Threat Grid Appliance Backup Overview and FAQ

Version: 2.2.4

Created: 7/10/2017

Cisco Systems, Inc. www.cisco.com

All contents are Copyright © 2015-2017 Cisco Systems, Inc. and/or its affiliates. All rights reserved.

Contents

Introduction

The Threat Grid appliance release 2.2.4 introduces a backup feature: Appliances now support backups to NFS-backed storage (NFSv4); initialization of data from the backup storage; reset of the appliance to an empty database state into which a backup can be loaded; and database restore functionality.

This document provides important backup information, beginning with a technical overview followed by Frequently Asked Questions with detailed information and step-by-step instructions.

Please read through this document carefully for important information before working any tasks, and contact Threat Grid Support (support@threatgrid.com) if you have any additional questions.

Technical Overview

NFS Requirements

- NFSv4 server (**NOT** v3!), with a shared directory configured to allow write access for nfsnobody (UID 65534).
- The NFSv4 server must be accessible via the Admin 10Gb interface.
- Storage. Lots of it. See details in the FAQ: *How much backup storage is required?*

Expectations

- **Included in the Backup** - This release of the Threat Grid appliance backup process covers customer-owned bulk data:
    - Samples
    - Analysis results
    - Databases (including users and organizations)
    - Configuration done within the Face or Mask portal UI
- **Not Included** - This release DOES NOT include configuration done inside the appliance OpAdmin interface.
- **PostgreSQL** - PostgreSQL base backup takes place on a 24-hour cycle. Database backup cannot be restored, and a warning will be displayed, until this has successfully completed at least once.
- **ElasticSearch** - ElasticSearch backup takes place incrementally, once every 5 minutes.
- **Freezer** - Freezer backup takes place on an ongoing basis, with a job following behind every 24 hours to handle any objects which were missed from the ongoing backup.
- **New Key Generation** - Generating a new key creates a new, independent backup store. Like the original, this new store is not valid until a base backup has taken place on a 24-hour cycle.

Backup Process

- NFS configuration is completed in the OpAdmin interface via a new step added to the setup wizard. A new menu entry is added for later access to NFS configuration.
- New encryption key generation and download, or upload.

IMPORTANT NOTE: the customer is responsible for backing up the encryption key and storing it securely! Backup is useless without this key!

Reset Process

CAUTION! Leveraging this process will destroy customer-owned data! Be very careful, and very certain! Read through the FAQ information before working any tasks.

- Before an appliance can be used as a restore target, it must be in a preconfigured state. Appliances ship in this state. However, getting one back to the preconfigured state once it has been configured requires explicit administrative action. This task is described in Step 1 of the FAQ entry, What is the process for restoring backed-up content?
- Note that reset is not the same as the secure wipe available in recovery mode; only the recovery-mode secure wipe is appropriate to completely remove customer-owned data from a machine before shipping it to a DLP reimaging center. The secure wipe in recovery mode is NOT a replacement for this reset: secure wipe renders an appliance unusable until reimaged, and reset prepares an appliance to restore a backup.

Restore Process

- Can only restore from the setup wizard.
- Set up the same NFS store as used previously, and the same encryption key as used previously, with a process identical to the original.
- The act of setting up an appliance with a prior NFS store and encryption key will trigger a restore.
- **IMPORTANT NOTE:** Only one server can be running with data from a given backup store active at a time!
- To test the restore process on a different Threat Grid appliance while your primary appliance is still operational, make a copy of a consistent snapshot of the backup store, and point a new appliance (with the encryption key uploaded) at that copy.

Frequently Asked Questions

What are the requirements for the NFS server's configuration?

- Must be running the NFSv4 protocol over TCP, accessible from the appliance's admin interface.
- Configured directory, must be writable by nfsnobody (UID 65534).
- The NFSv4 server must be accessible via the Admin 10Gb interface.
- Sufficient storage. See the question below, *How much backup storage is required?* for details.

The following mount parameters are unconditionally used: `rw,sync,nfsvers=4,nofail`

What are the failure modes for bad NFS configuration?

Invalid NFS configuration (or configuration pointing the service at an incorrectly-configured NFS server) will generally cause the process of applying configuration to fail. Correcting this configuration in OpAdmin and reapplying should result in success.

Is exposing files for write by `nfsnobody` secure?

The only processes running as `nfsnobody` or with write to `nfsnobody`, on the ThreatGRID appliance are those responsible for encryption of data. Plaintext data is exposed under distinct user accounts for different subtrees according to principal of least privilege; the PostgreSQL service on the appliance cannot access ElasticSearch data or the freezer; the ElasticSearch service cannot access PostgreSQL or freezer data; etc.

