Analyse de documents malveillants (partie 1) : introduction et MV

Français

Introduction

Ce (mini)cours s'adresse aux adeptes de la sécurité numérique et aux praticiens (assistance technique, formateurs, premiers intervenants, etc.) qui veulent en savoir plus sur les documents malveillants et la façon de les identifier. Ces documents peuvent être des pièces jointes à un courrier électronique, des fichiers stockés sur des lecteurs flash ou des téléchargements de sites Web spécifiques. Les principaux objectifs sont :

- Apprendre les bases du fonctionnement des formats de documents courants et la façon dont ils peuvent être utilisés comme armes, en mettant l'accent sur les fichiers au format document portable (PDF) et Microsoft Office (au moins à partir de MS Word, Excel et PowerPoint).
- Présenter quelques outils qui pourraient contribuer à repérer les signes de documents dangereux ou confirmer que leur ouverture est sans danger.
- Donner des conseils de sécurité et clarifier les doutes courants sur la façon de traiter les fichiers suspects.

Ce cours utilise le format de lectures courtes et de questionnaires dans la plupart des contenus couverts, où il sera nécessaire d'exécuter certains outils en fonction du contenu. Ce point sera traité dans la section Environnement de travail peu après l'introduction, les exigences générales étant :

Compléter les exercices proposés :

- Capacité de faire fonctionner 1) une machine virtuelle sur l'ordinateur en utilisant Virtualbox ou un logiciel similaire, ou 2) des scripts python (uniquement pour analyser les fichiers du cours, pas pour analyser des échantillons réels).
- Le temps de passer en revue l'ensemble du contenu (environ 2 heures)

Ce cours reprend les documents disponibles dans d'autres références et n'utilise que des outils qui sont librement disponibles. La majorité du contenu est inspirée par le travail de Didier Stevens au fil du temps, en particulier pour SANS, ainsi que d'autres références. Voici une courte liste :

- 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

Structure

1. Avertissements
2. Quelques considérations sur la modélisation des menaces
3. Pour chaque type de format de fichier (PDF, MS Office)
    1. Comment ils sont structurés (d'une manière plus technique)
    2. Comment ils peuvent être utilisés comme armes
    3. Comment nous pouvons faire une analyse introductive
    4. Quelques conclusions/faits sur le format du fichier
4. Quelques conseils généraux contre les menaces connexes
5. Et ensuite ?

Ensuite, une série d'avertissements utiles avant de commencer à aborder le contenu

Avertissements

Compte tenu de la nature de la tâche que nous allons effectuer après avoir maîtrisé le contenu fourni (l'analyse des fichiers malveillants et dangereux) et la complexité du sujet (que nous voyons comme une introduction à l'analyse des logiciels malveillants), nous recommandons fortement de lire cette section et d'accepter tous les points avant de passer à la suite.

1. **Ce cours est une introduction :** il est conçu pour les personnes sans expérience préalable en matière d'analyse des documents suspects. Dans la section Étapes suivantes, nous avons inclus une liste de ressources à lire et à consulter.
2. **Ce cours ne couvre pas de nombreuses techniques avancées :** il existe de nombreuses menaces spécifiques dont la complexité dépasse le cadre de ce document, et comme dans tout ce qui touche à la sécurité de l'information, il peut exister des menaces qui n'ont pas encore été découvertes et qui ne seront pas abordées dans ce cours. Nous recommandons de chercher de l'aide dans le cas où nous soupçonnons une menace avancée ou inconnue dans un fichier ou tout autre artefact, plus encore dans les sections ultérieures. Cela dit, ce cours nous aidera à mieux comprendre l'apparence des fichiers bénins plutôt que la structure de chaque document malveillant.
3. **Prenez vos précautions lors de l'analyse des fichiers réels :** les échantillons utilisés dans ce cours sont inoffensifs, mais si vous répétez les flux de travail présentés dans des échantillons réels sans la sécurité respective mesurée, il est probable que vous infectiez votre appareil. Veuillez ne pas exécuter de fichier suspect sur votre ordinateur principal, utilisez une machine virtuelle, un périphérique dédié ou un environnement afin de ne pas exécuter le fichier sur votre machine, mais seulement analyser ses propriétés.

Questionnaire d'avertissement

Question 1 : je comprends les risques liés à l'analyse de fichiers suspects, les conséquences possibles d'exécuter des programmes malveillants délibérément ou accidentellement, j'ai lu le contenu de cette page/section et je comprends les stratégies les plus courantes pour contrer ces menaces potentielles.

Question 2 : laquelle des options suivantes décrit le mieux ce que nous devons faire lors de l'analyse d'un fichier suspect authentique ?

- 1. Nous devrions démarrer une machine virtuelle (VM) ou un ordinateur dédié pour analyser le fichier et lui donner le moins d'accès possible à notre machine hôte et au reste du réseau.

- 2. Nous pouvons analyser le fichier dans notre propre ordinateur/environnement, mais sans accès à Internet.

- 3. Nous devrions analyser le fichier dans un ordinateur ou une machine virtuelle (VM) avec des systèmes d'exploitation moins courants comme Linux ou macOS.

À propos des modèles de menace

Lorsque l'on cherche des conseils sur la façon de traiter les fichiers suspects, l'approche proposée consiste généralement à éviter toute interaction avec les fichiers, par exemple :

- Ne pas ouvrir les fichiers inconnus.
- Ne pas interagir avec les fichiers suspects.
- Ne pas regarder les fichiers suspects dans les yeux.

Ou encore, nous pouvons trouver un autre type de conseil qui, bien qu'il soit suffisant pour la plupart des gens, pourrait être trompeur pour les utilisateurs sensibles comme les militants des droits humains ou les journalistes travaillant dans des environnements dangereux, ou simplement contre-productif, par exemple :

- Utiliser un antivirus est suffisant pour vous protéger contre les fichiers malveillants.
- Seuls les documents Microsoft Office comprenant des macros sont dangereux, vous pouvez donc traiter les autres types de fichiers sans trop vous inquiéter.
- Supprimer tout courrier électronique contenant des pièces jointes suspectes. Cette situation est particulièrement préoccupante dans certains scénarios, car si nous supprimons les e-mails et les pièces jointes de notre boîte de réception, nous perdons des éléments de preuve clés qui peuvent nous aider à évaluer si les artefacts sont effectivement malveillants ou ciblés, ce qui pourrait être une information inestimable.

En pratique, lorsque nous travaillons avec des communautés ciblées (notamment des journalistes), ne pas interagir avec les fichiers n'est pas une option. De nombreuses organisations, groupes et personnes doivent ouvrir des fichiers potentiellement dangereux

dans le cadre de leur travail, et ils le feront même en sachant les risques, voici quelques exemples :

- Des journalistes reçoivent une invitation à une conférence de presse.
- Des militants reçoivent un document justificatif comme preuve dans une affaire de violation des droits humains ou comme fuite d'informations.
- Une institution adverse envoie un document qui devrait être examiné et traité.

Un facteur supplémentaire à prendre en compte est que les acteurs de la société civile sont exposés à des menaces ciblées inconnues par les moteurs antivirus. Un autre facteur est que selon le type de l'attaque, d'autres formats de fichiers peuvent également être utilisés. Ces facteurs sont essentiels pour que les personnes qui aident les groupes vulnérables comprennent mieux comment les documents et autres formats de fichiers courants peuvent être utilisés comme armes, afin de donner des conseils utiles, mais aussi pour contribuer à analyser les dossiers spécifiques afin de comprendre et d'évaluer s'ils sont victimes d'attaques ciblées.

Avec tout cela à l'esprit, nous allons nous concentrer sur la compréhension de la façon dont les formats de fichiers standard sont structurés, comment repérer les attaques les plus courantes en les utilisant, et quelques mesures défensives mises à jour pour éviter d'être victimes de ce genre de menace.

