FORCEPOINT
POWERED BY Raytheon

Forcepoint Behavioral Analytics

Management of Personal Data

Forcepoint Behavioral Analytics– Management of Personal Data

CONTENTS

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 Behavioral Analytics product?” It is primarily intended for those involved in the procurement and privacy assessment of the Forcepoint Behavioral Analytics 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

Personal Data

Forcepoint Behavioral Analytics is designed to provide users with configuration options that will support their efforts to ensure their use of Forcepoint Behavioral Analytics is GDPR compliant. Forcepoint’s enterprise customers, not Forcepoint, are the data controller for any personal data processed by the SW. Moreover, Forcepoint is not a data Processor with respect to the enterprise customer’s Personal Data Processing in its on-premises implementation of the Forcepoint Behavioral Analytics. For purposes of this document, the terms “Controller”, “Processor”, and “Processing” have the meanings set forth in the GDPR.

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.

The data that is ingested into Forcepoint Behavioral Analytics is configured by the customer, and thus the personal nature of the data depends on what the customer sends into the system. In the “What Data is Used?” column below, the data used by each component will depend on what the customer chooses to input into the application.

Safeguarding Personal Data

Forcepoint uses industry-standard techniques to protect data held within the Forcepoint product, including personal data. This approach to data security ensures that the high-value data is unintelligible to any person who is not authorised to access it.

Personal Data

Data Set	What Data is Used?	Purpose	Is Anonymization Possible?	Storage, Flow & Protection	Retention
Ingest Pipeline	Event Data Events vary by customer, but typically can be e-mail, chats, phone calls, authentication requests, print activity, web requests, etc. Entity Data Identification information for customers.	Consumes events that are provided by the customer. The end result is that the data is converted into our common format and pushed into an Elasticsearch datastore to be used by the rest of the application stack.	Partially. We are able to pseudonymize using aliases for entities involved in the event (i.e. sender, recipients, chat room participants, etc) from certain users. This feature can be turned on and off by the customer, where the identity of the entity is hidden from the analyst. Unstructured event data (email bodies, for example) are not pseudonymized as the entities are not parsed. These entity values cannot be pseudonymized as they are required to run the analytics.	Data Protection Summary Here is a summary of the base security controls for the Forcepoint Behavioral Analytics Appliance / AWS installation for data at rest and any external communications. Physical Protection for Appliance & AWS - Some customers on CentOS 6 (EOS by Forcepoint Behavioral Analytics 1 yr ago), but includes CentOS 7 by default - Volume Encryption - product supports volume encryption on disk, filestore; default is disabled, but customer can configure. For additional data protection, customers should follow the Hardening Guide's recommendations. - Native/Attachment files are stored in either cloud storage (with optional inclusion of encryption techniques (S3)) or network file share (customer can optionally encrypt these volumes) - Inherit basic protections from AWS	In steady state: The data is persisted temporarily in a work queue and then removed as the data is processed. The data is persisted as it is indexed in the Elasticsearch data store. In error state: By default, errors sit in an error queue and will be retried. If the event continues to fail, it will be published to an Elasticsearch data store in a Monitoring environment where all application metrics are recorded within the application stack.

				environment
				Access Control
				- SSH Access for system administration - Enabled by default, Professional Services sets up using customer-provided IPs, Professional Services does not maintain separate access
				- APIs
				  ○ SSL/TLS certifications required to make API requests;
				  ○ Use client-side certifications for authentication, Data always encrypted in transit, using TLS
				  ○ At installation, or runtime certificates can be created as self signed or inherited from the customer and deployed per box. The certificates are managed with HashiCorp	

