Rapport d'analyse

Filpe Fortuanto, Doriane Kaffo, Pierre Kohler & Jonathan Zaehringer

November 4, 2019

Présentation SpiderOak

SpiderOak est une société proposant diverses solutions dans le partage de données et sauvegarde avec un fort accent sur la sécurité. Dans le cadre de cette recherche, seul le produit SpiderOak One Backup est intéressant au point de vue du backup.

Ce service propose de la synchronisation, du partage et du backup de documents personnels sur le `cloud`. L'argument principal du service provient du chiffrement point à point avec un stockage des données chiffrées sur leur serveur.

SpiderOak One Backup

Introduction

SpiderOak est initialement un système de backup sorti sous sa première forme en 2007. Une entreprise a émergé de ce produit qui a spécialisé son offre en différents services, tous basés sur la philosophie du `no knowledge`. Cela consiste à l'impossibilité de lire les informations utiles (données, message, etc.) d'un client par l'entreprise hébergeant le contenu. Ainsi, SpiderOak s'engage à garantir que le contenu ne peut être lu que par le client (chiffrement point à point).

Cependant, ces informations ne peuvent difficilement être prouvées, car SpiderOak ne fournit pas les sources de leur client/serveur. Une certaine volonté de fournir le code source de leur client SpiderOak One pour plus de transparence est présente. Malheureusement, il ne souhaite pas libérer leur code provenant d'une phase de startup. Cependant, l'article datant de deux ans, aucune ouverture du code n'est visible depuis sur le github. Hormis le client mobile qui est lui totalement open source. Leur organisation n'est que très peu active et compte que deux membres. Article référence de ces objectifs

Compatibilité

Le client SpiderOak One Backup est disponible sur les trois OS principaux (Windows, Mac et Linux). Il consiste en une interface simplifiée au possible proposant le minimum d'information nécessaire à son utilisation.

Il existe aussi un client mobile disponible pour iOS et Android qui permet essentiellement de visualiser les données et de les partager.

Fonctionnement

Support de fichier

SpiderOak One Backup est conçu pour faire une sauvegarde/synchronisation avec des fichiers personnels. Il n'est pas conçu pour faire un backup au niveau du système d'exploitation. Il supporte l'ensemble des fichiers types classiques, il ne gère pas les fichiers spécifiques à Linux.

Mise en place

La mise en place de SpiderOak One Backup, à l'instar de sa prise en main, se veut simple et rapide. Nous détaillerons donc ici l'installation et la configuration sur un système Debian (GNU/Linux), mais cette procédure sera très similaire sur n'importe quel système supporté.

Installation

En se rendant sur la page de téléchargement du site officiel de SpiderOak, on peut observer les différents systèmes supportés, à savoir Mac OS, Windows ainsi que Linux (DEB, RMP ou tarball).

Il suffit donc de choisir l'installateur approprié, de l'exécuter et de suivre la procédure d'installation, qui ne vous demandera aucun détail particulier.

Configuration

Connexion

Lors du premier lancement du logiciel SpiderOakONE, il vous sera demandé de vous identifier. Il est donc nécessaire de créer un compte afin de pouvoir bénéficier des services. Sachez qu'à la création d'un nouveau compte, il vous sera possible de tester gratuitement la solution pour 21 jours.

Prise en main

Onglet Dashboard Une fois connecté, le logiciel présentera en premier lieu une *Dashboard* exposant les activités en cours. Plus précisément, vous pourrez observer les transactions en cours (téléchargements / téléversements) relatifs aux trois solutions du logiciel (sauvegarde, synchronisation et partage).

Sur la gauche se trouve également un rappel de tous les appareils rattachés à l'utilisation de ce compte.

Onglet Sauvegarder Cet onglet permet la sélection des fichiers et dossiers que l'on souhaite sauvegarder (*backup*) sur la machine courante. Si la navigation à travers l'arborescence complète du système est possible, il est également proposé de sélectionner diverses catégories préétablies correspondant à des dossiers classiques des systèmes (p. ex. photos, film, musique, etc.).

Onglet Gérer Ce dernier contient l'ensemble des données présentent sur les serveurs distants de SpiderOak. Regroupées par machine, les différentes sauvegardes pourront être naviguées, téléchargées ou encore supprimées, et ce dans n'importe quelle version ayant existée.

Il vous sera également rappelé si un dossier fait également l'objet d'une synchronisation (dossier de couleur verte) ou d'un partage (couleur violette).