Questionnaire sur le modèle de menace

Question 1 : pour une organisation hautement ciblée recevant de nombreux documents Microsoft Office par e-mail, laquelle des options suivantes est vraie ? (Une seule est correcte)

- 1. Même si le logiciel antivirus (AV) indique que le fichier est sûr, il peut contenir des logiciels malveillants.
- 2. Ils doivent supprimer immédiatement toutes les pièces jointes suspectes, car il pourrait être dangereux de les conserver dans la boîte de réception.
- 3. Ils ne doivent ouvrir aucune pièce jointe provenant de sources inconnues.

Environnement : considérations générales

Pour exécuter la plupart des tâches de ce cours, nous utiliserons des outils de base écrits en langage de programmation Python, étant donné la grande compatibilité de Python avec tous les systèmes d'exploitation. Il existe d'innombrables façons de configurer un environnement et nous vous proposons une version spécifique, mais si vous êtes familiarisé(e) avec Python, l'analyse de logiciels malveillants et/ou la virtualisation, vous pouvez configurer une autre version mieux adaptée à vos besoins. La seule exigence indispensable serait d'avoir un

environnement isolé pour manipuler les artefacts dangereux (fichiers dans ce cas-ci). Il y a d'autres éléments à prendre en compte, mais cette exigence est la plus importante.

Environnement isolé et autres bonnes pratiques

Les échantillons utilisés dans ce cours sont inoffensifs et servent seulement à démontrer comment les fichiers sont structurés et comment repérer les signaux d'alerte. Cependant, si vous avez l'intention d'analyser des fichiers réels, il est probable que vous trouviez des fichiers infectés susceptibles de causer toutes sortes de problèmes, comme infecter l'ordinateur que vous utilisez, compromettre vos informations ou rendre votre appareil inutilisable, entre autres. Cela dit, il est courant d'avoir un environnement exclusif pour analyser et exécuter des échantillons suspects de manière contrôlée, donc si un problème se présente pendant la manipulation de l'échantillon, vous ne subirez aucun incident sur votre appareil ou les informations qu'il contient.

Un autre avantage d'avoir un environnement dédié est qu'après avoir manipulé des échantillons de logiciels malveillants, vous pouvez tout supprimer et recommencer sans craindre de perdre des fichiers non liés. Cela nous permet de planifier des moyens pratiques pour « réinitialiser » notre environnement à un état prêt à l'emploi avant chaque analyse.

L'une des stratégies les plus utilisées pour garantir un environnement isolé est d'utiliser des machines virtuelles (VM), qui émulent essentiellement un ordinateur complet à l'intérieur d'un autre ordinateur, y compris le système d'exploitation (OS), les disques durs, l'écran, etc. Les outils communs pour configurer et utiliser des VM sont Virtualbox et VMware Workstation Player, entre autres. L'utilisation d'un matériel dédié est également une option valable tant qu'elle est sécurisée en cas d'infection.

Un inconvénient potentiel pourrait être que certains logiciels malveillants incluent du code pour vérifier s'ils sont exécutés dans des environnements isolés afin d'éviter toute exécution, ce qui complique leur analyse. Cependant, le danger inhérent d'exécuter des logiciels malveillants dans nos environnements quotidiens exclut toute tentative, nous recommandons alors de chercher de l'aide, en vous concentrant sur des techniques qui ne reposent pas sur l'exécution des fichiers suspects, ou d'obtenir des informations sur la façon de mettre en place un environnement qui ressemble à une véritable machine pour les échantillons de logiciels malveillants. Pour cette ressource, cela ne devrait pas être un problème puisque nous n'exécuterons aucun code à partir de documents, mais si vous voulez apprendre et effectuer une analyse dynamique sur des fichiers suspects, cela pourrait s'avérer utile.

Autres considérations

Outre la bonne pratique d'avoir un environnement isolé, les autres pratiques courantes sont :

- **S'assurer que l'ordinateur que vous utilisez n'est pas connecté à Internet ou au réseau local :** surtout si vous ouvrez des fichiers suspects. La raison la plus fréquente est d'éviter de déclencher des signaux qui alerteront les opérateurs de logiciels

malveillants que le code est en cours d'exécution ou testé à partir d'autres données comme l'adresse IP ou le type de périphérique exécutant le logiciel malveillant. En outre, certains logiciels malveillants vont tenter de se propager sur le réseau local en essayant d'infecter d'autres appareils non visés à l'origine, il est donc courant d'isoler les dispositifs de test dans différents réseaux physiques ou virtuels (ou VLAN). N'oubliez pas que dans le cas d'une analyse d'un échantillon via son exécution, il est possible que le logiciel malveillant détecte qu'il n'a pas accès à Internet et évite toute exécution.

- **Si vous vous connectez à Internet, utilisez un VPN ou un outil similaire :** l'idée est de cacher votre emplacement réel au cas où le logiciel malveillant que vous analysez fonctionne et le signale à ses opérateurs. Encore une fois, il n'est généralement pas recommandé d'exécuter des logiciels malveillants sans prendre de mesures pour éviter toute communication potentielle avec leurs opérateurs. Cependant, l'utilisation d'un VPN pourrait être une bonne mesure en cas d'exécution accidentelle ou si d'autres configurations échouent à un moment donné.
- **Organiser un processus pour réinitialiser votre environnement à un état « propre »** : selon que vous utilisez une machine virtuelle ou un matériel dédié, il existe certains outils et fonctionnalités utiles pour réinitialiser l'environnement afin de garantir que la machine soit propre à chaque analyse d'un échantillon. Les VM utilisant des instantanés en sont un bon exemple et il existe des logiciels pour rétablir un ordinateur physique à un état précédent.
- **Tenez-vous-en à l'analyse statique :** en général, nous pouvons diviser l'analyse de logiciels malveillants selon que nous exécutons les échantillons ou non. L'analyse statique tente de disséquer les fichiers et autres artefacts pour recueillir autant d'informations que possible sans les exécuter, tandis que l'analyse dynamique exécute les échantillons pour voir quels changements ils apportent dans l'environnement de test. Selon le type de logiciel malveillant, un type d'analyse peut être plus utile que l'autre, mais en général, une analyse dynamique nécessitera plus de mesures pour protéger l'environnement de test et le réseau afin de pouvoir supporter l'exécution des véritables logiciels malveillants. Ce cours montre uniquement les techniques d'analyse statique.
- **Faites preuve de prudence lorsque vous publiez des échantillons ou d'autres renseignements sur les échantillons analysés :** en général, cela pourrait alerter les opérateurs des logiciels malveillants au sujet de notre analyse des logiciels malveillants, et les pousser à fermer l'infrastructure et nettoyer leurs traces pour rendre l'attribution plus difficile, etc. Cela s'applique à toute plateforme publique comme les réseaux sociaux et les sites Web, y compris certaines plateformes publiques sur lesquelles nous pouvons envoyer des fichiers pour les analyser dans le cloud à la recherche de signes d'alerte de la part des moteurs antivirus et de la communauté de la sécurité de l'information. Pour le dernier scénario, nous allons partager quelques exemples et techniques pour vérifier l'information dont nous avons besoin sans alerter personne.

Questionnaire sur l'environnement

Question 1 : laquelle des affirmations suivantes est vraie ?

- 1. L'exécution des échantillons nécessitera moins de mesures de sécurité que d'essayer de disséquer les artefacts pour obtenir des renseignements utiles.

- 2. Couper l'accès à Internet empêchera l'échantillon de logiciel malveillant d'informer ses créateurs au sujet de son exécution.

