opentext™

OpenText
ArcSight Log Anonymizer

Software Version: 0.6.0 Beta

Installation and User Guide for Log Anonymizer

Document Release Date: October 2023
Software Release Date: October 2023

Legal Notices

Open Text Corporation

275 Frank Tompa Drive, Waterloo, Ontario, Canada, N2L 0A1

Copyright Notice

Trademark Notices

Documentation Updates

The title page of this document contains the following identifying information:

- Software Version number
- Document Release Date, which changes each time the document is updated
- Software Release Date, which indicates the release date of this version of the software

To check for recent updates or to verify that you are using the most recent edition of a document, go to:

https://www.microfocus.com/support-and-services/documentation

Log Anonymizer

This guide provides information about installing and configuring the Log Anonymizer tool.

Intended Audience

This guide provides information for IT administrators who are responsible for managing the ArcSight software and its environment.

Additional Documentation

The ArcSight SmartConnector documentation library includes the following resources:

- Technical Requirements Guide for SmartConnector, which provides information about operating system, appliance, browser, and other support details for SmartConnector.
- Installation and User Guide for SmartConnectors, which provides detailed information about installing SmartConnectors.
- Configuration Guides for ArcSight SmartConnectors, which provides information about configuring SmartConnectors to collect events from different sources.
- Configuration Guide for SmartConnector Load Balancer, which provides detailed information about installing Load Balancer.

For the most recent version of this guide and other ArcSight SmartConnector documentation resources, visit the documentation site for ArcSight SmartConnectors 8.4.

Contact Information

We want to hear your comments and suggestions about this book and the other documentation included with this product. You can use the comment on this topic link at the bottom of each page of the online documentation, or send an email to MFI-Documentation-Feedback@opentext.com.

For specific product issues, contact Open Text Support for Micro Focus products.

Product Overview

Log Anonymizer is a tool that anonymizes Personally Identifiable Information (PII) defined in the security related log files. It will anonymize any text-based data, including IP, MAC, and email addresses to help ensure that PII data is obfuscated. The originating log file data is never altered.

Not having access to customer generated log files is a key contributor for increasing lag times in keeping parsers up-to-date with the latest versions of security device log data. The purpose of Log Anonymizer is to enable customers and partners to send anonymized log data to the ArcSight teams to keep the SmartConnector parsers up-to-date with the latest changes.

Anonymizer Features

Log Anonymizer uses a step-by-step wizard process that supports the anonymization of:

- Field names defined in log data generated and parsed by a SmartConnector.
- Field names defined in multiple text-based log formats and field structures such as CEF, JSON, and XML.
- Unstructured log data (for example, Syslog), including any new text-based log formats.
- Single log files, multiple selected log files, or all log files in a folder. The Anonymizer may be used for anonymizing log data for Intelligence analytics using the same anonymization key for all log data.
- View log data anonymization as you select fields to be anonymized. Anonymization of strings is case-sensitive, so ‘THOR’ and ‘thor’ are different strings and must be anonymized separately.