Using the `nfsnobody` account simplifies configuration, preventing the need to build an `idmap.conf` for each customer's site mapping local and remote account names together.

How much backup storage is required?

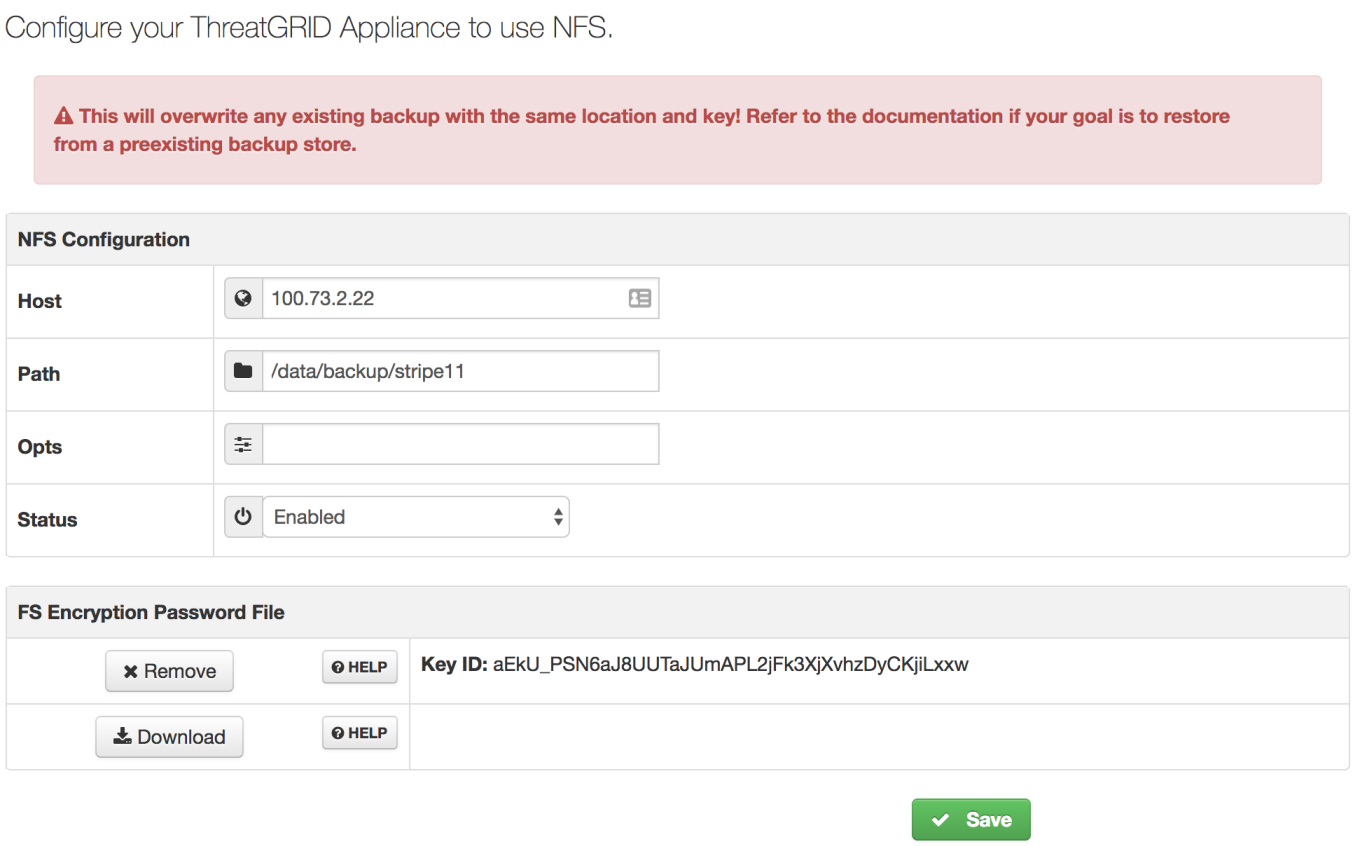
A backup store consists of the following components:

- **The Object Store.** In practice this will generally be the bulk of the storage in use. Data retention for the bulk component of a backup store follows the same policies and limits documented for the appliance release in use – for 2.2.x-series appliances, the document at http://www.cisco.com/c/dam/en/us/td/docs/security/amp_threatgrid/amp-threat-grid-appliance-data-retention-v2–2.pdf is applicable, and places maximum storage use for this component as 4.1TB.
- **The PostgreSQL database store.** This contains two full backups of the PostgreSQL store, and a chain of WAL logs sufficient to allow replay from the oldest of the retained full backups. This should be less than 500GB in total.
- **The ElasticSearch snapshot store.** This should be less than 1TB in total.
- **Total Storage.** Thus, given the above, a backup store should not require more than **5.6TB.**

What is the backup process?