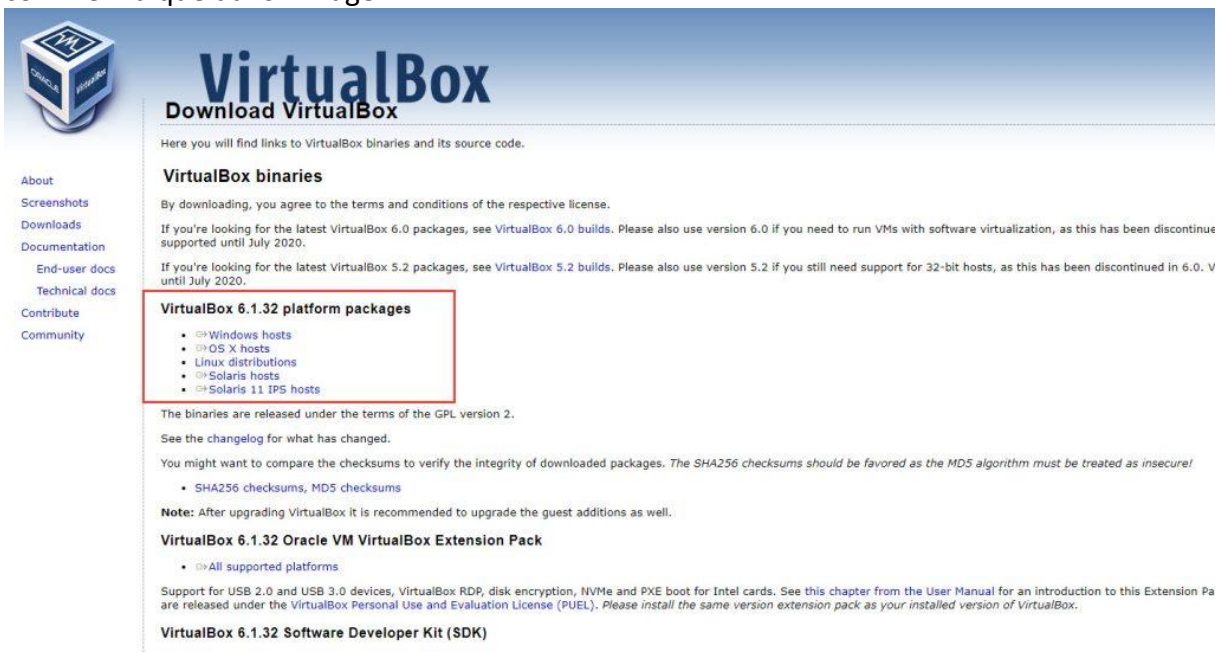
- 3. La façon la plus efficace d'analyser les logiciels malveillants consiste à utiliser des machines virtuelles, car si la machine devient infectée, nous pouvons la réinitialiser à partir de zéro.

Exemple d'environnement : Remnux + Virtualbox

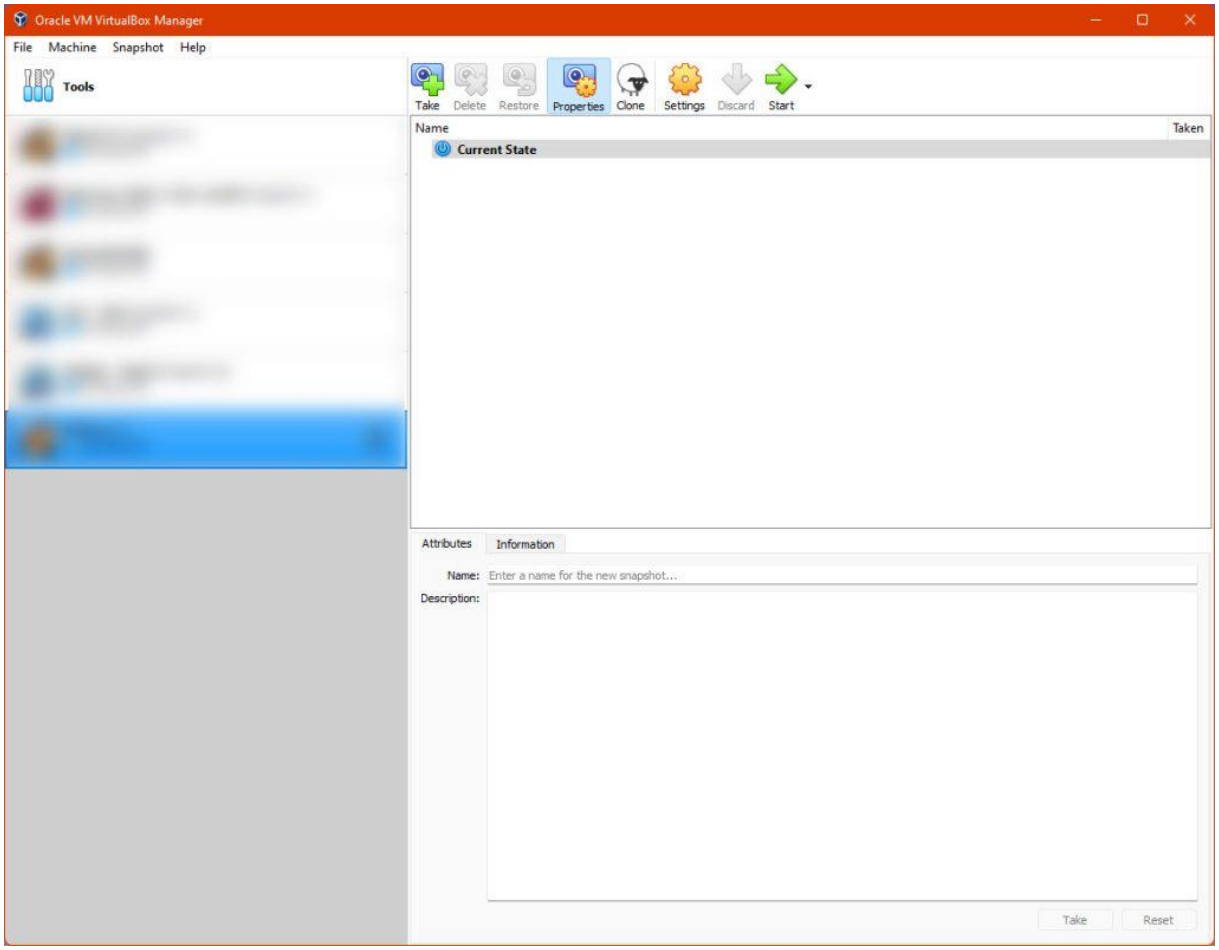
Si vous voulez un environnement fonctionnel prêt à l'emploi, nous vous recommandons d'utiliser Remnux, une machine virtuelle (VM) téléchargeable préconfigurée avec des outils utiles pour l'analyse de logiciels malveillants. Ici, nous allons utiliser Virtualbox pour virtualiser la machine Remnux. Si vous avez l'habitude de ce processus, n'hésitez pas à passer directement à la section suivante du cours.

Installer Virtualbox

Tout d'abord, nous aurons besoin d'un programme pour gérer nos machines virtuelles. Nous avons choisi Virtualbox, car cette solution est la plus utilisée et elle est compatible avec les trois principales plateformes (Windows, macOS et Linux), en plus d'être gratuite. Pour télécharger le programme d'installation respectif, visitez https://www.virtualbox.org/ et cherchez le gros bouton bleu. Ensuite, recherchez la section avec les paquets par plateforme comme indiqué dans l'image.