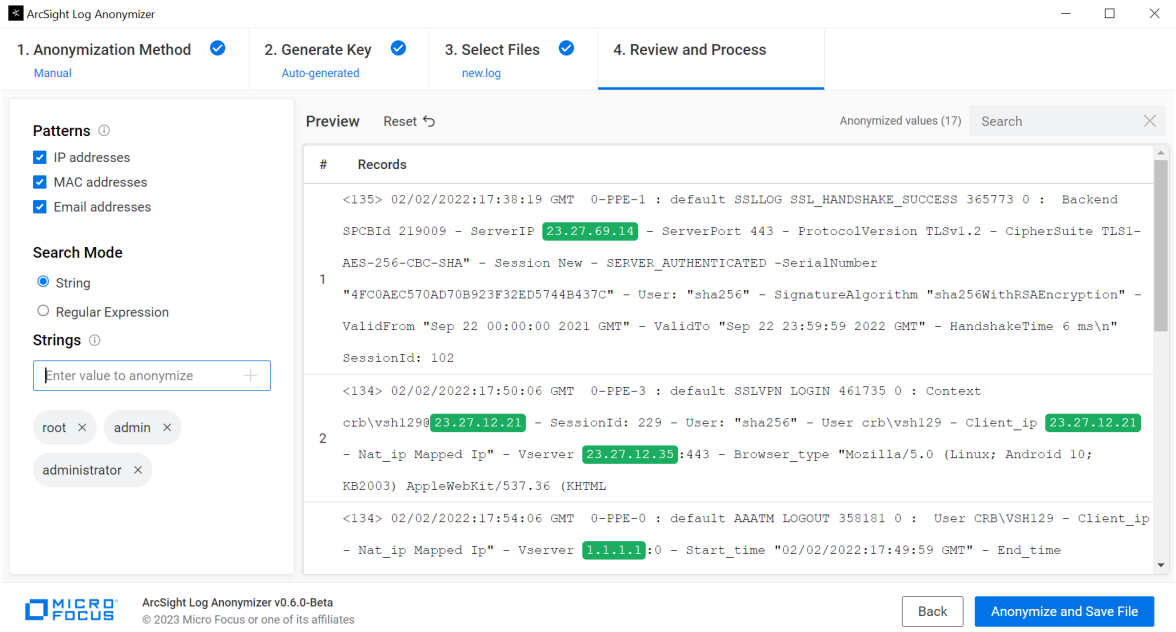
  

  Note: After a string or field is anonymized, if you do not like the resultant value, then you can re-anonymize the string value to a new value.

- Configurable and predefined text strings that are always anonymized (for example: ‘root’, ‘admin’, ‘administrator’, and so on.)
- Built-in scanning and anonymization of all IP, MAC, and email addresses. Note that IP Addresses might be the assigned values that are not valid (for example: 809.264.312.418).

Regardless of the anonymization option chosen, the source log file is never altered.

The Anonymizer supports previewing and anonymizing log record fields and strings prior to saving the anonymized file(s).



However, there are some limitations to previewing log record fields and strings as follows:

- Log files > 100 MB are not previewed when the **Anonymize Using a Connector** option is chosen.
- Log files > 200 MB are not previewed when the **Anonymize Manually** option is chosen.

Anonymizer Limitations

The Anonymizer has the following limitations:

- It supports only SmartConnector regex parsers. For non-regex parsers, you must use the **Anonymize Manually** option.
- It does not function if you do not have access to a valid SmartConnectors un-obfuscated parser folder.

Preparing Log Files Prior to Anonymizing

The contents of a security log file can be a few records to millions of records. Although the Anonymizer obfuscates data regardless of the number of records. Ensure that the log file that needs to be anonymized and sent to the ArcSight team for parser file updates contains, at a minimum, a single record of each log record type.

Consider using a text editor (such as notepad++) to delete unnecessary records from the log data source file prior to anonymizing, limiting the number of records to those that are new log record types or contain only changes.

Best Practices and Warnings

The following are the best practices for using the product and warnings you might encounter while using the product:

- Because anonymization is a CPU intensive operation, do not install and run the Anonymizer on a system that is running production or sensitive workloads.
- It is a best practice to copy the security logs to a different system from their source system, edit them with text editor to delete unnecessary log file data, and run the anonymization on a non-production machine.
- Large log files might take minutes to hours to complete, while small log files (< 10MB) might take seconds to complete. If you run anonymization for large files on your personal workstation, you might experience CPU resource contention issues.
- Ensure that you have sufficient disk space to store the anonymized log file and its compressed finished output (.zip file). Free space must be at least twice the size of the log file or files being anonymized.

Launching the Anonymizer

The Anonymizer is a Windows-based application. It is delivered as a self-contained .zip file. To launch it, just unzip it on the system you want the anonymization process to run on.

The zip file contains:

- **\bin** – Application binaries files and executable. Double-click **Anonymizer.exe** to launch the Anonymizer.
- **\doc** – Documentation.
- **\samples** – Sample log files for your initial testing and familiarity with the Anonymizer.
- **Readme_before_installing.txt** – Guidance before installing the Anonymizer.

