FORCEPOINT
POWERED BY Raytheon

Forcepoint UEBA

Management of Personal Data

Forcepoint UEBA– Management of Personal Data

CONTENTS

Disclaimer 2
General 3
Document Purpose 3
General Data Protection Regulation (GDPR) 3
Personal Data 3
Safeguarding Personal Data 3
Event or Element Data 4
Ingest Pipeline 4
Ingest Pipeline Logs 5
RedOwl Service for Entities (ROSE) 5
RedOwl Service for Entities (ROSE) Logs 5
Master Data Service 5
Master Data Service Logs 6
UI Services 6
UI Services Logs 7
Monitoring 7
Appendix A 8
Table 1: Entity Data – Common 8
Table 2: Entity Data - Attributes 8
Table 1: Event Data - Types 8
Table 2: Event Data – Generic Structure 9

Disclaimer

This document contains information regarding Forcepoint products and/or services. The information is Forcepoint's property. While every effort was made to ensure the content is up-to-date and accurate, the information is provided AS IS, without any representation or warranty, express or implied, and is subject to change without notice. Any references to future releases or functionality are forecasts and not intended to be commitments. Forcepoint assumes no liability for the use of this information.

General

Document Purpose

This document is designed to answer the question: “What personal data is stored in the Forcepoint UEBA product?” It is primarily intended for those involved in the procurement and privacy assessment of the Forcepoint UEBA product.

General Data Protection Regulation (GDPR)

The General Data Protection Regulation (GDPR) (Regulation (EU) 2016/679) was adopted on April 27, 2016 and came into effect on May 25, 2018. GDPR, which replaced the Data Protection Directive 95/46/EC, is a significant source for the privacy principles that guide Forcepoint’s privacy policies and processes, both internally and externally. Full details of the GDPR can be found in various sources, including https://ec.europa.eu/info/law/law-topic/data-protection/reform_en

The Forcepoint UEBA service is designed to enable GDPR compliant implementation. Consistent with GDPR’s principles, Forcepoint’s customers are considered to be the sole data controller. It should be noted that Forcepoint is neither the data controller nor data processor with respect to customer data stored in Forcepoint UEBA.

Personal Data

This document adheres to the definition of personal data as defined in article 4.1 of the General Data Protection Regulation, which defines 'personal data’ as any information relating to an identified or identifiable natural person (‘Data subject’); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier, or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural, or social identity of that natural person.

Safeguarding Personal Data

Forcepoint uses industry-standard techniques to protect data held within the Forcepoint product, including personal data.

Event or Element Data

Data Set	What Data is Used?	Purpose	Is Anonymization Possible?	Storage, Flow & Protection	Retention

Ingest Pipeline	Event Data
	
	Entity Data
		The Ingest Pipeline allows the continuous processing of events in which a pipeline service inserts event data into an Elasticsearch datastore to be used by the rest of the application stack.	Partial masking is possible. The entities involved in the event (i.e., sender, recipients, chat room participants, etc.) can be masked, but the body of those events cannot be.	Data Protection Summary
				
				Physical protection for appliance & Amazon Web Service (AWS)
				• Forcepoint UEBA supports volume encryption for data at rest when you configure CentOS.
				• Native/Attachment files are stored in either cloud storage (e.g., Simple Storage Service with optional encryption) or a network file share (optional encryption)
				• Basic protections are derived from the AWS environment.
				
				Access control
				• Secure Shell (SSH) access is enabled by default. PowerShell (PS) sets up access using your IP address. PS does not maintain separate access.
				• APIs
				  ○ Secure Sockets Layer (SSL)/Transport Layer Security (TLS) certifications are required to make API requests. Use client-side certifications for authentication. Data is always encrypted in transit using TLS.
				  ○ At installation, certifications are created and deployed per machine, and all credentials are stored in memory.
				  ○ No machines can be exposed by default except through the user interface or continuous
					In steady state, the pipeline queue is emptied,
					
					Event processing errors are sent to an error queue, where processing is retried. If the event continues to fail, it is published to a database in a monitoring environment.