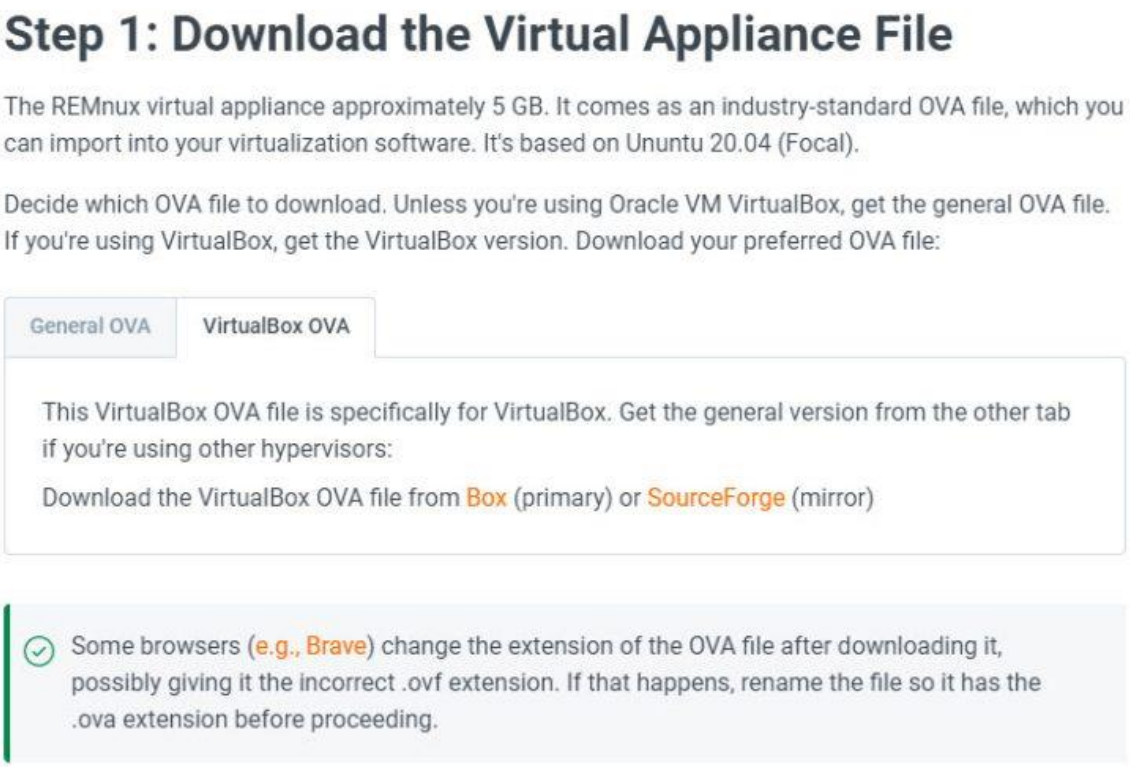
Cliquez ici sur votre plateforme, puis suivez les instructions. Ensuite, vous pourrez lancer Virtualbox et voir une fenêtre comme celle-ci :



Vous ne verrez rien dans la zone floue. À ce stade, nous sommes prêts à télécharger et installer Remnux.

Installer Remnux

Vous pouvez maintenant accéder à https://remnux.org/ et cliquer sur « Télécharger » dans la section correspondante. Il est possible que vous soyez redirigé(e) vers une autre page vous demandant de choisir si vous voulez télécharger un OVA Général ou un OVA Virtualbox. Dans notre cas, la deuxième option est la bonne.