1. Create the target directory according to the requirements noted above.
2. Complete the NFS Configuration page of the setup wizard in OpAdmin:

Fig. 1 - NFS Configuration Page



- **Host** - The address of the NFSv4 server for storing the appliance's backup data
- **Path** - The path to the Host server backup directory
- **Opts** - NFS mount options
- **Status** - Select *Enabled* from the dropdown (Pending Key).

3. Click **Save**. The page will refresh, with a FS Encryption Password Key ID now available.

   The first time you configure this page, options to **Remove** or to **Download** the encryption key become visible. **Upload** is available if you have NFS enabled but no key created. Once you create a key, **Upload** is changed to a **Download** button. (If you delete the key, the **Download** button becomes **Upload** again.)

   NOTE: If the key correctly matches the one used to create a backup, the *Key ID* displayed in OpAdmin after upload will match the name of a directory in the configured path. As already noted, backups cannot be restored without the encryption key.

4. Finish the remainder of the setup wizard as usual.

   The configuration process includes the process of mounting the NFS store, mounting the encrypted data, and initializing the appliances' local datastores from the NFS store's contents.

How frequently are backups run?

For bulk storage of samples, artifacts and reports, content is backed up continuously. Additionally, a pass is performed to look for and transfer missing content on a 24-hour cycle.

For the PostgreSQL database, a base backup is created on a 24-hour cycle, and incremental content is continually added thereafter - either as soon as a 16MB threshold of newly-written database content is reached, or not less than once every 5 minutes.

For the ElasticSearch database, content is incrementally added to the backup store on a 5-minute cycle.

Can backup frequency be controlled or tuned?

No.

As tuning these values would make estimates regarding storage usage, restore-process time, and performance overhead invalid, they are not presently tunable.

How long is backed-up data retained?

- For PostgreSQL, the last two successful backups and all WAL segments since those backups are retained.
- For ElasticSearch, the last two 5-minute snapshots are retained.
- For bulk storage, the same retention policy followed and documented for a single appliance is used for the shared store.

How is backup data encrypted?

Content is encrypted with gocryptfs, a 3rd-party open-source product.

Note that filename encryption is disabled for performance reasons. As samples and other content in Threat Grid are not stored with their original names under any circumstances, this does not leak customer-owned data.

What is the reset process?

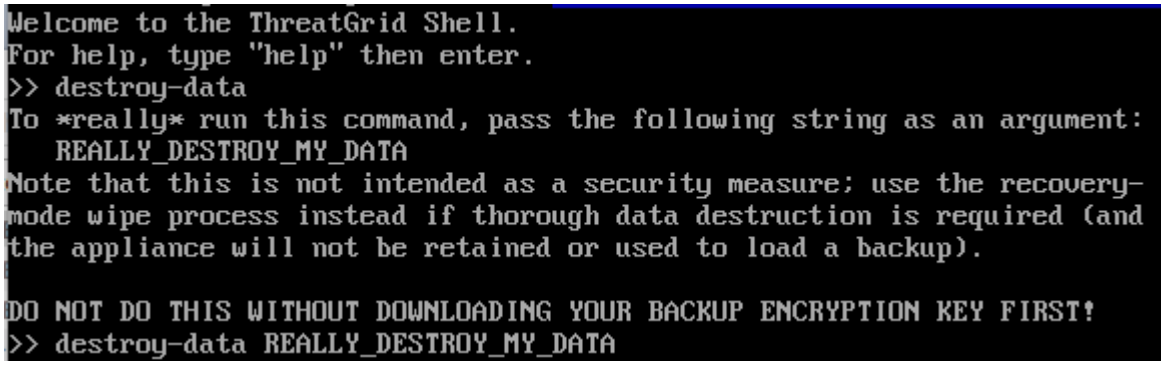
Before an appliance can be used as a restore target, it must be in a preconfigured state. It arrives in this state from manufacturing, but if the appliance has been used, you will first need to restore the system to a preconfigured state as described in Step 1:

1. **If not restoring to a system fresh from manufacturing:**

   The restore target appliance must be returned to the preconfigured state by clearing pre-existing data and NFS-related configuration from the system:

   - Access the `tgsh-dialog` configuration interface, either via the appliance's TTY or via SSH.
   - Select the `CONSOLE` option to enter tgsh. (Note that entering tgsh via recovery mode is not suitable for this use case.)
   - At the tgsh prompt, enter the command `destroy-data`. Carefully read and follow the instructions provided with the prompt.
   - CAUTION! There is NO *Undo* from this command:

Fig. 2 - The `destroy-data REALLY_DESTROY_MY_DATA` command and argument



2. **If another system is actively writing to the backup being restored:**

   (For example, if this is a test restore of content being written by a second, master appliance actively used in production.) Generate a consistent, writable copy of the datastore, and point your appliance doing the test restore at this writable copy rather than at the store which is being continuously written.