Data Set	What Data is Used?	Purpose	Is Anonymization Possible?	Storage, Flow & Protection	Retention
				integration tool (software upgrades only) in AWS. **NOTE**: Accessing SSH per network address translation is required for proof of concepts using bastion host/JumpBox in an AWS virtual private cloud.	
Ingest Pipeline Logs	Logs of Event Data	Error, application, and debug logs for pipeline activities		See the Data Protection Summary above.	By default, logs are retained for 15 days with daily rotation.
RedOwl Service for Entities (ROSE)	Name Identifiers	ROSE is capable of storing and managing customer specified reference information such as a name, work email, personal email, Skype handle, IP address, phone number, etc.; as well as other attributes, such as salary, supervisor, office location, etc.. Customers determine the reference information to be stored and managed based on their specific use case	Masking is possible. However, ROSE stores both the masked version of an entity (e.g., "Blue Garden Bird") as well as the real identity (i.e., entity's full name) so that the UI can allow super users to view event senders.	See the Data Protection Summary above. ROSE stores data within the database or Elasticsearch and not directly on a disk.	Entity attributes and identifiers tend to be persistent for the course of the installation as they are referenced by new events as ingested. By default, the data is stored in a Postgres database and replicated to Elasticsearch with customer specified retention period.
RedOwl Service for Entities (ROSE) Logs	Logs Records of ROSE updates of new/deleted information	Error, application, and debug logs for ROSE activities.		See the Data Protection Summary above.	By default, logs are retained for 15 days with daily rotation in the following location: /var/log/ro-rose Logs are replicated to Elasticsearch monitoring instance.
Master Data Service	Event Data Element Data	The Master Data Service maps raw values to known entities, periodically retrieving this mapping from Elasticsearch. The processor resolves the roles for each event.	Partial masking is possible. The entities involved in an event (i.e., sender, recipients, chat room participants, etc.) can be masked, but the body of those events cannot be masked. Master Data Service prevents unauthorized users	See the Data Protection Summary above.	Data is stored in the database with no automated retention by default. Retention is configurable. Typical event retention is 6-12 months. Elasticsearch is used for querying data, applying analytics to the data, and

Data Set	What Data is Used?	Purpose	Is Anonymization Possible?	Storage, Flow & Protection	Retention
			from viewing the event content.		applying meta information to the data. The original data is not modified.
Master Data Service Logs	Logs	Error, app, and debug logs for Master Data Service activities.		See the Data Protection Summary above.	By default, logs are retained for 15 days with daily rotation in the following location: /var/log/ro-mds Logs are replicated to Elasticsearch monitoring instance.

UI Services	UI Services allows you to view the following user data
	
	• Event data
	• Element data
	• Entity data
	• User preferences
	• User settings
	• Saved search queries
		Primary user interface (web app) where customer interacts with ingested data	Partial masking is possible. The entities involved in the event (i.e., sender, recipients, chat room participants, etc.) can be masking, but the body of those events cannot be masked. The UI Services prevent unauthorized users from viewing event content.	See the Data Protection Summary above.
				
				Logical Access Controls
				
				• Multi-factor authentication (MFA) is not supported beyond Security Assertion Markup Language (SAML) requirements.
				• Login with single sign-on (SSO) using public key infrastructure (PKI) or SAML
				• Login with a username and password
				  ○ Password complexity is enforced.
				  ○ No password rotation enforced
				  ○ Passwords are stored in a Postgres database (bcrypt as default, unique salt with SHA-512 available).
				  ○ HTTPS
				• Authorization roles limit the graphs or charts that are visible. Entitlements/SafeSearch limits the data that the charts or graphs display.
				  ○ For example, you cannot view your manager's salary data.
					Data is stored in the database and UI Services use Elasticsearch.

Data Set	What Data is Used?	Purpose	Is Anonymization Possible?	Storage, Flow & Protection	Retention
UI Services Logs	Specific Logs/Reports are built for the customer specific compliance needs. Log information can include: ○ Sender ○ Recipient ○ Subject ○ Timestamp for email, chat, or text ○ Events	Audit, request, error and debug logs for UI Services activities.		See the Data Protection Summary above.	By default, logs are retained for 30 days with daily rotation in the following location: /var/log/ro-ui Logs are replicated to Elasticsearch monitoring instance.
Monitoring	Logs from all these services are replicated in a monitoring instance: • Ingest Pipeline Logs • ROSE Logs • Master Data Service Logs • UI Services Logs	Instance where logs from the various services are replicated. It also monitors system operations and performance, as well as log files to track backend activity. This helps you to understand functional details, efficiently address bugs, and identify potential need for scaling in your system.		See the Data Protection Summary above.	Data is stored in Elasticsearch and can be manually deleted based on your specific compliance needs.