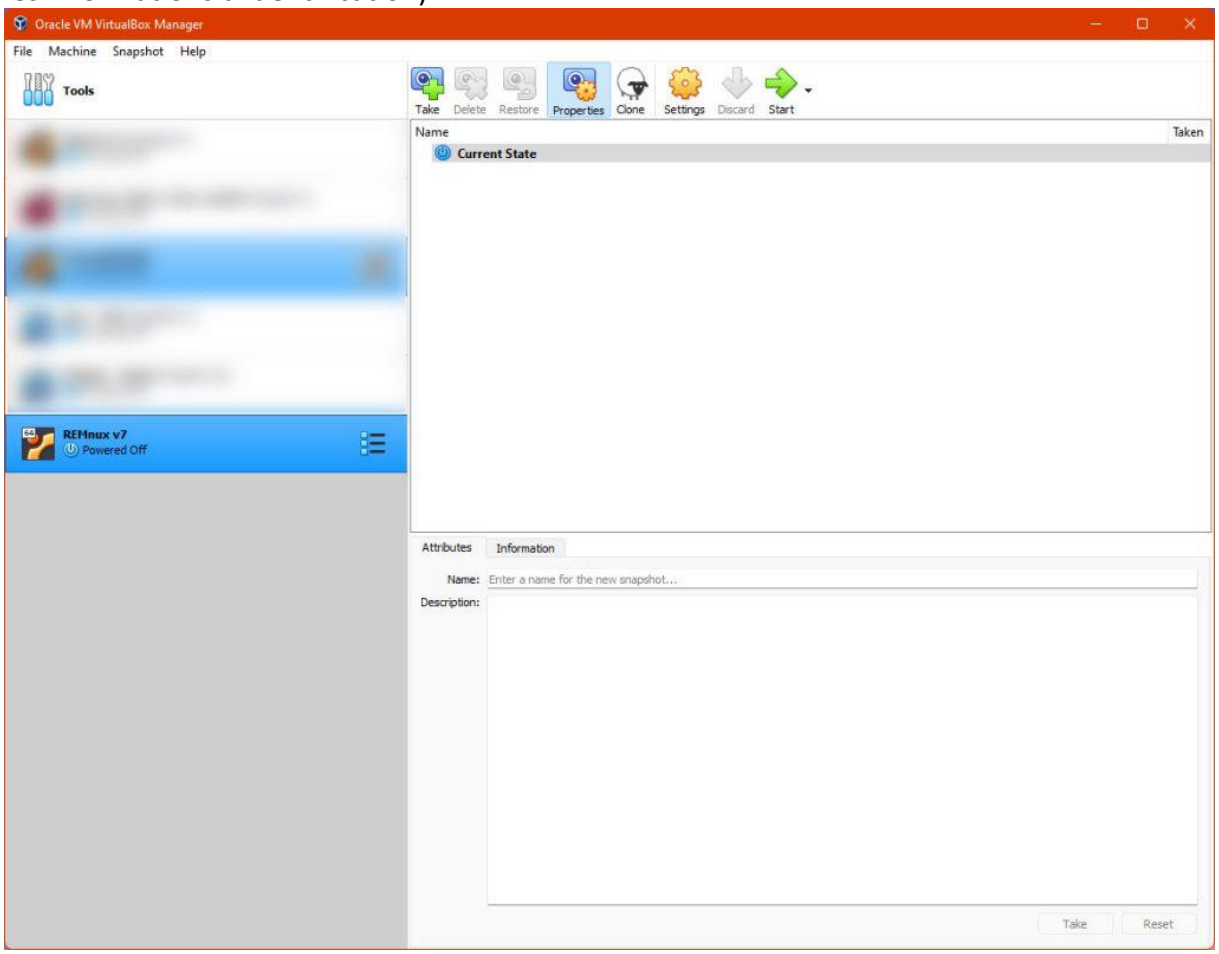


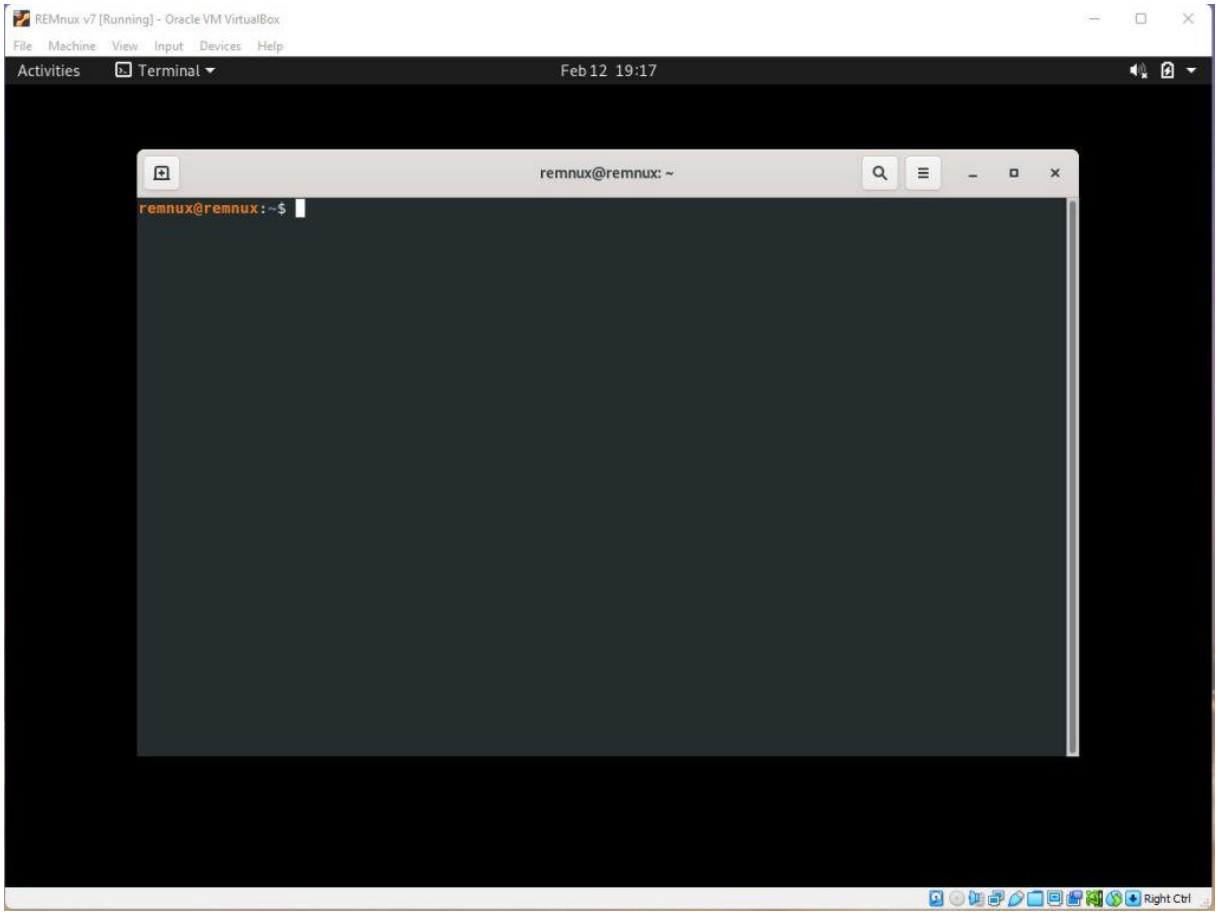
Après avoir téléchargé le fichier, il est recommandé de vérifier si le fichier a été téléchargé correctement. Pour ce faire, nous devons vérifier le hachage associé du fichier. Le hachage est un sujet dense que nous encourageons à apprendre et à appliquer (il est également très utilisé dans l'analyse des logiciels malveillants), mais pour l'instant, nous pouvons le résumer comme un processus mathématique qui transforme une partie de données (comme un texte ou un fichier) en code alphanumérique. Ce code doit être unique aux données analysées, et la moindre modification de celles-ci entraînera un changement important du hachage. Vérifiez donc que notre fichier téléchargé a le même hachage que celui publié sur le site de Remnux, cela garantira que le fichier a été téléchargé sans aucun problème. Si le hachage est différent, cela peut indiquer que le fichier a été corrompu à cause d'un processus de téléchargement défectueux ou qu'il ne s'agit pas du bon fichier. Cela peut être dû à une erreur de notre part lors du choix de la bonne version, ou dans un scénario plus distant, à la modification du fichier en faveur d'une version malveillante. Soyez sur vos gardes. Une référence rapide sur la façon de vérifier les hachages est disponible à https://technastic.com/check-md5-checksum-hash/

Télécharger le fichier

Après avoir vérifié que notre fichier a été téléchargé sans problème, nous pouvons l'importer dans Virtualbox. Des instructions sont disponibles sur la page Remnux où nous avons téléchargé la VM. Cependant, il suffit de double-cliquer sur le fichier .ova et un assistant nous guidera à travers le processus d'importation. Nous pouvons tout laisser tel que suggéré dans la configuration proposée. Au final, nous devrions voir la machine Remnux dans notre fenêtre Virtualbox. Cliquez sur « Démarrer » pour lancer notre machine dans une

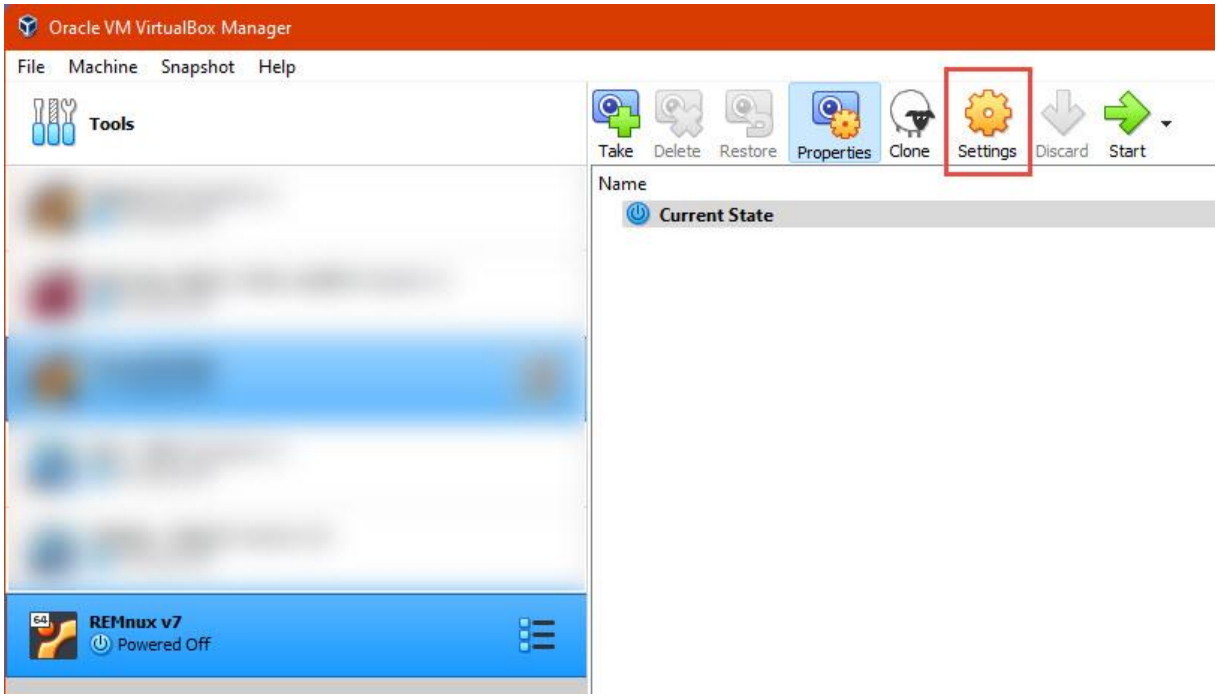
fenêtre séparée. Il s'agit d'une machine Linux, et pour se connecter l'utilisateur est *remnux* et le mot de passe *malware* (cependant, il est possible que la session s'ouvre sans demander les informations d'identification).