				○ Vault, which is installed in the tool ○ Note - for POCs, using bastion host/jumpbox in AWS VPC; requires accessing SSH per NAT	
Ingest Pipeline Logs	Logs of Event Data This may include metadata associated with e-mail, chats, phone calls, authentication requests, print activity, web requests, etc Logs of Entity Data Identification information for customers.	Error, app, and debug logs for the application. These logs are intended to be used by system administrators to assist in troubleshooting issues.		See Data Protection Summary above	Default - Rotation: yes - Rotation Freq.: Daily - Log Retention: 90 days - Log Location: /var/log/ro-qw These settings are configurable by the customer, limited only by the customer's storage space.
Entity Management Service	Name Identifiers	Stores and manages referential information regarding an entity (i.e. Chris Lenoir), all their identifiers (i.e. work email, personal email, Skype handle, Bloomberg handle, IP address, phone number, etc) as well as all their attributes (i.e. salary, supervisor, office location, etc). This information is provided by the customer and is generally relevant to their use case.	The service stores both the pseudonymized version of the entity (e.g. Blue Garden Bird) and also the real identity (e.g. Chris Lenoir) so that the UI can allow super users to see who actually sent an event. The pseudonymization feature enables or disables the analyst's ability to see unmasked, or real, data. The purpose of this is to hide the real identity of the entity until an analyst has deemed an entity of interest.	See Data Protection Summary above Data is stored within Postgres or Elasticsearch and not directly on disk by this service.	Entity attributes and identifiers tend to be persistent for the course of the installation as they are referenced by new events as ingested. - There is no automated entity removal process. Entity information can be purged with a manual process. - Stored in Postgres and replicated to Elasticsearch
Entity Management Service Logs	Logs Records of when entity service record is updated with new/deleted information	Error, app, and debug logs for the application. These logs are intended to be used by system administrators to assist in troubleshooting issues.		See Data Protection Summary above	Default - Rotation: yes - Rotation Freq.: Daily - Log Retention: 90 days - Log Location: /var/log/ro-rose

					- Logs also replicated to Elasticsearch Monitoring instance
Entity Publisher Service	Entity and risk data	Retrieves entity information from Elasticsearch, publishes to DLP and uses to enrich events as they are ingested	No, the incoming data is pseudonymized by default. Entities are ingested with an ID from Forcepoint DLP that is already considered pseudonymized. That pseudonymized ID is used to publish risk level back to Forcepoint DLP. If a customer changes that from the default, then a non-pseudonymized entity identifier, such as email, could be published.	See Data Protection Summary above	This temporary storage is only used while it is being processed.
Entity Publisher Service Logs	Logs	Error, app, and debug logs for the application. These logs are intended to be used by system administrators to assist in troubleshooting issues.		See Data Protection Summary above	Default - Rotation: yes - Rotation Freq.: Daily - Log Retention: 90 days - Log Location: /var/log/ro-ups
Master Data Service	Event / Element Data	This sits in front of Elasicsearch and brokers requests from other services (the User Interface, entity service, etc) for information regarding user events. It provides additional analytic processes, can trigger back-end jobs to run against the Elasticsearch data, and can manipulate the event data being returned – such as restricting what data the UI receives, based on who is requesting it.	Partially. This is part of the entity alias pseudonymization feature. We are able to pseudonymize the entities involved in the event (i.e. sender, recipients, chat room participants, etc) but we are not able to pseudonymize the body of those events. However, the Master Data Service and the UI have mechanisms to prevent un-authorized users from seeing those bodies. Only authorized analysts are able to see real entity identities.	See Data Protection Summary above	Typical event horizons are six months to one year before data is rolled off. - Data stored in Postgres DB - Uses Elasticsearch - for querying data, applying analytics to the data, applying meta information to the data (but not altering the original data) - No automated retention - Customer configurable by writing cron jobs to clean up data

Master Data Service Logs	Logs	Error, app, and debug logs for the application. These logs are intended to be used by system administrators to assist in troubleshooting issues.		See Data Protection Summary above	Default - Rotation: yes - Rotation Freq.: Daily - Log Retention: 15 days - Log Location: /var/log/ro-mds - Logs replicated to Elasticsearch Monitoring instance. These settings are independent of the mds log settings. See below. These settings are configurable by the customer, limited only by the customer's storage space.

UI Services	Viewing User Data
	
	- Displays Event Data
	- Displays Element Data
	- Displays Entity Data
	- User Preferences
	- User Settings
	- Saved search queries
		This is the primary user interface (web app) that customers use to interact with their data.	Partially. We are able to pseudonymize the entities involved in the event (i.e. sender, recipients, chat room participants, etc) but we are not able to pseudonymize the body of those events. However, the Master Data Service and the UI have mechanisms to prevent un-authorized users from seeing those bodies.	See Data Protection Summary above
				
				Logical Access Controls for analysts
				- No MFA support beyond SAML requirements
				- Login - SSO
				  ○ LDAP PKI or SAML
				- Login - UN/Pass
				  ○ Password complexity enforced
				  ○ No password rotation (could be enforced via Professional Services)
				  ○ Passwords encrypted and stored in Postgres DB (default algorithm bcrypt, SHA-512 with salt also available)
				  ○ HTTPS
					- Stored in Postgres
					- Uses Elasticsearch

