# Backups

This chapter describes Threat Grid Appliance requirements, expectations, data retention policy, and procedures for backups and restore. It includes the following topics:



# Threat Grid Appliance Backups

The Threat Grid Appliance (v2.2.4 or later) supports encrypted backups to NFS-backed storage, initialization of data from such storage, and reset to an empty-database state into which such a backup can be loaded.

**Note** Reset is different from the Wipe Appliance process; it is used to allow an appliance to be shipped off customer premises without information leakage, and is for backup preparation. The wipe process appropriate for that purpose already exists in the recovery bootloader, but is not suitable for preparing a system to restore a backup.

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

**Note** Filename encryption is disabled for performance reasons. Samples and other content in Threat Grid are not stored with their original names under any circumstances so this does not leak customer-owned data.

We strongly encourage consulting the documentation prior to use. Extended documentation regarding the backup functionality is available, and we strongly encourage consulting it prior to use. For additional technical information and instructions see the *Threat Grid Appliance Backup Notes and FAQ*, and the *Cisco Threat Grid Appliance Setup and Configuration Guide* on Cisco.com.

# NFS Requirements

The following NFS requirements must be met for encrypted backups to NFS-backed storage:

- Must be running the NFSv4 protocol over TCP, accessible from the Threat Grid Appliance admin interface.
- Configured directory must be writable by nfsnobody (UID 65534).
- The NFSv4 server must be accessible via the Admin 10-Gb interface.
- Sufficient storage must be available (see Backup Storage Requirements).
- The following mount parameters are unconditionally used: `rw`, `sync`, `nfsvers=4`, `nofail`

**Note** These parameters do not need to be manually entered; manually entering any parameters that conflict with them is explicitly unsupported and may result in undefined behavior.

- Invalid NFS configuration (or configuration pointing the service to 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.
- Exposing files for write by **nfsnobody** is secure. The only processes on the Threat Grid Appliance running as **nfsnobody** or with write to **nfsnobody**, are those responsible for encryption of data. Plain text 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.
- Using the **nfsnobody** account simplifies configuration, preventing the need to build an **idmap.conf** for each customer site, mapping local and remote account names together.

# File System

The Threat Grid Appliance (v2.7 and later) uses XFS as the primary file system, instead of the ZFS file system that was used on older appliances that have not been reset. This change does not affect pre-existing appliances except as otherwise noted (see Data Reset Process).

# Backup Storage Requirements

Total storage required for a backup store should not require more than 5.6 TB. A backup store consists of the following components:

- **Object Store** - This is normally 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 Threat Grid Appliance release in use and places maximum storage use for this component as 4.1 TB. See the Threat Grid Appliance Data Retention Notes.

- **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 500 GB in total.
- **Elasticsearch snapshot store** - This should be less than 1 TB in total.

# Backup Expectations

The following backup expectations should be considered.

### Included in Backup

The initial release of the Threat Grid Appliance backup process includes the following customer-owned bulk data:

- Samples
- Analysis results, artifacts, flagging
- Application-layer (not OpAdmin) organization and user account data.
- Databases (including users and organizations)
- Configuration done within the Face or Mask portal UI

### Not Included in Backup

The following is not included in the initial release of the Threat Grid Appliance backup process:

- System logs
- Previously downloaded and installed updates
- Configuration inside the appliance OpAdmin interface, including SSL keys and CA certificates

### Other Expectations

Other considerations about the backup process include:

- 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 backup takes place incrementally, once every 5 minutes.
- 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.
- 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 Data Retention

During a backup, data is retained as follows:

- **PostgreSQL** - The last two successful backups and all WAL segments since those backups are retained.
- **Elasticsearch** - The last two 5-minute snapshots are retained.
- **Bulk Storage** - The same retention policy followed and documented for a single Threat Grid Appliance is used for the shared store.

If you want to retain historical data for longer periods, it is strongly recommended that you use a NFS server with filesystem- or block-layer snapshot support.

Database base backups are only retained until a new base backup has been successfully created.

**Note** Backup copies of the virtual machine images are created on the RAID-1 storage array, to be used in the event of a reset following a bulk-storage failure. Early Cisco Threat Grid Appliance models (based on the UCS C220-M3 platform) have less storage than later models, and are more likely than other units to have less than 25 percent of disk space remaining available on the RAID-1 file system after installing Threat Grid Appliance v2.9, which will trigger a service notice.

For later model hardware, being at less than 25 percent of remaining storage on the RAID-1 array after installing the v2.9 release is not normal and should be raised to customer support.

## Strictly Enforce Retention Period Limits

The **strict_retention** option in **tgsh** (v2.6 or later) allows you to strictly enforce the retention period limit by not storing artifacts from analysis for more than fifteen (15) days. When this option is enabled, files are deleted during the first nightly pruning on which they are more than 15 days old.

**Note** The time period of 15 days cannot be configured or changed.

Artifacts refers to the samples themselves and other things generated from them. Artifacts do not include the analysis report HTML, which is subject to its original limits as otherwise documented. Artifacts also do not include database entries and search indexes.

The **strict_retention** option is disabled (false) by default. To enable the hard-pruning of artifacts after 15 days, set the option to true in **tgsh**:

**configure set strict_retention true**

# Backup Process Overview

The backup process on Threat Grid Appliance consists of the following steps:

**Step 1** Create the backup target directory according to the NFS Requirements.

**Step 2** Complete the **NFS** page in OpAdmin (**Configuration** > **NFS**), as described in the *Cisco Threat Grid Appliance Setup and Configuration Guide*.

