# Configuring full traffic logging

Topic 45054 / Updated: 22-April-2015

| **Applies To:** | Websense TRITON AP-WEB with Web Cloud Module |
|---|---|

This paper is intended for users of Websense TRITON AP-WEB with Web Cloud Module who want to download full Web traffic logs for retention and analysis. Once you have enabled traffic logging in TRITON AP-WEB, you can schedule a regular process to download the logs and save them to a location of your choice. Logs are retained in the cloud service for 14 days.

**Important**

Full traffic logging is an add-on for Websense TRITON AP-WEB with Web Cloud Module, and is separate from the standard Web reporting functionality. Standard reporting data is retained for 90 days and can be accessed through standard and custom reports; full traffic logs, once enabled, are retained for 14 days and are accessible through download only.

The section *Setting up full logging*, page 1 provides step-by-step instructions for setting up full traffic logging in the Cloud TRITON Manager, accessing the log files, and understanding the sample download script provided by Websense. *Downloading log files*, page 5 describes the issues you must be aware of when downloading the logs, and how to schedule the download process.

The section *File format definition*, page 8 describes the contents of a log file, with examples.

## Setting up full logging

The full traffic logging feature is not available by default. To make it available in your account, contact Support.

Once the feature is available, you must configure it in the Cloud TRITON Manager in order for TRITON AP-WEB to start generating and retaining your detailed log information.

# Configuring portal options

To set up full traffic logging in the Cloud TRITON Manager:

- Create a new administrator contact. We strongly recommend that the log download process has its own user name and password to gain access to the TRITON AP-WEB service. This keeps the process separate from your other administration tasks and enables you to establish longer password expiration policies.
- Enable traffic logging, either for your whole account or for specific policies.

## Creating a new administrator contact

To create the new contact:

1. On the main toolbar, click **Account**.
2. Click **Contacts**.
3. In the Contacts section, click **Add**.
4. Enter identifying information for the new contact in the **First name** and **Surname** fields. For example, “Traffic” and “Logging.”
5. Click **Submit**.
6. In the User Name field, click here to add a user name.
7. Enter a password for the contact. It must conform to the password policy on the main Contacts page.
8. Enter a password expiration date for the contact. To avoid having to regularly update it, this should be different than the regular account settings; it should span a longer period. The maximum period is 365 days.
9. Under **Account Permissions**, check the **Full Traffic Logging** box, and any other permissions you want to give this “user”. You can act as an administrator from this logon.

**Note**

If you give this contact only the **Full Traffic Logging** permission and nothing else, the user name and password cannot be used to log on to the Cloud TRITON Manager. The **View Reports** permission is the minimum permission a user needs to be able to log on.

10. Click **Submit**.

## Enabling full traffic logging

To enable log retention for your account:

1. On the main toolbar, click **Web**.
2. Under **Settings**, click **Full Traffic Logging**.
3. Click **Edit**.
4. Check the **Enable full Web traffic logging** box.

The text on this page states the conditions for using full traffic logging – namely, that all data is retained for only 14 days, and if you do not download any files for a period of 14 days, full traffic logging is automatically disabled. For more information, see *Troubleshooting*, page 11.

This page also contains a link to a sample script that you can use to download and store your log files. You can edit this script to suit your needs. For more information, see *Setting up a download script*, page 4.

5. Click **Submit**.

By default, all web policies have the logging setting that you define at the account level. If you want to change the log retention for a particular policy:

1. On the main toolbar, click **Web**.
2. Select **Policy Management > Policies**, and click the name of the policy you want to configure.
3. On the General tab, click **Edit**.
4. Under Full Traffic Logging, change the selection in the drop-down list from **Use policy-wide default** to either **Enabled** or **Disabled**. This overrides the account-level setting.
5. Click **Submit**.

You can view the logs available for your account by going to https://sync-web.mailcontrol.com/hosted/logs and logging on with the user name and password that you set up with the Full Traffic Logging permission. If you access this site immediately after you have set up full traffic logging, you will see only an empty XML script, but once TRITON AP-WEB has started to retain your logs, the page will show all available log files for download.