Once the appliance is in a preconfigured state, it can function as the target for the backup store as described in the answer to the next question, *What is the restore process?*, below:

What is the restore process?

Required: As already noted, backups cannot be restored without the encryption key.

Upload the Backup Encryption Key - In the NFS Configuration page of the setup wizard in OpAdmin, click **Upload** to retrieve the backup key previously generated when configuring the server on which the backup was created.

- If the key correctly matches the one used to create a backup, the Key ID displayed in OpAdmin after upload will match the name of a directory in the configured path.
- The install wizard checks for a directory matching the backup key, and if it finds one, will begin restoring the data into that location.
- The amount of time required to restore data depends on the size of the backup and other factors. In testing, a 1.2GB restore simply fly by, while a 1.2TB restore required 16+ hours. NOTE: There is no progress bar, so on lengthy restores it may appear that the install has hung; be patient.
- The restored data looks just like the original data.
- NOTE: The system is unavailable for sample submission during the restore process.

How do I know if the backup has been successfully restored?

OpAdmin will report that the install was successful and the appliance will start up.

What are the backup-related service notices that can occur?

Network storage not mounted. Check that the network filesystem being used as a backend is fully operational, and try reapplying configuration through opadmin or rebooting your appliance.

Network storage not working. Check that the network filesystem being used as a backend is fully operational; if the system does not recover within 15 minutes of correcting any problems with the NFS server, try rebooting your appliance.

Backup filesystem access failure. Contact customer support.

No PostgreSQL backup found - This is a normal condition between the point in time when a backup store has been configured and the point in time when the first base backup (run automatically on a 24-hour cycle) takes place. Note that until this is complete, a backup is not considered complete and cannot be restored. *If and only if* this message persists for more than 48 hours, contact customer support.

Newest PostgreSQL base backup more than two days old - This indicates that the system has not been successful in generating a new base backup for PostgreSQL. If unremediated, this can result in unbounded usage on the backup store (to retain a full chain of writes necessary to restore from an increasingly-old backup point), and unacceptably long processing time needed for a restore to take place. Contact customer support.

Backup Creation Messages: - These reflect errors detected when starting or triggering a backup.

ES Backup (Creation) Inactive - Indicates that when ElasticSearch was started, the backup store was unavailable. This can be remediated by rebooting the appliance, or (if NFS and the encryption service are now functional) by logging into tgsh and running the command `service restart elasticsearch.service`.

Backup Maintenance Messages: - These reflect errors detected when checking status of previously-created backups.

ES Backup (Maintenance) snapshot (…) status FAILED - This indicates that in the most recent attempt to update the backup of the ElasticSearch database, no indices could be successfully written. Check that the NFS server is functional and has free space; if no issue can be identified and the issue persists, contact customer support.

ES Backup (Maintenance) snapshot (…) status INCOMPATIBLE - Should only occur immediately after an appliance upgrade installing a new version of ElasticSearch; will be displayed until the backup store has been upgraded to be compatible with this new release. Restoring from an INCOMPATIBLE backup may require customer service assistance, should a failure occur while in this state.

ES Backup (Maintenance) snapshot (…) status PARTIAL - Contains one of two messages in the body: *No prior successful backups seen, so retaining.* (if we're keeping a partial backup as better than none at all); or *Prior successful backups exist, so removing.* (if we're discarding that partial backup with the intent to retry later).

ES Backup (Maintenance) - Backup required (…)ms - Occurs if a backup requires more than 60 seconds. This is not necessarily an error: ElasticSearch performs periodic maintenance which can cause significant write load even on idle systems. However, if it takes place consistently when under periods of low load, investigate storage performance or contact customer service for assistance.

ES Backup (Maintenance) - Unable to query ElasticSearch snapshot status - ElasticSearch could not be contacted; and this failure took place after a backup creation was successfully started. Generally, this will occur in conjunction with other appliance failures, and remediation should focus on those issues. If this error is seen when the appliance is otherwise fully functional and does not go away of its own accord, contact customer support.

How can I recover from an interrupted restore process?

