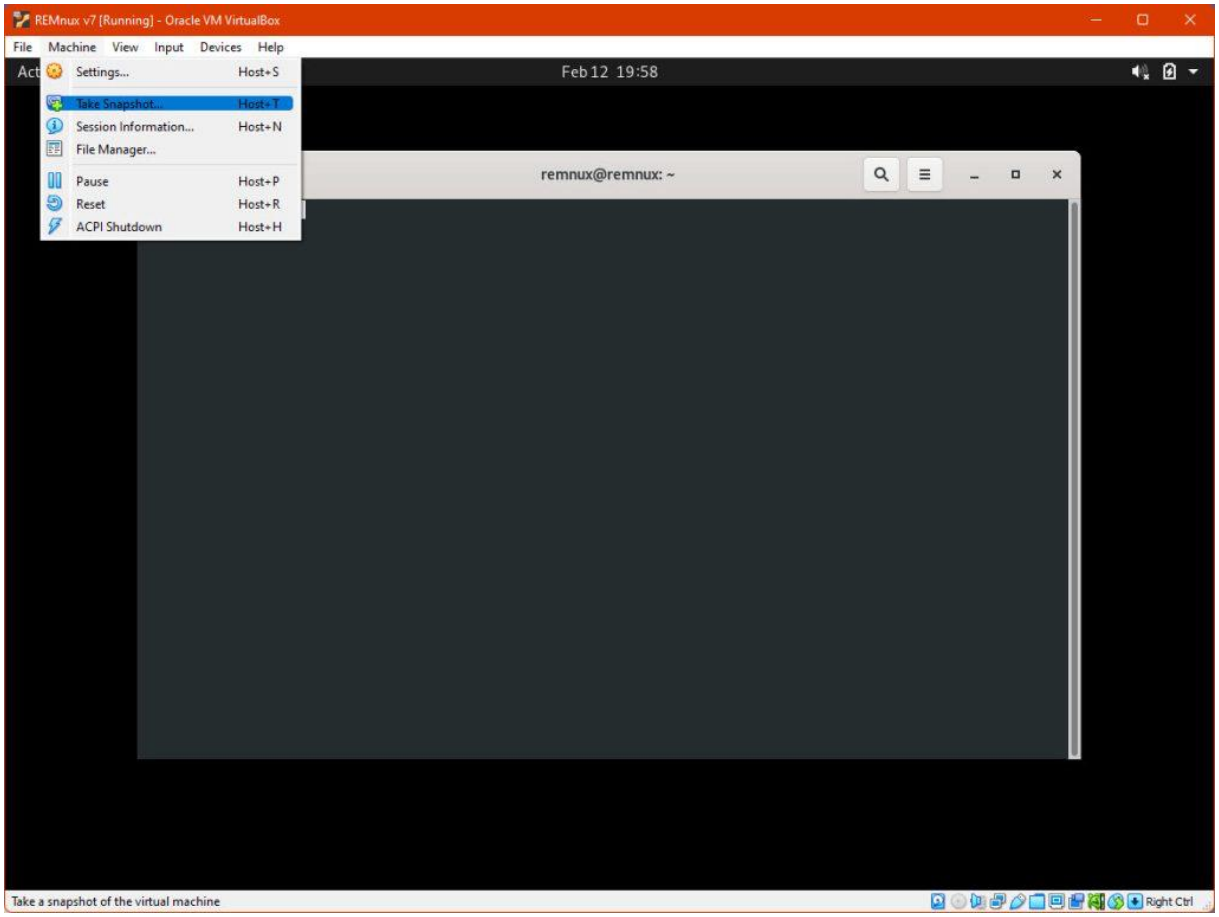
- **Portapapeles compartido y Arrastrar y soltar:** Esto nos permitirá compartir el portapapeles entre nuestra computadora y la VM. Esta opción puede ser desactivada, configurada como unidireccional o bidireccional, como se sugiere en la imagen. De manera similar, se puede habilitar arrastrar y soltar archivos entre el sistema host y el guest. Para algunos, deshabilitar las carpetas compartidas y habilitar el arrastrar y soltar sólo de "Host a Guest" es la opción más segura para proteger nuestras computadoras físicas, similar con el portapapeles compartido, Sin embargo, en algunos momentos podríamos necesitar extraer información de la VM.



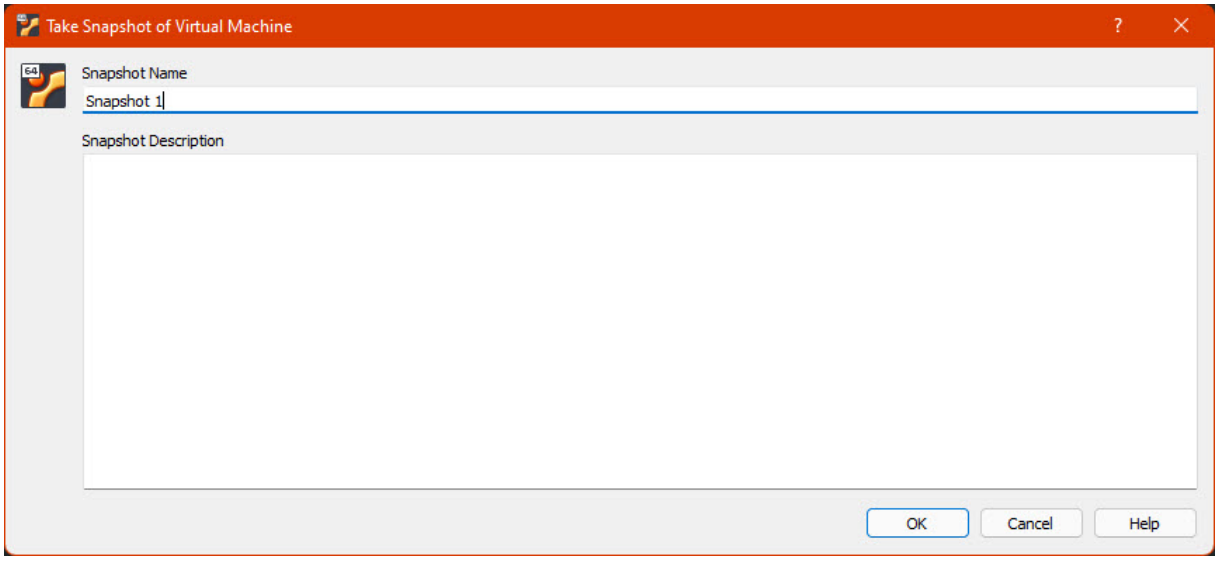
Configuraciones adicionales en Virtualbox – Instantáneas