Each file name has the following format:

hosted_<*SourceID*>_<*AccountID*>_<*ClusterIP*>_<*Version*>_<*Epoch*>_<*SequenceNo*>.gz

Where:

| File name section | Description |
|---|---|
| SourceID | Internal string to identify where in the cloud service the log data was created. |
| AccountID | The TRITON AP-WEB internal identifier for your account. |
| ClusterIP | The IP address for the cluster that processed your web requests. |
| Version | Current version number for the log file format. |
| Epoch | The UNIX time for each log file, representing the time intervals between each generated log. |
| SequenceNo | If your logs exceed a certain size, the cloud service will deliver multiple log files for the configured time interval. This number identifies each log in the given period. |

For example, for log files from a cluster with the IP address 10.12.14.16 generated every 10 minutes, you might see the following:

```
hosted_xxxx1_1234_10.12.14.16_1_1236779400_1.gz
hosted_xxxx2_1234_10.12.14.16_1_1236780000_1.gz
hosted_xxxx1_1234_10.12.14.16_1_1236780000_2.gz
hosted_xxxx3_1234_10.12.14.16_1_1236780000_3.gz
hosted_xxxx1_1234_10.12.14.16_1_1236780600_1.gz
```

## Setting up a download script

To download the log files and save them to a location of your choice, you can either use the sample Perl script supplied by Websense or create a script of your own. To save the sample script to your network:

1. On the main toolbar, click **Web**.
2. Under **Settings**, click **Full Traffic Logging**.
3. Click **Edit**.
4. Click the sample script link, and save the file to a location of your choice. By default the file is named full_traffic_log_download.pl.

**Warning**

Websense, Inc., provides the sample log download script as a convenience to its customers, but does not provide support for customization and will not be responsible for any problems that may arise from editing the script.

The script can be run on Windows or Linux, and does the following:

- Connects to TRITON AP-WEB using the URL specified in the script
- Optionally reports the log files available for download
- Downloads the available log files to a location of your choice, or by default to the directory where the script is located
- Optionally checks the md5sum of each downloaded file to verify the file's integrity before deletion from the server
- Uses the HTTP DELETE method to request that the cloud service delete the downloaded files

**Note**

Running the script on Windows requires ActivePerl, which you can download from http://www.activestate.com/activeperl/downloads.

To see the ActivePerl modules required for running the script, open the script in a text editor. The modules are listed at the beginning of the file.

If you customize the sample script or choose to write your own script, you must always include the DELETE method to remove the downloaded files from the server. This is because files are only retained for 14 days, and any files that have not been deleted after 7 days will trigger a warning email. For more information, see

*Troubleshooting*, page 11.

# Downloading log files

To download log data when it is available, run the script that you have set up. If you are using the sample script provided by Websense, the available parameters to use with the script are described below. Some parameters have a short form (for example, `-v`) and a long form (for example, `--verbose`). For these parameters, both options are listed.

| Parameter | Description | Examples |
|---|---|---|
| `-u <`*username*`>`<br>`--username` | Mandatory. Defines the logon user name for connecting to the cloud service. This must be an administrator contact with Full Traffic Logging permissions. | `-u FTL_user@example.com` |
| `-p <`*password*`>`<br>`--password` | Mandatory. This is the password for the specified user name. | `-p Ft2010Logs` |
| `-v`<br>`--verbose` | Optional. Runs the script in verbose mode, which displays progress messages. | Verbose mode provides feedback on the script's progress, for example:<br>• Downloading filelist from <host name> as <user name><br>• No files available to download<br>• Downloading <file> to <file name location> |
| `-h <`*hostname*`>`<br>`--host` | Optional. Defines the host name to connect to. This is specified in the script by default, so you would only need this option if you have edited the script to remove it, or if you have been given a different URL to connect to. | `-h https://sync-web.mailcontrol.com` |
| `-d <`*file path*`>`<br>`--destination` | Optional. Defines the destination directory for the downloaded log files. If not specified, the files are downloaded into your current working directory. | `-d /cloudweb/logs` |
| `-m`<br>`--md5sum` | Optional. Checks the md5sum of each downloaded file. The MD5 hash is commonly used to verify the integrity of files (i.e. to verify that a file has not changed as a result of file transfer or disk error), and can therefore be used to check the files before they are deleted from the server. | n/a |
| `-l`<br>`--list-only` | Optional. Displays a list of available log files without downloading them. | n/a |