Vous pourrez également obtenir un lien public de téléchargement d'un **fichier**.

Onglet Synchroniser L'onglet de synchronisation permet simplement d'ajouter des règles de synchronisation entre différents dossiers préalablement sauvegardés (à l'aide le l'onglet *Sauvegarder*). Il vous sera donc possible de conserver un état identique entre ces dossiers, et ce même sur des machines distantes.

Onglet Partager Enfin, cet onglet permet la mise en place d'un partage sécurisé.

À la création d'un nouveau partage vous seront demandés un nom explicite et un identifiant unique pour le partage, un mot de passe et une description.

Finalement, après sélection du dossier à partager et finalisation du processus, une salle de partage sera créée et accessible via un lien précis ainsi que le mot de passe préalablement défini.

Présentation Tarsnap

Tarsnap est un service de sauvegarde en ligne sécurisé et efficace créé en 2008 par Colin Percival. Il est un fournisseur de sauvegarde hors site. Il nous donne la possibilité de créer notre propre sauvegarde. L'un de ses principaux avantages est que les sauvegardes effectuées sont cryptées avec une clé secrète accessible uniquement à vous. Cela signifie que ni Tarsnap, ni les personnes n'ayant accès à ses serveurs, ne peuvent déchiffrer vos sauvegardes à moins de disposer de votre clé secrète. Alors que les sauvegardes elles-mêmes garantissent que vous ne perdez pas les données importantes de votre serveur, le cryptage vous évite de vous inquiéter de la possibilité pour un attaquant de récupérer vos sauvegardes et d'accéder aux données sensibles qu'ils contiennent.

Champs d'utilisation

Tarsnap, de par l'absence d'interface graphique, se voit immédiatement orientée pour les entreprises, et non pour les particuliers.

L'utilisation principale de Tarsnap serait de faire des sauvegardes à intervalle régulière, et fréquente (journalières ou hebdomadaires), de services et données critiques à une entreprise.

En effet, de par le modèle des coûts de celui-ci, la fréquence de sauvegarde n'affecte pas fortement le coût mensuel de Tarsnap.

Il est également bon de noter que Tarsnap est adéquat pour la sauvegarde de données sensibles (p. ex. les données bancaires), de par leur politique de stockage "Zero-Knowledge".

Tarsnap open source?

Le code source du client Tarsnap est disponible afin que l'utilisateur puisse voir exactement comment les données sont protégées. Cependant, il n'est pas distribué sous une licence Open Source. Certains des composants réutilisables ont été publiés séparément sous une licence BSD tels que :

- La fonction de dérivation du script key : conçue pour être beaucoup plus sécurisée contre les attaques par force brute que par d'autres fonctions telles que PBKDF2 ou bcrypt .
- La kivaloo data store : La kivaloo est un ensemble d'utilitaires qui forment ensemble un magasin de données associant des clés allant jusqu'à 255 octets avec des valeurs allant jusqu'à 255 octets.
- Le spiped secure pipe daemon : un utilitaire permettant de créer des canaux cryptés et authentifiés de manière symétrique entre les adresses de socket, de manière à ce que chacun puisse se connecter à une adresse (par exemple, un socket UNIX sur localhost) et établir de manière transparente une connexion à une autre.

Contrairement à de nombreux systèmes de sauvegarde, Tarsnap utilise un concept d' archives et ne nécessite pas des connaissances en gestion de sauvegarde. Son outil d'archivage est le tar. L'archive Tarsnap créée ne peut plus être modifiée, par conséquent ne peut être écrasée. Chacune d'elles doit être créée avec un nouveau nom.

Les exigences de tarsnap

Tarsnap n'a pas d'interface utilisateur graphique. Il n'y a que des binaires pré-construits officiels pour quelques systèmes d'exploitation. Il est supporté par les systèmes d'exploitation suivants :

- Linux
- BSD
- MacOS: utiliser un gestionnaire de paquets.
- Windows: il est uniquement pris en charge via Cygwin et le sous-système Windows pour Linux.
- La plateforme indépendante de tarsnap: en utilisant son code source.

Il est à noter que Tarsnap dispose d'un code source de chacun de ces systèmes d'exploitation dont il faut compiler pour l'installer.