The **\samples** folder contains the following sample log files and subfolders:

Log File Name	Format	Parser	Description
2021-12-20-13-16-22.885.cef	CEF	None	This is a structured CEF log as an example of what a vendor might provide, but it has no parser yet. Use the **Anonymize Manually** option.
allevents.xml	XML	None	This is a structured XML log as an example of what a vendor might provide but it has no parser yet. Use the **Anonymize Manually** option.
logs.json	JSON	None	This is a structured JSON log as an example of what a vendor might provide but it has no parser yet. Use the **Anonymize Manually** option.

unstructured.log	unstructured	None	This is an unstructured log as an example of what a vendor might provide, but it has no parser yet. Use the **Anonymize Manually** option.
\aixaudit\aixaudit_syslog.log and \aixaudit\aixaudit_syslog_45kb.log	CEF	Choose the 'syslog' connector and 'aixaudit_syslog'	Use the **Anonymize Using a Connector** option.
\whatsup\whatsup_syslog.log	CEF	Choose the 'syslog' connector and' whatsup_syslog'	Use the **Anonymize Using a Connector** option.

Operating System Requirements

Windows 10 or later

Using the Anonymizer

To launch and use the Anonymizer, browse to the installation folder and double-click the **\bin\anonymizer.exe** executable.

Anonymizer Methods

The following methods are available to anonymize the log files individually or all log files available in a folder:

- **Anonymize Using Connectors**

  Use this method if there is a SmartConnector and Parser for the log source, even if the log source data is not fully recognized and parsed by the current version of the parser.

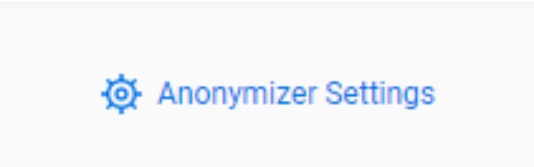
  When this method is chosen, the Connector and its parser must match the log data source. For example, an AIXauditd log will require a parser file specifically written for AIXauditd and will neither recognize nor anonymize a Cisco Pix log.

- **Anonymize Manually**

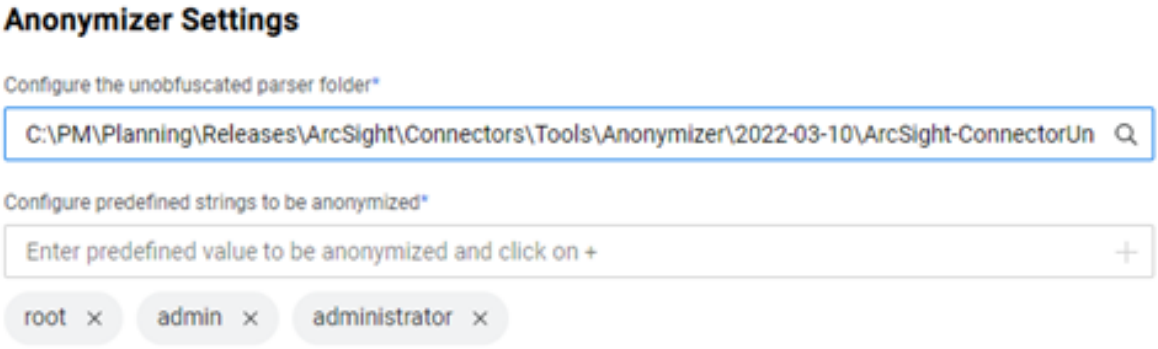
  Use this method to manually anonymize all occurrences of the selected PII text strings. This method is independent of parser availability for the log source.

Anonymizer Settings

You must configure the Anonymizer to anonymize any files. To open the configuration dialog, click **Anonymizer Settings**.



The following dialog is shown.



Configuring the Un-obfuscated Parser Folder

The Anonymizer works only if it has access to a valid ArcSight SmartConnectors unobfuscated parser folder. The un-obfuscated parser folder is shipped with the SmartConnectors and requires an ArcSight license to access it.

To configure the un-obfuscated parser folder:

1. Open the **ArcSight-ConnectorUnobfuscatedParsers-v.r.sp.build.x.zip** file present in the SmartConnector installation media.
2. Copy the Un-obfuscated parser folders to your system and unzip them.
3. In the **Configure the unobfuscated parser** folder field in the **Anonymizer Settings** dialog, specify the location of the unzipped folder obtained in the previous step.