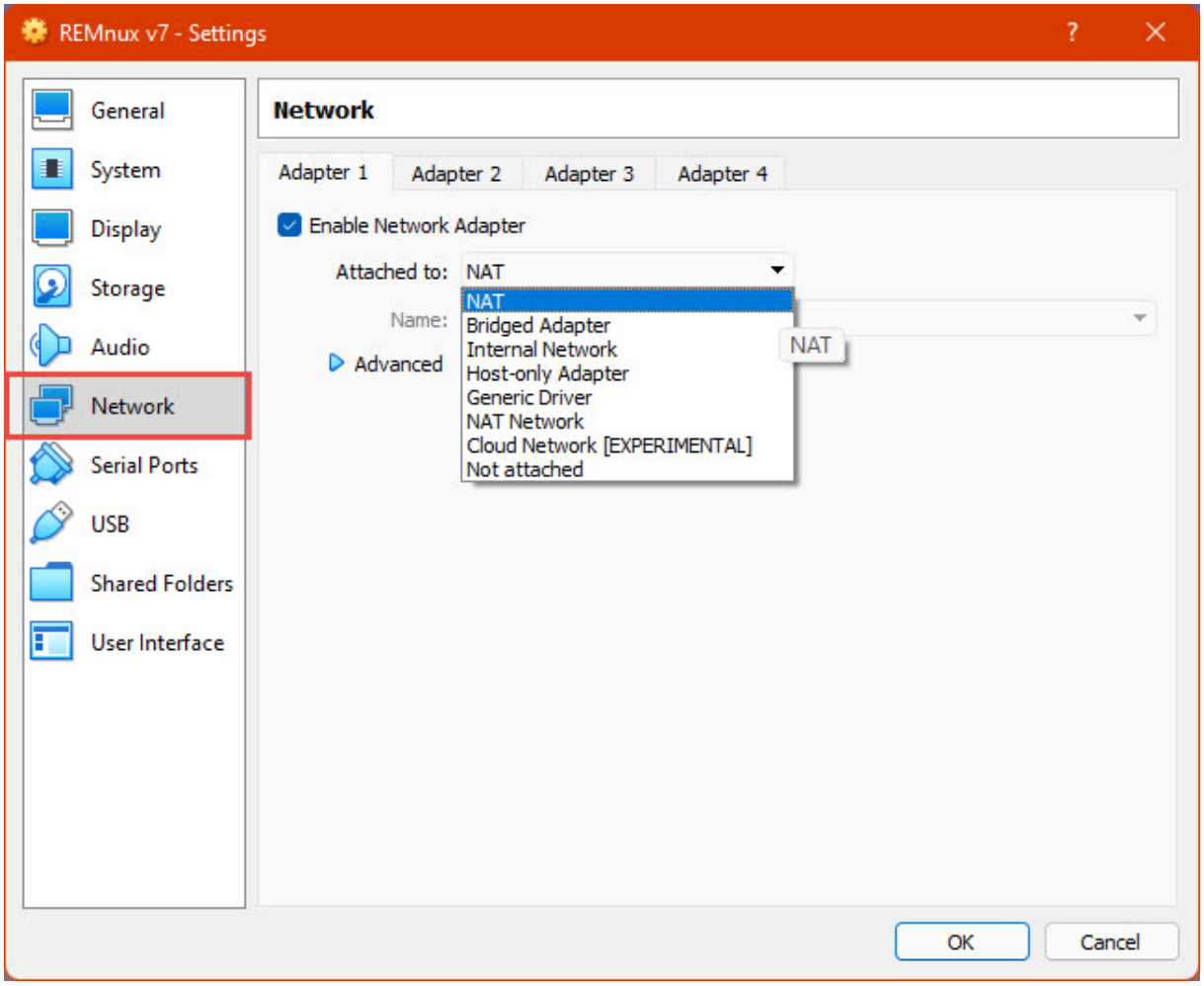


Configurations supplémentaires sur Virtualbox : réseau

Étant donné que nous allons analyser des fichiers potentiellement dangereux, il n'est pas conseillé d'exécuter la machine d'une manière qui puisse communiquer avec le reste de notre réseau. Bien que la stratégie spécifique puisse varier en fonction du style de l'analyste, la configuration a surtout lieu dans l'écran des interfaces de notre VM. Avec notre machine Remnux éteinte, nous cliquons sur le bouton « Paramètres » dans la barre d'outils.



Ensuite, dans la section « Réseau », vous verrez une série d'options, les plus importantes sont :

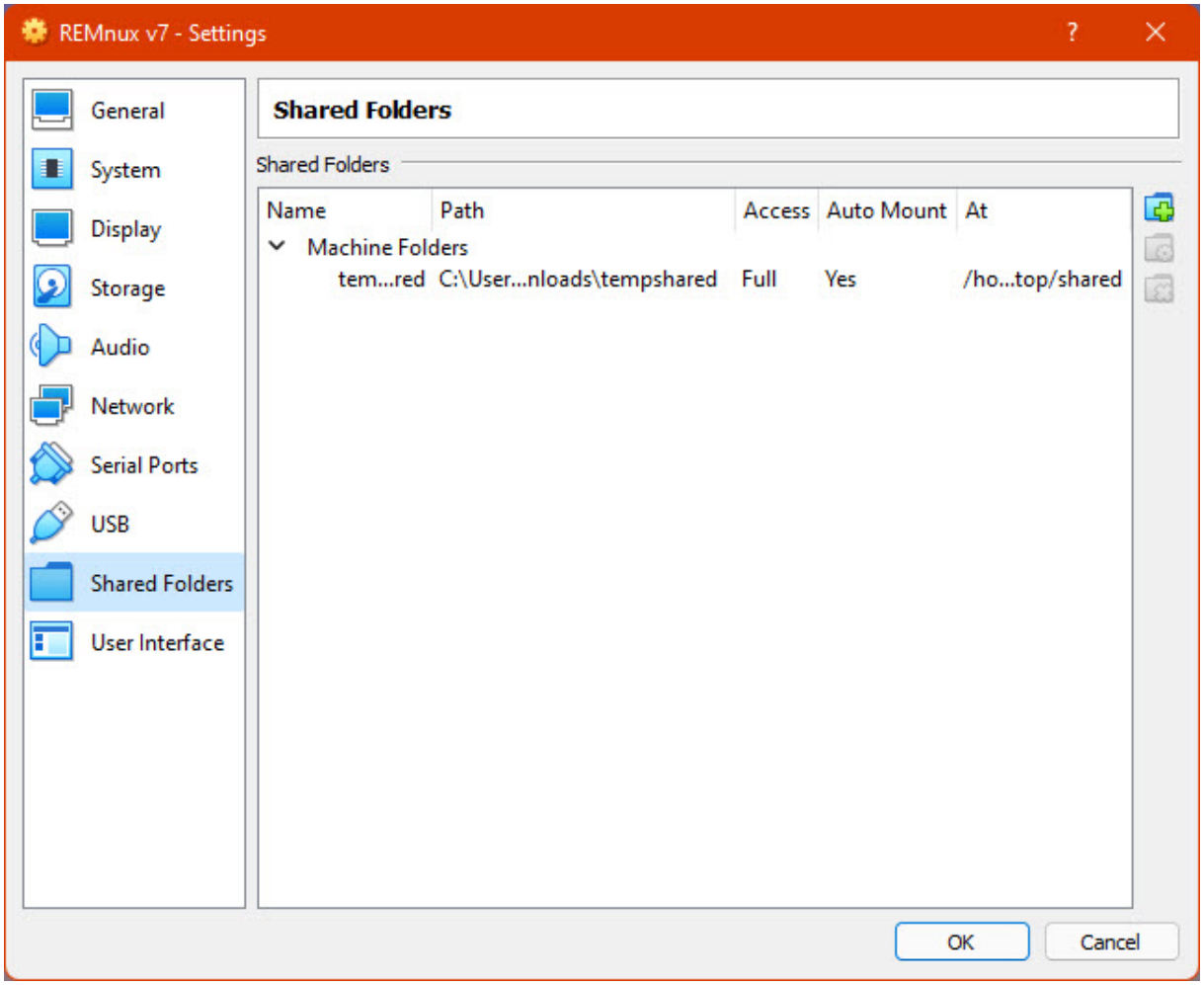


- **Activer l'interface réseau :** la désactivation de cette option éliminera toute connectivité entre notre VM et d'autres périphériques via le réseau (y compris le nôtre, nous la gérerons en utilisant l'interface graphique). Elle émulera l'absence de matériel permettant la connexion aux autres réseaux dans la VM.
- **Mode d'accès réseau - NAT :** la configuration par défaut émulera un nouveau réseau pour la machine virtuelle, cela lui permettra d'accéder à Internet, mais aussi à d'autres périphériques de notre réseau. Cette option n'est pas recommandée pour le type d'utilisation que nous allons donner à notre machine virtuelle.
- **Mode d'accès réseau - Accès par pont :** cela permettra de partager l'adaptateur réseau de notre ordinateur hôte physique à la VM en le considérant comme un autre appareil sur notre réseau. Cette option n'est pas non plus recommandée pour notre cas d'utilisation.
- **Mode d'accès réseau - Accès privé hôte :** cette option attache la VM à un réseau qui n'est connecté qu'à notre machine hôte et aux autres VM ayant la même configuration. Bien que cela puisse être utile dans certains cas, cela peut également exposer notre machine à une activité malveillante.
- **Mode d'accès réseau - Réseau interne :** similaire à l'option précédente, mais notre machine hôte ne sera pas accessible. Cela est utile lorsque nous voulons voir comment deux ou plusieurs machines interagissent entre elles.
- **Mode d'accès réseau - Aucun accès :** cette option émule un adaptateur réseau sans câble connecté.