Una función muy útil de Virtualbox es guardar una versión de la VM a la que podamos regresar en cualquier momento en el futuro. Por ejemplo, si configuramos la máquina Remnux para analizar malware, podría ser útil guardar una instantánea antes de comenzar el análisis, de forma que cuando hayamos terminado, podamos restaurar la VM a la instantánea guardada para estar seguros de que la máquina no está infectada, y estemos listos para continuar el análisis.

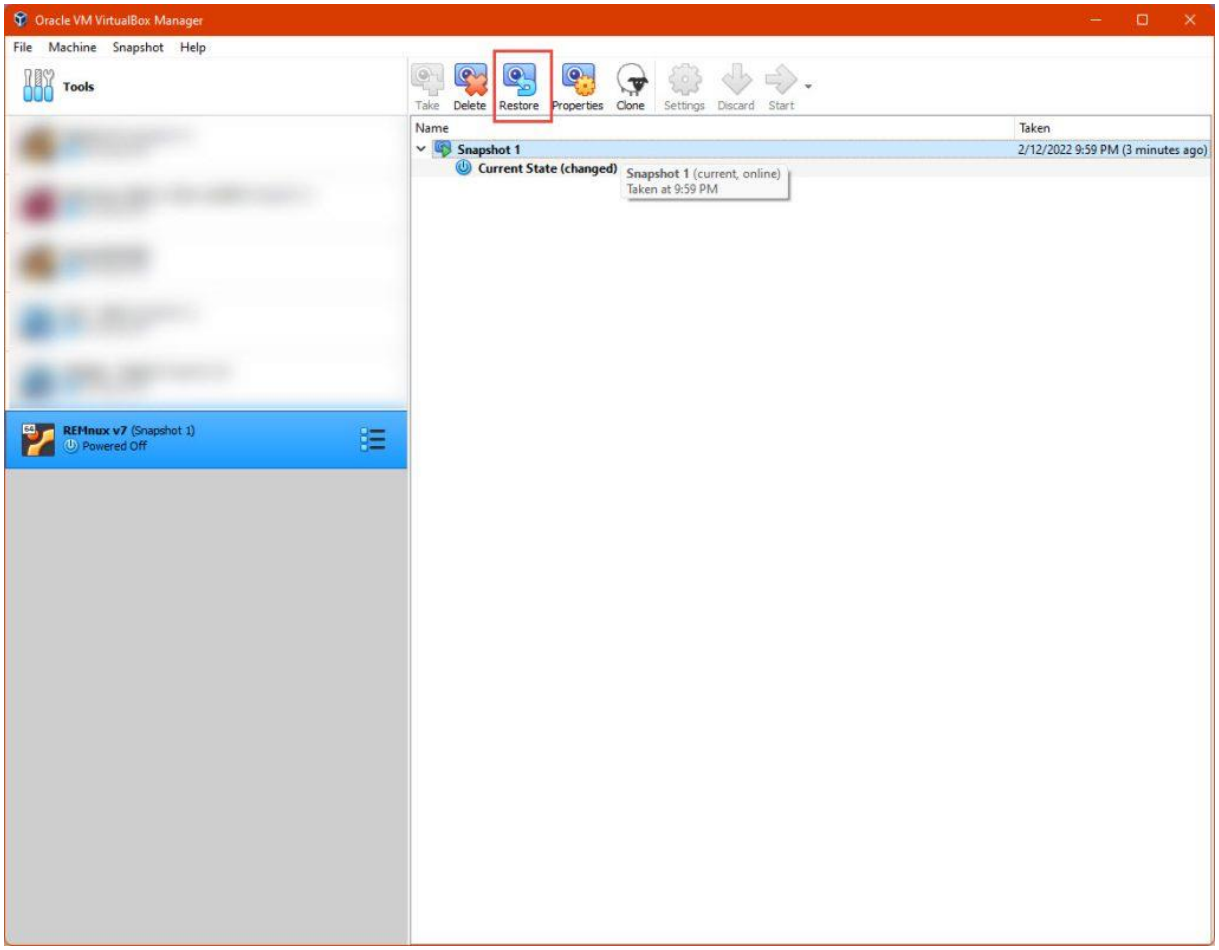
Para guardar una instantánea, con la máquina en el estado deseado, haga clic en "Máquina" y luego en "Tomar instantánea"



A continuación, seleccione un nombre y haga clic en "Aceptar". Esto tardará algún tiempo en crearse y después estará disponible en la sección Instantáneas de la pantalla principal de Virtualbox para nuestra VM.



Podemos usar el botón "Restaurar" en la pantalla correspondiente



Y ahora, ¿qué sigue?

Dado que podemos manejar lo básico con Virtualbox, podemos aprender sobre Remnux mientras entendemos y analizamos nuestro primer formato de archivo: PDF.