Configuring Predefined Strings for Anonymization

The predefined text strings that are anonmyzied by default are "root", "admin", and "administrator".

To add a new string, specify the string in the **Configure predefined strings to be anonymized** field in the **Anonymizer Settings** dialog, then press Enter or click **+**.

To remove a default or a newly added string, click **x** against the string.

Anonymization Key

All Anonymizer methods require you to set an alphanumeric key that is used to anonymize log text. You can either provide your own key by typing it in, or you can generate a 64-character key. The key can be downloaded and saved to a text file in case you want to use the same key when anonymizing other log files.

Note: For the ArcSight Intelligence product, logs typically spanning a 30-to-45-day duration are required to determine analytic baselines and anomalies. You must use the same anonymization key for log files spanning multiple days. For ArcSight SmartConnectors, you can use the same anonymization key or different anonymization keys when anonymizing individual log files.

Anonymization Pattern Matching

The Anonymizer scans log records for specific patterns and then anonymizes the text that matches any of these patterns. You can toggle the pattern-matching on and off by checking which patterns you would like to be enabled. The supported patterns are:

- IP addresses
- MAC addresses
- Email addresses

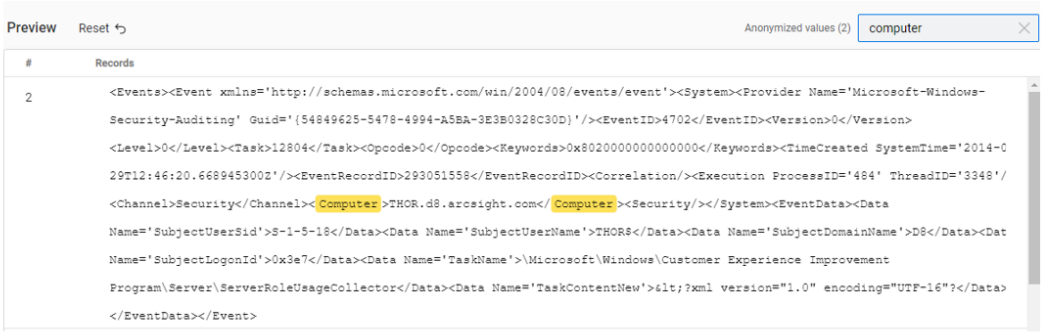
For IP addresses, both IPv4 and IPv6 addresses are recognized and anonymized.

Anonymization of Specific Fields

The Anonymizer enables you to search for specific fields in logs and anonymize those fields. The following are the ways by which you can search for fields to anonymize:

- **Preview Page Search Mode**

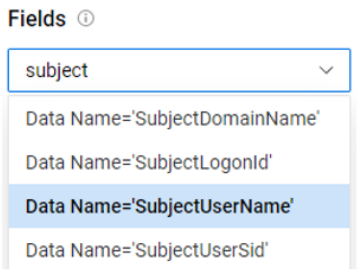
  The Search Mode option is a case-insensitive search option available in the upper right corner of the **Preview** page. With this option, you specify a text in the search box, and only those records that contain the text are displayed, with every instance of the text in each of the records highlighted in yellow.

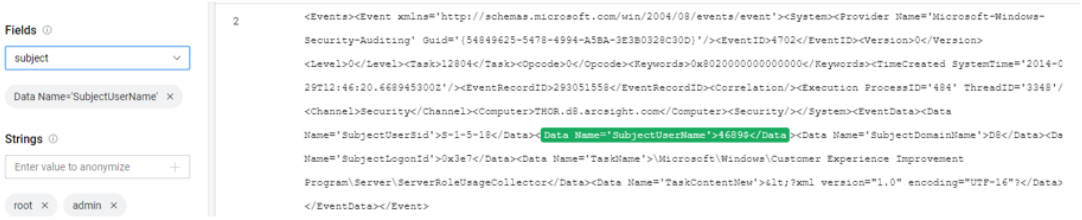


  To exit search mode, press **X** or delete the specified text from the search box.