Appendix A

ENTITIES

A person, place, or thing for which a pattern of behavior can be identified. The customer can designate employees as entities, but can just as easily use an office, a building, or a computer as an entity. Every event (email, chat, printer job, etc.) has an entity associated with it. The entity can be an end user the customer wishes to track closely, or it could be someone that the customer either doesn't know or have a desire to know (like the sender of bulk email).

Table 1: Entity Data – Common

Every entity can have the following data associated with it, depending on your specific needs.

Attribute	Requirement	Description
actorId	Mandatory	The name of an entity (e.g., Jane Doe)
alsoKnownAs	Mandatory	All the ways we can identify this entity (e.g., email addresses, phone numbers, IP addresses, chat handles, etc.)
attributes	Optional	See Table 2

Table 2: Entity Data - Attributes

Any number of attributes can be related to an entity, depending on your specific needs. The data is intended to improve the UEBA analytics, so the data is generally limited to only what is necessary. The following common examples of attributes is by no means meant to be exhaustive.

Name	Type
Salary	Integer
Department	String
Office Location	String
Supervisor	String
Supervisor's Supervisor	String
Business Unit	String
Start Date	Date
Job Title	String
Number of Reports	Integer

EVENTS

A communication (email, chat, text, etc.) or action (printing, file copy, door access, etc.) performed by an entity. Events represent the largest amount of data in the system. Event types vary depending on your needs, and it's even possible that the same data type (like email) may vary from customer to customer. Because of all the variations, an exact definition of event data is not possible, but some consistencies can be identified.

Table 1: Event Data - Types

The following examples of event types are common:

- Communications
  - Email
  - Chat (e.g., Jabber, Slack, Skype, Lync, Symphony, Bloomberg)
  - Phone (audio files and/or transcripts)
- Activity
  - Printing
  - Web requests
  - Data movement (file copies, deletions, etc.)

- Authentication

Table 2: Event Data – Generic Structure

All the events we receive can be fit into this generic structure.

Name	Requirement	Purpose
Type	Mandatory	Defines the event type (e.g., email, chat, or print)
Roles	Mandatory	Generic way of defining which entities are involved in an event. In an email, sample roles would be sender, recipients, CC, or BCC. For a file copy event, the user account performing the copy is the relevant role. Roles are stored two ways: the raw identifier (email address, IP address, chat handle, etc.) and the resolved identifier (the entity that owns the email address/handle/etc.), if available.
IngestDate	Mandatory	A system date that defines when the event was ingested into Forcepoint UEBA
Timestamp	Mandatory	The date/time when the event occurred
Attachments	Optional	If the event has any attachments (e.g., email or chat room), the filename, file size, and file type are recorded. A file with extractable text is also stored here, so that users can search on it. The actual attachments (PDF, zip, doc, etc.) are stored in a separate location.
Body	Optional	Most communication events have a body (e.g., email, chat, text message, etc.), but not all event types do.
Subject	Optional	Most communication events have a subject (e.g., email, chat, etc.), but not all event types do.
Attributes	Optional	A collection of key/value pairs of meta data taken directly from the event. A Data Movement event, for instance, may have attributes like: operation (copy, delete, move, rename), file size, source folder, destination folder, and file type.

Meta data can be applied to an event, but it does not alter the original data in any way and would not generally contain any personally sensitive information. The following fields can be added to an event.

Name	Requirement	Purpose
AllEntityNamesResolved	Mandatory	A collection of all the entity names already mentioned in the Roles object
AllEntityNamesResolvedCount	Mandatory	A count of the entities in AllEntityNamesResolved
AllEntityNamesRaw	Mandatory	A collection of all the entity addresses already mentioned in the Roles object
AllEntityNamesRawCount	Mandatory	A count of the entities in AllEntityNamesRaw
Features	Optional	Applied by our “Feature” analytics
AttachmentCount	Optional	A count of the number of attachments in this event