Infrastructure

Le service Tarsnap est construit sur la plate-forme solide fournie par Amazon Web Services. Une fois que les données ont atteint le serveur Tarsnap, elles sont stockées dans Amazon S3 . et le serveur Tarsnap n'accuse réception de la demande du client qu'une fois que les données ont été stockées sur des disques dans plusieurs centres de données géographiquement divers. Cependant, S3 en lui-même ne fournit ni les garanties de cohérence ni la journalisation requise par Tarsnap. Pour les fournir, le serveur Tarsnap implémente un système de fichiers structuré en journal sur S3, mais conserve toutes les métadonnées pertinentes sur un EC2.

Conserver des (méta) données sur une instance EC2 signifie que nous devons accepter la possibilité que l'instance EC2 meure; cependant, l'utilisation d'un système de fichiers structuré en journal facilite la résolution du problème : toutes les métadonnées sont implicites dans les entrées de journal individuelles. Par conséquent, pour régénérer les métadonnées, il suffit de relire les entrées de journal à partir de S3. En fait, cela fournit un très bon "filet de sécurité" pour se protéger contre tout problème imprévu dans le service Tarsnap : si tout échoue, je peux effectuer un redémarrage complet du service en jetant tout sauf les données stockées sur S3, puis en reconstruisant tout l'état transitoire.

La sauvegarde

Lors de la création d'archives, Tarsnap prend des flux de données d'archives et les divise en blocs de longueur variable. Ces blocs sont comparés et tous les blocs en double sont supprimés avant d'être téléchargés sur le serveur Tarsnap. Tarsnap conserve un cache local lui indiquant les blocs précédemment stockés et l'utilise lors de la création d'archives supplémentaires. Ainsi, stocker deux archives contenant les mêmes données ne prend que très légèrement plus d'espace (par conséquent une petite quantité de temps système par archive) que de stocker une archive unique. Si les fichiers changent entre les archives, seules les modifications devront être téléchargées lors de la création de la prochaine archive. La taille d'une sauvegarde n'est pas fixée.

Tarsnap utilise le même ou moins de stockage qu'un système traditionnel de sauvegarde complète et incrémentielle, en offrant la possibilité de créer et supprimer les archives indépendamment les uns des autres. Du coup, il utilise beaucoup moins de bande passante et de stockage que les sauvegardes incrémentielles, car Tarsnap évite de stocker plusieurs copies des segments non modifiés de fichiers. Il conserve un cache contenant les chemins, les numéros d'inode, les tailles et les heures de modification des fichiers. Au cas où les fichiers n'auraient pas été modifiés, il évitera de passer du temps à les lire à partir du disque.

De par le fonctionnement des sauvegardes, tarsnap ne permet au client que de faire des backup "complets", et non pas en incrémentaux, bien que ne stockant que les données uniques de ces nouvelles sauvegardes sur ses serveurs.

En vue de ce système de sauvegardes distantes, il est également envisageable de faire des sauvegardes aussi souvent que nécessaire, afin de satisfaire les nécessitées en termes de RPO, sans que les coûts de stockage explosent dramatiquement.

Finalement, dû à l'utilisation sous-jacente de tar par Tarsnap, il est relativement aisé de récupérer un seul fichier depuis une sauvegarde, il suffit d'exécuter la commande suivante : `tarsnap -x -f <backup_name> <path_of_file_to_restore>`

La sécurité de tarsnap

Elle est construite autour de la notion de fichier de clé requis par le code du client Tarsnap pour pouvoir effectuer toute opération. Cela permet à Tarsnap d'éviter plusieurs faiblesses liées à l'utilisation de mots de passe comme l'utilisation des mots de passe pas assez fort, la saisie du mot de passe par l'utilisateur peut être observée, l'impossibilité d'utiliser les mots de passe pour déterminer les clés asymétriques. Par contre on utilise toujours les phrases secrètes par exemple si le fichier de clé est volé.

Le modèle de sécurité de tarsnap est celui du "Zero Knowledge"

Les fichiers de clés Tarsnap contiennent deux types de clés : les clés d'**authentification**, utilisées pour prouver au serveur que le titulaire est autorisé à écrire, lire ou supprimer des données, ainsi que de signer et vérifier l'intégrité de celles-ci ; et des clés de **chiffrement**, qui sont utilisées pour chiffrer et déchiffrer des archives. Cette séparation garantit que même si le service Tarsnap est compromis, les données seront protégées de toute divulgation et altération non autorisée.