- **Field Name Search**

  When anonymizing a structured log file (such as CEF, JSON, or XML), you can specify a text in the **Fields** search box. The search result is a list of fields or tags that begin with the specified text. You can then select the required field or tag from the list for anonymization.





Anonymization Using Connectors

This option uses a SmartConnector parser to format the log data into **FieldName=FieldValue** pairs for the fields that are recognized by the parser. You might also anonymize the default string values or specific strings as required.

Perform the following steps to anonymize the data using connectors:

1. Double-click the **\bin\Anonymizer.exe** file to launch the Anonymization tool.
2. Select the **Anonymize Using Connectors** option.
3. Select the connector from the list.
4. (Conditional) If you choose the **syslog** connector, you must next choose which syslog connector is associated with the log source.
5. Enter your Anonymization Key or generate the key automatically. To generate a 64-character key, click **Generate**. The maximum length of the Anonymization Key is 64 characters.
   - To copy the key, click **Copy Key**.
   - To download the key and save it to a file, click **Download Key**.

6. Choose one of the following options to determine the log files for anonymization:
   - **Select Specific Log Files**

     Browse for and select one or more log files to anonymize. After selecting the files, if you need to remove any of them, click x for the corresponding file.

     If you have selected only one log file to anonymize, you need to view the log file and anonmyize manually by clicking Preview Log.

     If you have selected more than one log file, you need to determine which fields and strings need anonymization. All occurrences of the fields and strings in all the selected files are anonymized. You cannot manually adjust the anonymization in any of the files.

   - **Select All Log files in a Folder**

     Browse for the folder to anonymize. When you choose this option, all the files in the selected folder are anonymized, except for subfolders, .zip, and .exe files.

     You need to determine which fields and strings need anonymization. You cannot manually adjust the anonymization in any of the files.

     Warning: Ensure that only the required log files are available in the folder. Existence in the folder is sufficient to have the file anonymized.

7. Click **Save Anonymized Files**. The files are anonymized and saved to a **.zip** folder. The zipped folder also contains the parser that was used for anonymization.

After reviewing the contents of the .zip folder, you can share the .zip folder with OpenText to help us improve our connectors and parsers using your anonymized PII data.

Anonymize Manually

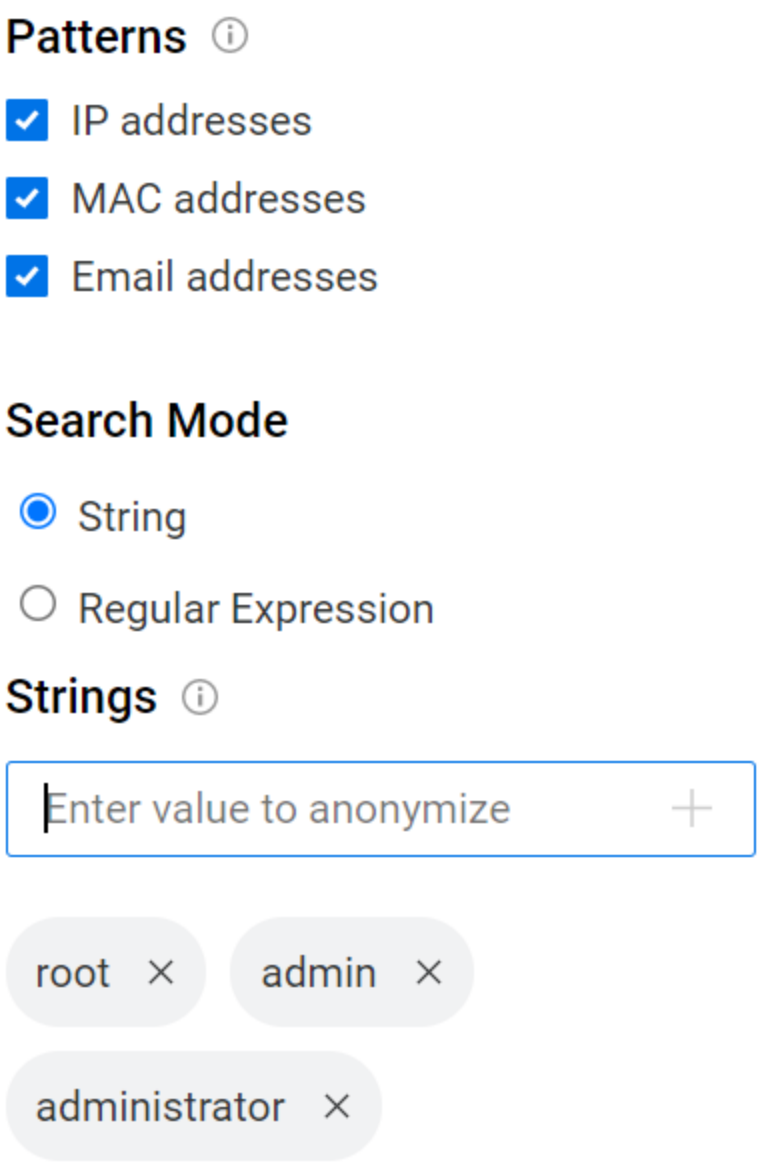
The Anonymizer recognizes both structured and unstructured log formats. Manual anonymization on structured logs that are formatted in either CEF, JSON, or XML enables anonymization based on field names or tags defined in the log data. Unstructured log files do not conform to either of these formats. For examples of these log formats, see "Log Format Examples" on page 17.