| Parameter | Description | Examples |
|---|---|---|
| `--proxy` <*proxy details*> | Optional. Specifies an HTTP proxy to use if you are having difficulty connecting to the cloud service. The proxy must be in the form http://*username:password@host:port* | `--proxy http://jsmith:Abc123@proxy_server:80` |
| `--format=` <*format*> | Optional. Creates a new data file containing the original downloaded data rewritten in the desired format. The new file's name has the relevant data format as a suffix. Note that when this parameter is used, by default the original *.gz file from the source server is not saved to the destination directory.<br>Valid data formats are:<br>csv : Comma Separated Values<br>cef : ArcSight Common Event Format<br>wc3 : WC3 Extended Log file Format (http://www.w3.org/TR/WD-logfile.html) | `--format=csv` |
| `--keepgz` | Optional. Use in conjunction with the `format` parameter to download and keep a copy of the *.gz data file in the destination directory. This overrides the default behavior of the `format` parameter. | `--format=csv --keepgz` |
| `--delete` | Optional. Deletes the original data file from the source server following download. The default option is to delete the file from the server. | n/a |
| `--nodelete` | Optional.Keeps the original data file on the source server after download. This parameter is provided for testing purposes while configuring the `format` parameter described above, enabling you to download a file in different formats. Note that files are still only retained for 14 days, and you will still receive a warning after 7 days if a downloaded data file remains on the server. | n/a |
| `--man` | Optional. Displays the list of parameters with their descriptions. | n/a |
| `--help` | Optional. Displays a brief description of the program's purpose. | n/a |

We recommend that if you wish to analyze the downloaded log files, due to the volume of data you should import the information into a database. For more information about the downloaded data, see *File format definition*, page 8.

# Scheduling log file download

Once you have run an initial download and determined the parameters you want to use in your script, you should set up a scheduled service to run automatic downloads in the background.

We recommend that you download the log files at least once a day. To avoid periods of high network traffic, it is advisable to select a random time for the download (for example, somewhere between 10 and 50 minutes past the hour).

## Scheduling on Windows

Before scheduling downloads from the cloud service, you must first ensure that the Windows Task Scheduler service is started. To check this:

1. Select **Start > Control Panel > Administrative Tools**.
2. Double-click **Services**.
3. In the Services window, scroll down to Task Scheduler.

   If the status is Started, you need do nothing. Otherwise, do one of the following:

   a. If the status column is empty, right-click Task Scheduler and select **Start.**

   b. If the status is Paused, right-click Task Scheduler and select **Resume**.

To schedule the log file download:

1. Select **Start > Control Panel > Scheduled Tasks**.
2. Double-click **Add Scheduled Task**.
3. Work through the Scheduled Task Wizard as follows:
   - Browse to the location where you have stored your script.
   - Select how often to perform the task (daily is advisable).
   - Select a time to start the task, and the start date.
   - Enter your network user name and password (**not** the user name and password you set up in the Cloud TRITON Manager). Windows requires these details to run the task successfully.
   - Check the **Open advanced properties for this task** box, and then click **Finish**.
4. The Advanced Properties for the task are displayed. On the Task tab, add the -u, -p, and -d parameters to the end of the **Run** field, before the closing quotes, as well as any other parameters you want to use. The **Run** field might look similar to this:

   ```
   "\\server\users\jsmith\hosted_logs\full_traffic_log_download
   .pl -u FTL_user@example.com -p Ft2010Logs -d /hostedweb/
   logs"
   ```

5. Click **OK**.

## Scheduling on Linux

Create a cron job to schedule your script to run at the times you want. For more information in Linux, see `man cron` and `man crontab`.

# File format definition

The log files downloaded from the cloud service are in JavaScript Object Notation (JSON) format. For more information about JSON, see http://www.json.org/.

Each log file contains multiple lines, with one request per line. Each line is enclosed in square brackets.