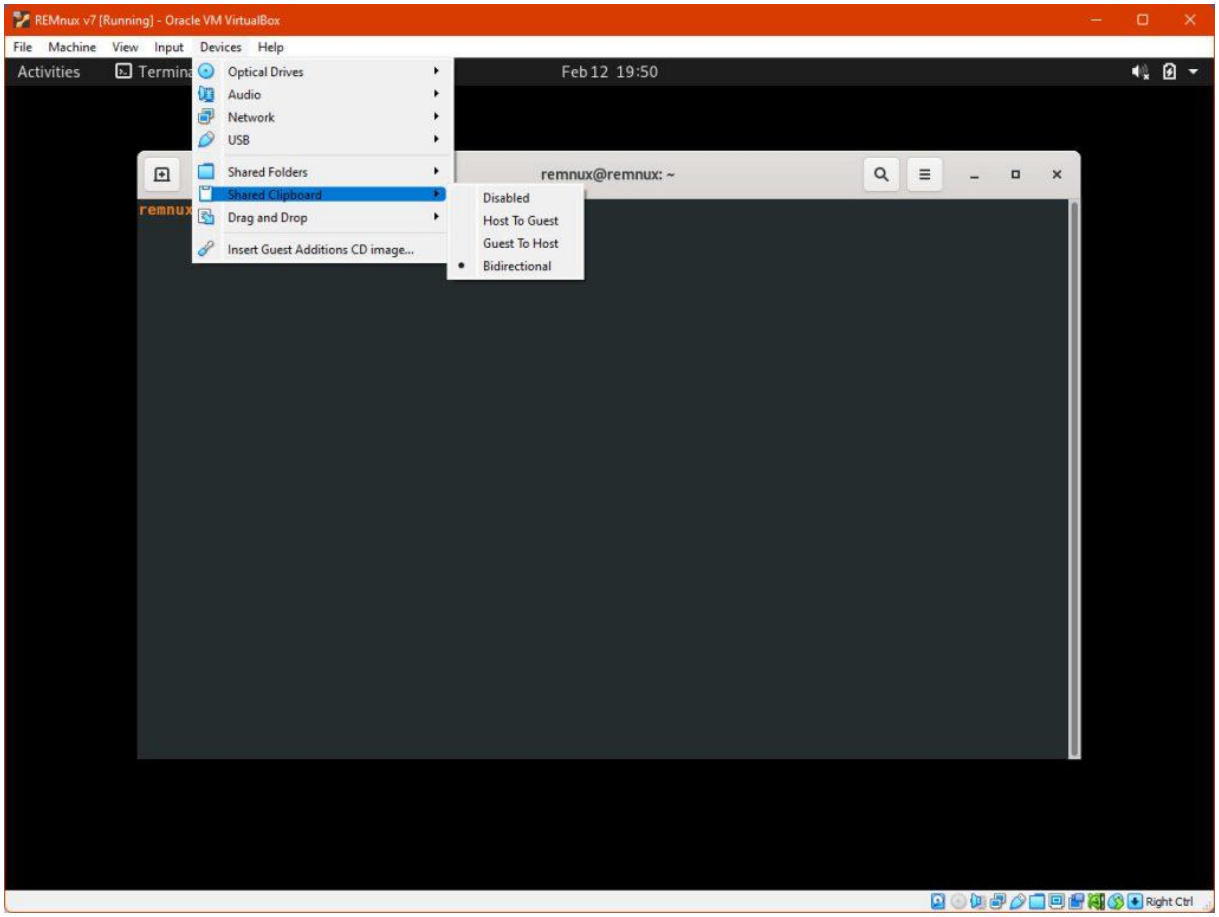
Selon l'utilisation que nous allons donner à notre machine, nous pouvons garder NAT activé dans la configuration initiale pour lui permettre d'accéder à Internet afin de télécharger des outils, etc. Nous pouvons ensuite modifier la configuration sur Sans accès, Réseau interne ou désactiver l'adaptateur avant de commencer notre analyse.

Configurations supplémentaires sur Virtualbox : partage d'informations avec la machine hôte

Il est très courant de partager des fichiers et d'autres données entre notre ordinateur et la machine virtuelle. Encore une fois, il existe différentes approches que nous pouvons adopter : **Dossiers partagés :** cela s'apparente à un dossier partagé sur le réseau, nous pouvons synchroniser un dossier entre notre hôte et notre système invité (la machine virtuelle). Le partage d'échantillons de logiciels malveillants n'est pas toujours recommandé, car cela ouvrira un espace sur notre ordinateur qui sera contrôlé par notre VM et qui pourra être infecté au cours de notre analyse. Pour configurer les dossiers partagés, une section dédiée est disponible dans les paramètres.



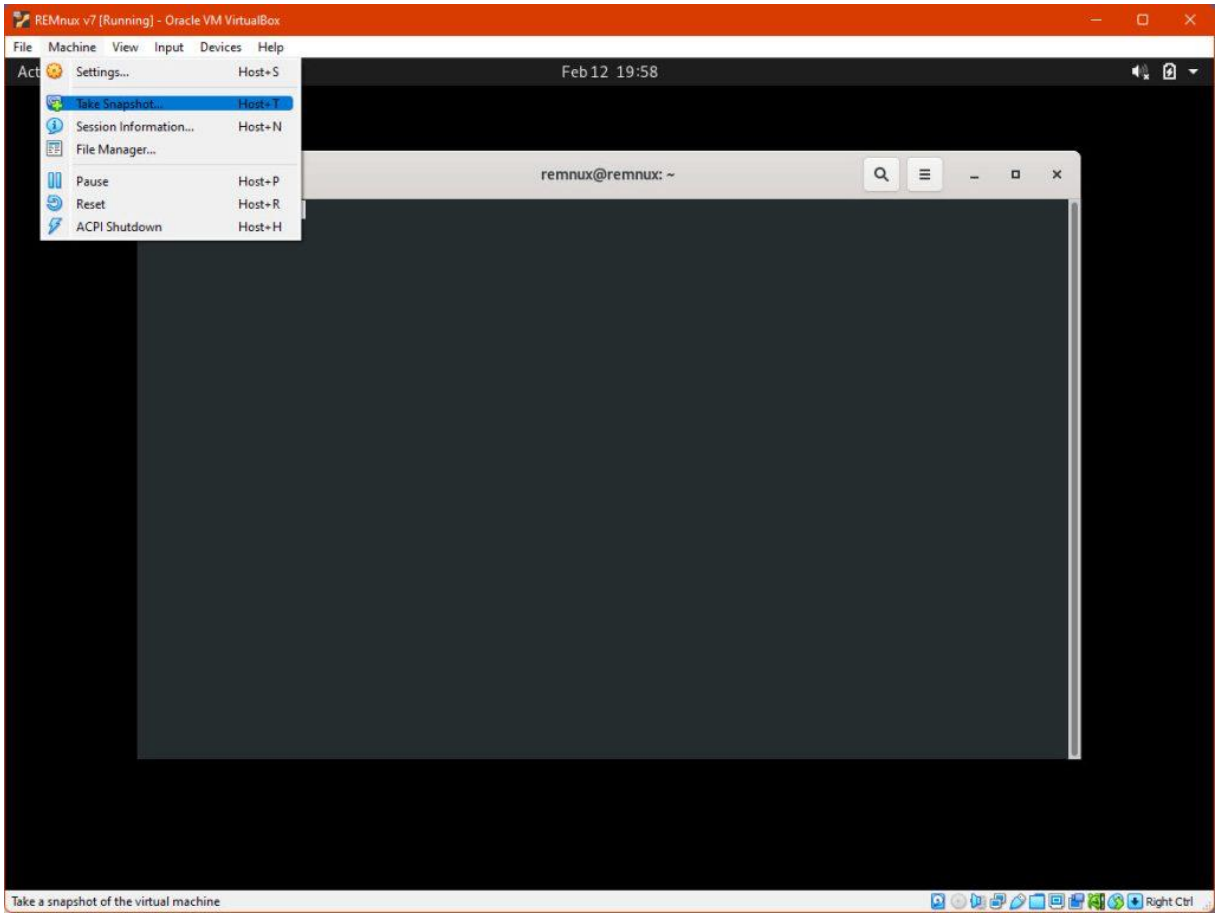
- **Presse-papiers partagé et glisser-déposer :** cela nous permettra de partager le presse-papiers entre notre ordinateur et la VM. Cette option peut être désactivée, de façon unidirectionnelle ou bidirectionnelle comme suggéré dans l'image. Cela s'applique également au glisser-déposer de fichiers entre l'hôte et le système invité. Pour certains, la désactivation du partage des dossiers et l'activation de la fonction glisser-déposer uniquement à partir de « Hôte vers l'invité » est l'option la plus sûre pour protéger les ordinateurs physiques, comme le partage du presse-papiers. Cependant, nous pourrions avoir besoin d'extraire des informations de la machine virtuelle à certains moments.