				○ Authorization Roles that are granted to users by the customer, which define the abilities of those users, such as what data can be seen ○ Entitlements / SafeSearch limits what data is within the charts/graphs based on customer configurations ▪ e.g. cannot see your manager's salary data	
UI Services Logs	Logs - Specific Logs / Reports purpose built for customer compliance needs. Could include: ○ sender ○ recipient ○ subject ○ timestamp ○ for email, chat, text events	Audit, request, error and debug logs for the application. These logs are intended to be used by system administrators to assist in troubleshooting issues.	No, it is not practical as actual log data is required for assessing system performance.	See Data Protection Summary above	Default - Rotation: yes - Rotation Freq.: Daily - Log Retention: 30 days - Log Location: /var/log/ro-ui - Logs replicated to Elasticsearch Monitoring instance
Monitoring Service	Logs from various services are replicated here. - MDS Logs - Entity service Logs	This environment gives the customer visibility into what their Forcepoint Behavioral Analytics stack is doing. Both at a hardware level (CPU, memory, etc) but	No, it is not practical as actual log data is required for assessing system performance.	See Data Protection Summary above	Data is stored in Elasticsearch. Retention is based on customer needs, and records can be manually deleted. By default, retention is set to 90 days.

	- Ingest Pipeline Logs - UI Service Logs	more critically it also is where logs from the various services are replicated. The purpose of this is to assist system administrators and analysts with troubleshooting.			There is no automated purging of data. Customers should follow the Hardening Guide's recommendations to further protect personal data.

Appendix A

ENTITIES

A person, place or thing of which a pattern of behaviour can be gleaned. Most customers use their employees as the entities, but they could just as easily use an office, a building or a computer as an entity, for instance. Every event (email, chat, printer job, etc) has an entity associated with it. The entity can be someone the customer wishes to track closely, or it could be someone they either don't know or have no desire to know (like the sender of bulk email).

Table 1: Entity Data – Common

Every entity has this data associated with it. The information comes either from the customer, or in the case of unknown entities

Attribute	Requirement	Purpose
Entity Display Name	Mandatory	The name of the entity (i.e. Chris Lenoir)
Aliases	Mandatory	All the ways we can identify this entity (email addresses, phone numbers, IP addresses, chat handles, etc)
Attributes	Optional	See Table 2

Table 2: Entity Data - Attributes

Customers may provide us with any number of attributes related to an entity. The data is intended to improve the Forcepoint Behavioral Analytics analytics, so the data is generally limited to only what's necessary. Below are some common examples of attributes, and by no means is this list meant to be exhaustive.

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

EVENTS

A communication (email, chat, text, etc) or action (printing, file copy, door access, etc) performed by an entity. Events are provided by the customer and represent the largest amount of data in our system. Event types vary from customer to customer, and it's even possible for the same data type (like email) to vary between customers. Because of all the variations, it's not possible to provide an exact definition of our event data, but there are some consistencies that can be called out.

Table 1: Event Data - Types

Here are some examples of event types we see from customers.

- Communications
  - E-mail
  - Chat (Jabber, Slack, Skype, Lync, Symphony, Bloomberg, etc)
  - Phone (can include 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	To define what the event is – e-mail, chat, print, etc
Roles	Mandatory	A role is a generic way of defining what entities are involved in the event. On an email, this would be sender, recipients, CC, BCC, etc. On a file copy, there is only the user account performing the copy. There are two ways Roles are stored: the raw identifier (e-mail address, IP address, chat handle, etc) and the resolved identifier (which entity owns that address/handle/etc) if available.
IngestDate	Mandatory	A system date defining when the event was ingested into Forcepoint Behavioral Analytics
Timestamp	Mandatory	The date/time when this event occurred
Attachments	Optional	If the event has any attachments (like in an email or chat room), the filename, file size and file type are recorded. If the file has extractable text, it 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 (like an email, chat, text message, etc), but not all event types do.
Subject	Optional	Most communication events have a subject (like an email, chat, etc), but not all event types do.
Attributes	Optional	This is a collection of key/value pairs of meta information 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, file type.

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

Name	Requirement	Purpose
AllEntityNamesResolved	Mandatory	Simply a collection of all the entity names already mentioned in the Roles object.
AllEntityNamesResolvedCount	Mandatory	A count of the entities in AllEntityNamesResolved
AllEntityNamesRaw	Mandatory	Simply 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 on this event