Lorsqu'un système est enregistré auprès du service Tarsnap via la tarsnap-keygen utilitaire, toutes ces clés sont générées et stockées dans un fichier de clés unique. Mais à l'aide de l' tarsnap-keymgmt utilitaire, il est possible de créer des fichiers de clé "restreints" pouvant être utilisés pour créer n'importe quel sous-ensemble de: lecture, vérification et décryptage des archives. Cryptage, signature et écriture d'archives. Lecture et suppression d'archives individuelles. (il est impossible de supprimer une archive tout en laissant les autres intactes sans pouvoir également lire les archives.). Supprimer toutes les archives stockées par un système. Cela permet par exemple de configurer un système pour stocker automatiquement les archives quotidiennement sans pouvoir les lire ou les supprimer, ce qui garantit que même si quelqu'un pénètre dans le système et le fait préférable de tout supprimer, vos données seront toujours en sécurité (par exemple, dans le cas d'un ransomware avancé).

Installation de Tarsnap

Sur des machines de type Debian (Ubuntu ou Mint, par exemple), le site de Tarsnap explicite l'ajout de la clé de package de

Tarsnap au système, l'ajout de leur repos à la liste de sources APT, et l'installation de Tarsnap lui-même via apt-get.

Leur site web fournit également des instructions pour compiler et installer Tarsnap via le code source, afin de pouvoir l'installer sur d'autres systèmes non supportés.

Il faudra ensuite créer un compte sur le site de Tarsnap, et déposer au minimum 5$ USD sur son compte afin de pouvoir l'utiliser.

Configuration de Tarsnap

Tarsnap est configurable via un fichier de configuration, localisé à 2 endroits possibles :

- `/etc/tarsnap.conf`, si Tarsnap a été installé via leur package .deb
- `/usr/local/etc/tarsnap.conf`, si Tarsnap a été installé via le code source.

Le fichier de configuration par défaut est suffisant pour une utilisation traditionnelle.

Il faudra ensuite créer une clé d'authentification Tarsnap, authentifiant la machine auprès des serveurs de Tarsnap, à l'aide de la commande suivante, en rentrant son mot de passe lorsqu'il est demandé.

```
sudo tarsnap-keygen \
    --keyfile /root/tarsnap.key \
    --user <account_email> \
    --machine <machine_name>
```

Attention ! Le keyfile d'authentification généré est également utilisé pour le chiffrement des données sauvegardées, si celui-ci est perdu, les données seront définitivement perdues.

Il est recommandé de sauvegarder ce keyfile en 3 places, sur plusieurs supports et sur plusieurs lieux physiques.

Utilisation de Tarsnap

Utilisation basique

L'utilisation recommandée de Tarsnap pour créer des backups est via des scripts automatisés via cron, mais dans le cadre de ce guide d'utilisation l'automatisation via cron ne sera pas couverte.

La commande pour backup le dossier `/tmp/toto` peut est comme suie :

```
tarsnap -c -f "$(uname -n)\
    -$(date +%Y-%m-%d_%H-%M-%S)" /tmp/toto
```

- le paramètre <-c> spécifie la création d'une archive distante
- le paramètre <-f> spécifie le nom de l'archive distante. Dans ce cas, le nom distant sera le nom de la machine, suivi de la date actuelle, précise à la seconde.

NB : chaque keyfile dispose d'un namespace, et par conséquent, de dossier racine diffère, donc il est possible de créer plusieurs archives de même nom uniquement si l'on dispose de différents keyfiles.

- Il est ensuite possible de lister tous les backups distants à l'aide de la commande :

```
tarsnap --list-archives
```

- Il est possible de restaurer un backup via :

```
tarsnap -x -f <backup_name>
```

- Il est possible de supprimer un backup avec :

```
tarsnap -d -f <backup_name>
```

Utilisation avancée

- Il est possible de lister les fichiers contenus dans un backup avec :

```
tarsnap -tv -f <backup_name>
```

- Il est possible de restaurer un fichier particulier à l'aide :

```
tarsnap -x -f <backup_name> <path_of_file_to_restore>
```

- Il est possible de tester, sans rien upload, combien de données seraient utilisées (et donc, combien d'argent sera dépensé) par le backup d'un fichier/dossier :

```
tarsnap -c -f <backup_name> --dry-run --print-stats \
    <path_of_file_or_folder>
```

- Suite à une perte du dossier de cache, ou une migration système, il est possible de le restaurer via :

```
tarsnap --fsck
```