The following table describes the fields that comprise each request.

| Field | Description |
|---|---|
| DateAndTime | The time that a request occurred on the proxy, in seconds in UNIX time. |
| AccountID | The TRITON AP-WEB internal identifier for your account. |
| UserID | The web user's ID, usually their email address. |
| ClientIP | The client's external Internet IP address, shown as a number without the separating periods. |
| RequestCount | The number of requests for a particular site. This will default to 1 per log entry. |
| RequestSize | Size of the request in bytes. |
| ResponseSize | Size of the response in bytes. |
| Disposition | The disposition code of the request. For an explanation of the codes, see *Disposition codes*. |
| Categories | A comma-separated list of category IDs. To see how the ID numbers relate to category names, go to https://sync-web.mailcontrol.com/hosted/categories?version=2. (Note that this URL is for logs generated with 2015 Release 1 and later. To see ID numbers and category names for logs generated prior to that release, go to https://sync-web.mailcontrol.com/hosted/categories.) |
| Protocol | The protocol used in the request – for example HTTP, HTTPS, or FTP. |
| Port | The port number used for the request. |
| DestinationIP | The IP of the requested address, shown as a number without the separating periods. |
| URI | The full URL of the page requested by the user. |
| AnalyticID | Defines the analytic applied to the request. Can be one of the following:<br>• 1, 2 - Real-Time Security Scanning (RTSS)<br>• 4, 5, 6 - Advanced Detection (AD)<br>• 10 - Antivirus (AE)<br>• 11 - Real-Time Classification (RTC)<br>• 13 - Malicious iFrame Detection (MIDE)<br>• 14 - Malicious PDF Detection (SPIE)<br>• 15 - Advanced Secure Hash (ASH)<br>• 18 - Meta-analytic Detection (ICE) |
| ReasonCode | The reason code assigned to the request. For an explanation of the codes, see *Reason codes*. |
| ContentStripping | This field is blank in this version of the log file. |

| Field | Description |
|---|---|
| ReasonString | This is an internal signature ID string. |
| FileType | One of the following groups: 'unknown', 'text', 'executable', 'image', 'multimedia', 'document', 'suspicious', 'archive', 'ria', 'mime'. |
| PolicyName | Name of the policy used to filter the request. |
| ContentType | Content-Type of the response. The default value is an empty string. |
| RemoteHost | The host name of the origin server. |
| Method | HTTP method used in the request. |
| ProxyTime | The total delay, in milliseconds, due to filtering the transaction through the proxy. |
| OriginTime | The time taken, in milliseconds, to receive the request from the origin server. |
| ResponseTime | The total response time for the transaction, in milliseconds. |

# Disposition codes

The following table explains the meaning of the disposition codes used in the log files.

| Code | Description |
|---|---|
| 2 | Page blocked |
| 3 | Page filtered and permitted |
| 4 | Request filtering refused to service the request |
| 6 | Could not connect to requested site |
| 9 | Blocked, and user access to service disabled |
| 12 | Continue/confirm request |
| 13 | Quota request |
| 14 | Request permitted without filtering |

# Reason codes

The following table explains the meaning of the reason codes used in the log files.

| ID | Analytic | Name | Version |
|---|---|---|---|
| 1 | | Generic | 7.0 |
| 1 | Real-time security scanning | Generic | 7.0 |
| 2 | Real-time security scanning | Suspicious | 7.0 |
| 3 | Real-time security scanning | Exploit | 7.0 |
| 4 | Real-time security scanning | Redirection | 7.0 |
| 5 | Real-time security scanning | Obfuscation | 7.0 |
| 6 | Real-time security scanning | Evasion | 7.0 |