**Step 3** Download the encryption key that is generated once you complete the NFS configuration. You need this key to restore the backup data.

**Important** The customer is responsible for backing up the encryption key and securely storing it. Threat Grid does not retain a copy. Backup cannot be completed without this key.

**Step 4** Reset the backup restore target as described in Reset Threat Grid Appliance as Backup Restore Target.

**Step 5** Restore the backup data as described in Restore Backup Content, on page 7.

# Backup Frequency

The backup frequency of data is as follows:

- For bulk storage of samples, artifacts and reports, content is continuously backed up. 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 16-MB 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.

Backup frequency cannot be controlled or tuned because doing so would make estimates regarding storage usage, restore-process time, and performance overhead invalid.

# Reset Threat Grid Appliance as Backup Restore Target

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.

**Caution** Performing this process will destroy customer-owned data. Read all of the documentation before performing any tasks, and be very careful before proceeding. .

**Note** Reset is not the same as the secure wipe that is 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. However, the secure wipe in recovery mode is not a replacement for this reset: secure wipe renders an appliance unusable until reimaged, while this reset prepares an appliance to restore a backup.

# Data Reset Process

The data reset process was updated in Threat Grid Appliance v2.7 and later and is now more comprehensive. While the Wipe process (in the recovery bootloader menu) is still required for a firm guarantee of the destruction of all customer-related data, the reset process now clears operating system logs and other state which was previously left in place.

A successfully reset Threat Grid Appliance now has a new randomly-generated password displayed on its console (identical to behavior in newly-installed state). This improved process now reboots multiple times, and can be invoked from recovery mode (as opposed to the prior process, which could only be successfully invoked when booted into regular operation).

If a Threat Grid Appliance has its data reset, the datastore will be changed from a ZFS file system to a XFS file system. This improves forward compatibility and provides OS-level support for I/O usage monitoring on a per-service basis.

The data reset process now also requires sufficient storage to contain all content necessary for a fresh install on the system SSDs. Any pre-existing data is only deleted after the presence and validity of this content has been ensured. It is possible that systems that have been in use for an extended period (particularly first-generation hardware), may not have sufficient space immediately available. If this is the case, customer support can assist, if needed.

# Return Target Appliance to Preconfigured State

If you are 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.

**Step 1** Access the TGSH Dialog via the Threat Grid Appliance TTY, or SSH.

**Step 2** Select the **CONSOLE** option to enter **tgsh**.

**Note** Entering **tgsh** via Recovery Mode is not suitable for this use case.

**Step 3** 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. All data will be destroyed.

***Figure 1: The destroy-data REALLY_DESTROY_MY_DATA Command and Argument***

```
Welcome to the ThreatGrid Shell.
For help, type "help" then enter.
>> destroy-data
To *really* run this command, pass the following string as an argument:
   REALLY_DESTROY_MY_DATA
Note that this is not intended as a security measure; use the recovery-
mode wipe process instead if thorough data destruction is required (and
the appliance will not be retained or used to load a backup).

DO NOT DO THIS WITHOUT DOWNLOADING YOUR BACKUP ENCRYPTION KEY FIRST!
>> destroy-data REALLY_DESTROY_MY_DATA
```

The following data is destroyed:

- Samples

- Analysis results, artifacts, flagging
- Application-layer (not OpAdmin) organization and user account data
- Databases (including users and organizations)
- Configuration done within the Face or Mask portal UI
- NFS configuration and credentials
- The local copy of the encryption key used for NFS

# Appliance Actively Writing to Backup Being Restored

If another system or Threat Grid Appliance is actively writing to the backup that is being restored, for example, a test restore of content being written by a second master Threat Grid Appliance actively used in production, return that Threat Grid Appliance to the preconfigured state.

**Step 1** Generate a consistent, writable copy of the datastore.

**Step 2** Point the Threat Grid Appliance that is doing the test restore to the writable copy instead of to the store which is being continuously written.

Once the Threat Grid Appliance is in a preconfigured state, it can function as the target for the backup store as described in Restore Backup Content.

# Restore Backup Content

**Important** The system is unavailable for sample submission during the restore process.

Perform the following steps to restore the backup content:

**Step 1** In the OpAdmin portal, click **Configuration** > **NFS** to open the **NFS** page.

**Step 2** 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 portal should 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, begins restoring the data to that location.

**Note** There is no progress bar. The amount of time required to restore data depends on the size of the backup and other factors. In testing, a 1.2-GB restore is quick, while a 1.2-TB restore required over 16 hours. For large restores it may appear that the install has hung so be patient. OpAdmin will report that the restore has succeeded, and the appliance will start up.

**Step 3** Confirm that the restored data looks the same as the original data.

# Backup and Restore Notes

- Sample submission is unavailable during the restore process.
- Backups can only be restored from the setup wizard.
- Set up the same NFS store and encryption key, as previously used, with a process identical to the original process.
- Setting up a Threat Grid Appliance with a prior NFS store and encryption key will trigger a restore.
- To test the restore process on a different Threat Grid Appliance while the primary Threat Grid Appliance is still operational, make a copy of a consistent snapshot of the backup store and point the new Threat Grid Appliance (with the encryption key uploaded) to it.

**Important** Only one server can be running with data from a given backup store active at a time.

# Backup-Related Service Notices

The following service notices may be displayed during the backup process:

- **Network storage not mounted** - Check that the network file system 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 file system 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 file system 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.