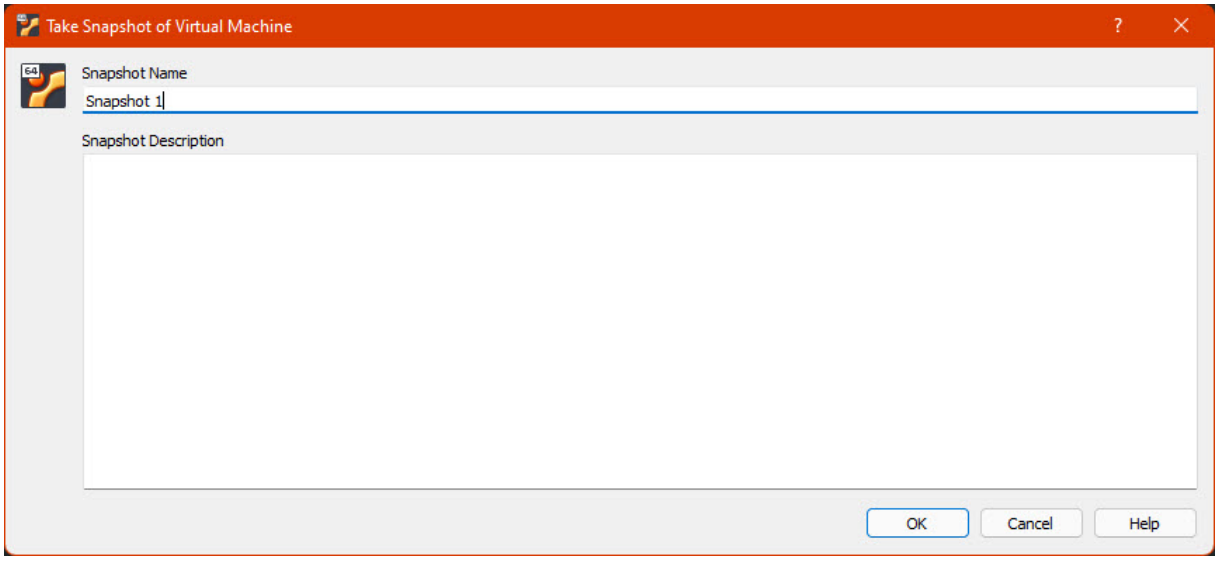
Configurations supplémentaires sur Virtualbox : instantanés

Une fonctionnalité très utile de Virtualbox est d'enregistrer une version de la VM à laquelle nous pouvons revenir à tout moment dans le futur, donc si nous configurons la machine Remnux pour analyser les logiciels malveillants, nous pourrions sauvegarder un instantané avant de commencer l'analyse. Ainsi, une fois que nous terminons, nous pouvons rétablir la VM à l'instantané enregistré pour nous assurer que la machine n'est pas infectée et prêt pour poursuivre l'analyse.

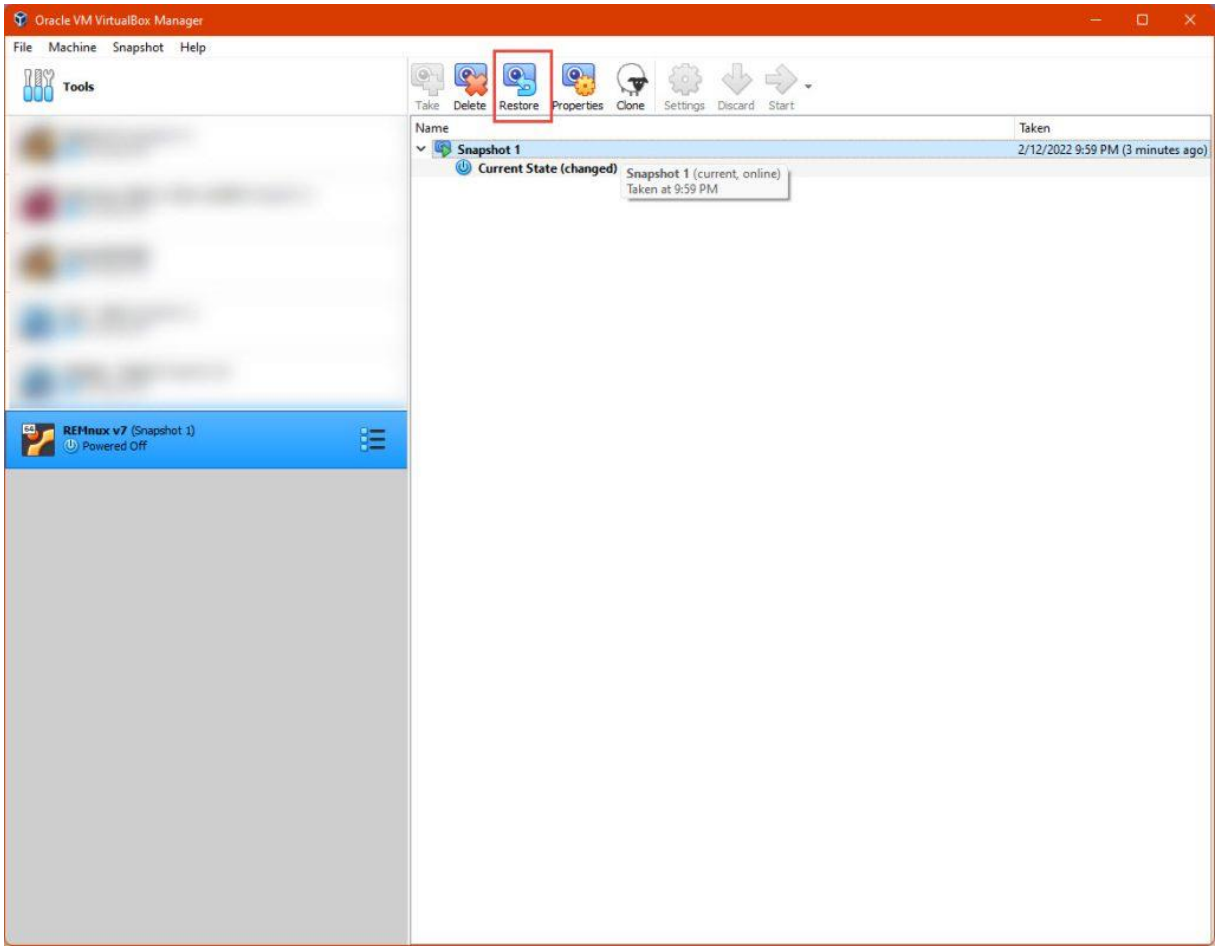
Pour enregistrer un instantané, avec la machine dans l'état souhaité, cliquez sur « Machine », puis sur « Prendre un instantané ».



Sélectionnez ensuite un nom et cliquez sur « OK ». La construction de l'instantané peut prendre du temps. Une fois l'opération terminée, il sera disponible dans la section « Instantanés » de l'écran principal de Virtualbox pour notre VM.



Nous pouvons utiliser le bouton « Restaurer » sur l'écran respectif.



Et ensuite ?

Maintenant que nous gérons les bases de Virtualbox, nous pouvons nous familiariser avec Remnux tout en comprenant et en analysant notre premier format de fichier : les PDF.