| ID | Analytic | Name | Version |
|---|---|---|---|
| 7 | Real-time security scanning | Couterfeit | 7.0 |
| 8 | Real-time security scanning | Spam | 7.0 |
| 9 | Real-time security scanning | Hijacked | 7.0 |
| 10 | Real-time security scanning | Defaced | 7.0 |
| 11 | Real-time security scanning | Tools | 7.0 |
| 12 | Real-time security scanning | Infostealer | 7.0 |
| 13 | Real-time security scanning | Backchannel traffic | 7.0 |
| 14 | Real-time security scanning | Remote control | 7.0 |
| 15 | Real-time security scanning | Installer | 7.0 |
| 16 | Advanced Detection | Malicious Packed | 7.0 |
| 17 | Advanced Detection | Generic Malicious | 7.0 |
| 18 | Advanced Detection | Trojan | 7.0 |
| 19 | Advanced Detection | Virus | 7.0 |
| 20 | Advanced Detection | Worm | 7.0 |
| 21 | Advanced Detection | Infected | 7.0 |
| 22 | Advanced Detection | Adware | 7.0 |
| 50 | Advanced Detection | Zipbomb | 7.0 |
| 300 | Malicious PDF Detection | Suspicious document | 7.7 |
| 301 | Malicious PDF Detection | Suspicious uncategorized document | 7.7 |
| 302 | Malicious PDF Detection | Document with active content | 7.7 |
| 400 | Malicious iFrame Detection | Malicious iFrame detection | 7.7 |
| 700 | Advanced Secure Hash | Generic | 7.8 |
| 800 | Meta-Analytic Detection | Generic | 7.8 |
| 900 | AppID | Generic | 7.8 |
| 10001 | Antivirus | Virus | 7.1 |
| 10002 | Antivirus | Adware | 7.1 |
| 10003 | Antivirus | Application | 7.1 |
| 10004 | Antivirus | Backdoor | 7.1 |
| 10005 | Antivirus | Bomb | 7.1 |
| 10006 | Antivirus | BootVirus | 7.1 |
| 10007 | Antivirus | Denial | 7.1 |
| 10008 | Antivirus | Dialer | 7.1 |

| ID | Analytic | Name | Version |
|---|---|---|---|
| 10009 | Antivirus | Downloader | 7.1 |
| 10010 | Antivirus | Exploit | 7.1 |
| 10011 | Antivirus | Intended | 7.1 |
| 10012 | Antivirus | Joke | 7.1 |
| 10013 | Antivirus | Macro | 7.1 |
| 10014 | Antivirus | MassMailer | 7.1 |
| 10015 | Antivirus | MisDisinfection | 7.1 |
| 10016 | Antivirus | NetWorm | 7.1 |
| 10017 | Antivirus | P2Worm | 7.1 |
| 10018 | Antivirus | Proxy | 7.1 |
| 10019 | Antivirus | PasswordStealer | 7.1 |
| 10020 | Antivirus | Remote | 7.1 |
| 10021 | Antivirus | Risk | 7.1 |
| 10022 | Antivirus | Spyware | 7.1 |
| 10023 | Antivirus | Tool | 7.1 |
| 10024 | Antivirus | Trojan | 7.1 |
| 10025 | Antivirus | HiddenProcess | 7.5 |
| 10026 | Antivirus | Injected Code | 7.5 |

# Troubleshooting

Your download script attempts to connect to the cloud service to download full traffic logs at an interval that you configure. If your script is unable to make the connection, or if it is unable to retrieve the log files after connecting, the following problems may occur:

- The cloud service stores log files for only 14 days. After that period, the files are deleted, and cannot be recovered. When this occurs, your organization is no longer able to access and analyze Web filtering activity recorded in those logs.
- Depending on the volume of Internet activity that your organization sends through the cloud service, log files may grow quickly. If your script is unable to download log files for a day or more, the bandwidth required to download the files and the disk space required to store them may be substantial.

To address this issue:

- Check that your scheduling service (Windows Task Scheduler, or crontab on Linux) is running. If you are using Windows Task Scheduler, check that it is using your most recent network password to run the task.
- Your script may be prevented from accessing the cloud service due to network problems, either affecting Internet or internal network connections. Use a browser

or the **ping** utility to verify that the machine running the script can connect to the Internet.

- If the script is connecting to the cloud service but cannot retrieve log records, verify that there is not a problem with the cloud service. Check the administrative email address associated with your full traffic logging account.
- Check that your TRITON AP-WEB password has not expired.

If you do not download traffic logs for a period of 7 days, a notification email is sent to all administrative contacts with Full Traffic Logging permission enabled, and all policy administrators where full traffic logging is enabled for the policy. The email warns that logging will be disabled if you do not download logs for 14 days. Further notifications are sent after 10 and 13 days, and after 14 days you will be notified that full traffic logging has been deactivated and traffic logs are no longer being generated for your account.