Perform the following steps to anonymize manually:

1. Double-click the **\bin\Anonymizer.exe** file to launch the Anonymization tool.
2. Select the **Anonymize Manually** option.

3. Enter your Anonymization Key or generate the key automatically. To generate a 64-character key, click **Generate**. The maximum length of the Anonymization Key is 64 characters.
    - To copy the key, click **Copy Key**.
    - To download the key and save it to a file, click **Download Key**.
4. Browse to and select the log file name to anonymize. Click **Preview Log and manually anonymize the log**. The log file details are displayed record-by-record.
5. In the event log shown, select the patterns, field names, string values, and custom regex values you want to anonymize. The matching text is highlighted in green. Check the patterns you want scanned for and enter the string/custom regex values you want to be anonymized.
    - For CEF, JSON, and XML formatted logs, choose the field names or tags you want to anonymize. All occurrences of the field names or tags are anonymized. Or, click **X** next to it to reset its value.
    - To add a string, type it into the Strings dialog box and press **+**. All occurrences of the strings are anonymized. Or, click **X** next to it to remove its value.
    - To add a custom regex, type a regular expression into the custom regex dialog box and press **+**. All occurrences of the strings that match the regular expression are anonymized. Or, click **X** next to it to remove it.
    - To add a pattern, select the check box of the pattern. To remove a pattern clear the check box.

Note: After a string or field is anonymized, if you do not like the resultant value, you can re-anonymize the original string or field by specifying it in the **Strings** dialog box, and clicking **+**.



6. To discard all your anonymization changes or to start over, click **Reset**.

   A warning is displayed. Click **OK**, to revert file to its original state without anonymization.
7. Use the **Search** text box to search for the required text within the file. Records containing the text are displayed in the table. If there are no matches, no records are shown. To exit search mode and see the full log, you must remove all text in the **Search** field.
8. Click **Save Anonymized File** to save the anonymized file to a .zip file.

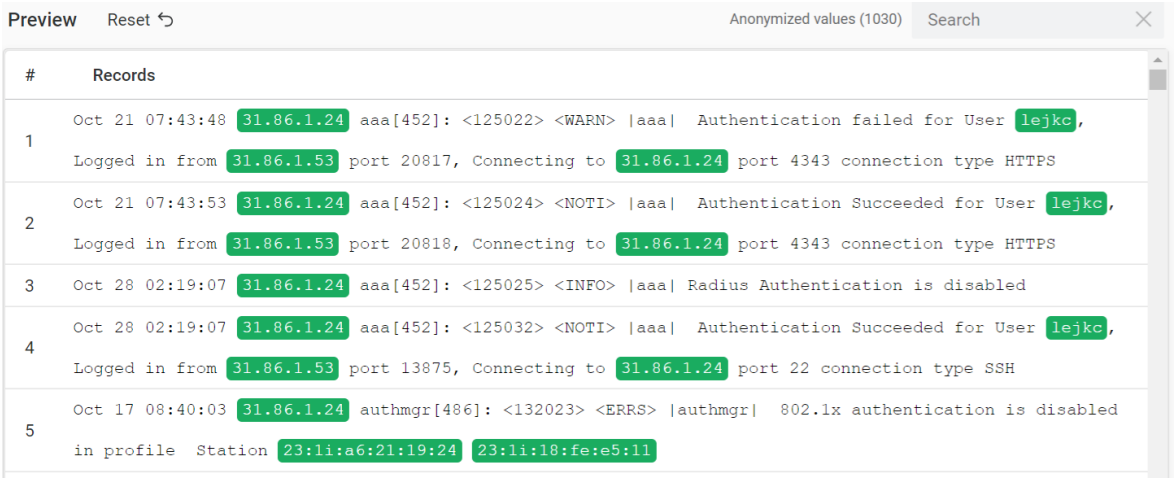
9. To close the current log and go back to the main screen, click the **Close Log File** icon in the upper right corner and go back to the main screen. Changes performed to the file are not saved.