If the restore process is interrupted by a reboot or other unexpected error, use the `destroy-data` command documented as part of the process of resetting an appliance to be ready to load a backup before attempting backup restore again.

Can I restore an older backup?

Old backup snapshots (and PITR records) are retained only for disaster recovery scenarios when the most recent backup is unusable. Functionality for restoring them is not customer-accessible, and scenarios where they need to be used (because the general restore process fails) should be grounds for escalation to Threat Grid appliance engineering.

If you need to retain older backups in a customer-accessible form, use a NAS with snapshot support from a vendor offering such capabilities for the NFS store. Note the guidance on cross-version compatibility given elsewhere.

Can the user select a different backup set for each service (i.e., PostgreSQL, ElasticSearch, and bulk storage)?

No.

Which data is included in backups?

All configuration and content created through the primary ThreatGRID application. This includes:

- Samples, analysis reports, artifacts, flagging, and other application-level content generated through the Threat Grid application.
- Application-layer (not OpAdmin) organization and user account data.

Which data is NOT included in backups?

- System logs
- Previously downloaded and installed updates
- Configuration entered through OpAdmin, including SSL keys and CA certificates

Which data is destroyed by the `destroy-data` command?

- All data listed above under *Which data is included in backups?*
- NFS configuration and credentials.
- The local copy of the encryption key used for NFS.

Which data *is not* destroyed by the `destroy-data` command?

All data listed under *Which data is NOT included in backups?*, except for NFS configuration and the local copy of the encryption key used for NFS (both of which are in fact destroyed).

Will the set of content included in backups be extended in the future?

While no release date for this functionality is promised or guaranteed, future releases are likely to include configuration information from OpAdmin, excluding network configuration, in backups.

Which configurations are affected by the `destroy-data` command?

NFS config is the only OpAdmin configuration affected by `destroy-data`.

What is the process for configuring backup on an appliance that was previously set up without it?

While NFS is accessible in the setup wizard for OpAdmin, it is also accessible post-installation:

1. Log into OpAdmin
2. Navigate in the menu to *Configuration*, then *NFS*
3. Complete the *Host* and *Path* fields (as you would on a new install), and select **Enabled (Pending Key)**.
4. Click **Save**.
5. To generate a new key click **Generate**, followed by **Download** to download it (whereafter it needs to be kept in a safe place); OR, to upload a previously-generated key, click **Upload**.

Can Threat Grid Support recover a lost encryption key for a backup?

No.

The encryption key for a backup store can always be downloaded while an appliance writing to that backup store is operational. However, if that appliance is not operational, **there is no key escrow mechanism or other means of recovering the key**.

Always be certain to download your key and store it securely!

Which guarantees, if any, exist regarding which appliance versions can restore a backup created by a given version?

A backup can be restored by any equal or newer appliance point-release within the same minor-version series as that which created it, or the immediately following minor release, **unless any contrary rule is explicitly documented in the Release Notes.**

For example:

We guarantee that a version 2.2.4 backup can be imported by 2.2.4, 2.2.5, etc.; or by 2.3.0. We *do not* guarantee that a 2.2.4 backup can be imported by 2.4.0 (a two-minor-version jump), or by 2.2.3 (an earlier point release in the same series).

Again - the above guidance is valid *by default*, but the release notes should be reviewed carefully for any variances from same.

Can I point multiple appliances against the same backup store with the same encryption key?

No.

This is not supported, and will cause data corruption. In a future release, multiple appliances will be able to use the same backup store with the same key if and only if they are members of a cluster; until this functionality is available, and unless it is in use, multiple appliances MUST NOT use the same backup store and key.

Can I force a PostgreSQL base backup to be performed outside the 24-hour cycle?

Yes.

Note that use of this functionality on a scripted, cronned, or otherwise regular basis is not recommended: The use of WAL logs to replay content added after a base backup makes a higher-than-default frequency of base backup execution is unnecessary under most circumstances.

1. Access the `tgsh-dialog` configuration interface, either via the appliance's TTY or via SSH.
2. Select the `CONSOLE` option to enter tgsh. (Note that entering tgsh via recovery mode is not suitable for this use case).
3. At the tgsh prompt, enter the command `service start tg-database-backup.service`.

This command will not return until the backup being invoked has completed or failed; pressing Ctrl-c or closing the tgsh session will not prevent the backup started by `service start tg-database-backup.service` from continuing its execution.

Can I mount a backup store read-only?

No.

The processes for interacting with a backup store are written to assume write access; this is why `rw` (like `nfsvers=4` and `sync`) is one of the mandatory mount options. If you want to ensure that a prior backup snapshot is retrievable, use a NFS store with snapshot functionality available, and leverage this functionality.