After reviewing the contents of the .zip folder, share the .zip folder with OpenText to help us improve our connectors and parsers.

Log Format Examples

This section contains some sample anonymized logs formatted from unstructured, CEF, JSON, and XML log sources.

Unstructured Log Example



CEF Formatted Log Example

JSON Formatted Log Example

```
16                  "device": "Computer",
17                  "id": "0oaq2bbc2ROsU1J194x6",
18                  "ipAddress": "14.223.908.21",
19                  "geographicalContext": {
20                          "city": null,
21                          "state": "California",
22                          "country": "United States",
23                          "postalCode": null,
24                          "geolocation": {
25                                  "lat": 34.0544,
26                                  "lon": -118.244
27                          }
```


XML Formatted Log Example


```
#   Records
1   <?xml version="1.0" encoding="utf-8" standalone="yes"?>
2   <Events><Event xmlns='http://schemas.microsoft.com/win/2004/08/events/event'><System><Provider Name='Microsoft-Windows-
    Security-Auditing' Guid='{54849625-5478-4994-A5BA-3E3B0328C30D}'/><EventID>4702</EventID><Version>0</Version>
    <Level>0</Level><Task>12804</Task><Opcode>0</Opcode><Keywords>0x8020000000000000</Keywords><TimeCreated SystemTime='2014-0
    29T12:46:20.6689453002'/><EventRecordID>293051558</EventRecordID><Correlation/><Execution ProcessID='484' ThreadID='3348'/
    <Channel>Security</Channel><Computer>4689.22.ba08393e.416</Computer><Security/></System><EventData><Data
    Name='SubjectUserSid'>3-5-1-82</Data><Data Name='SubjectUserName'>4689$</Data><Data Name='SubjectDomainName'>D8</Data><D
    Name='SubjectLogonId'>18140</Data><Data Name='TaskName'>\Microsoft\Windows\Customer Experience Improvement
    Program\Server\ServerRoleUsageCollector</Data><Data Name='TaskContentNew'>&lt;?xml version="1.0" encoding="UTF-16"?</Data>
    </EventData></Event>
```

Beta Component Support and Contacts

This release is a Beta version and is not supported by the ArcSight Customer Support Team. It is supported directly by ArcSight R&D and Product Management.

Contact the people listed below for assistance and feedback on the Log Anonymizer

- Feedback -
  - Product Manager - Wayne Dalesio (wdalesio@opentext.com)
  - Product Manager - Prentice Hayes (phayes@opentext.com)
- Assistance - R&D Engineers:
  - Rohith Nandakumar (rnandakumar@opentext.com)
  - Sagar SM (ssm@opentext.com)
  - Sherin Joy (sjoy@opentext.com)

Send Documentation Feedback

If you have comments about this document, you can contact the documentation team by email. If an email client is configured on this computer, click the link above and an email window opens with the following information in the subject line:

Feedback on Installation and User Guide for Log Anonymizer (Log Anonymizer 0.6.0 Beta)

Just add your feedback to the email and click send.

If no email client is available, copy the information above to a new message in a web mail client, and send your feedback to MFI-Documentation-Feedback@opentext.com.

We appreciate your feedback!