Forcepoint

Data Loss Prevention Management of Personal Data

Table of Contents

Disclaimer 3

General 3

DLP Custom Dictionaries 4

DLP Incidents 5

Structured DLP Fingerprinting 7

Unstructured DLP Fingerprinting 8

Incident Forensics 10

Active Directory Data / Resource Resolver 12

Appendix A 13

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 Forcepoint Data Loss Prevention?" It is primarily intended for those involved in the procurement and privacy assessment of Forcepoint Data Loss Prevention.

Privacy Laws

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, along with other applicable data privacy laws, guide the principles that are incorporated in 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.

Forcepoint DLP is designed to comply with applicable data privacy principles, including those contained in GDPR. Consistent with these principles, Forcepoint's customers are considered to be the sole data controller. Forcepoint is the data processor with respect to customer data transferred through or stored in Forcepoint DLP.

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. Full details on Forcepoint's privacy policy and processes can be found at: https://www.forcepoint.com/legal/forcepoint-trust-hub

DLP Custom Dictionaries

Data Set	What Personal Data is Used?	Purpose	Is Pseudonymization Possible?	Storage, Flow & Protection	Retention
Any content that the administrator added to the dictionary manually or by importing a CSV	The custom dictionary may hold any data that the DLP admin added to it. The intended usage of dictionaries is for short phrases (i.e., numbers, single words or few words). Note: It is not the intended usage of dictionaries to hold PII, and there are other methods of protecting PII (i.e., Fingerprinting). However, customers may still use dictionaries for this purpose.	The custom dictionary contains data that is considered sensitive by the DLP admin and is intended to be matched by the DLP policy engine and blocked/alerted	The device information is not pseudo anonymized presently as noted in "what personal data is used?" It is not intended to hold PII.	The dictionary is created on-premises and uploaded to DLP agents via SOAP-XML over a secure (TLS 1.2) connection. The dictionary entries are saved in the customer's SQL server. It is the customer's responsibility to keep that server secure and limit the access to the SQL DB. A reflection of the dictionary is sent to all DLP agents and endpoints as an encrypted file. The data of the dictionary can be accessed via the DLP UI only by administrators with specific policy permissions. Dictionaries are uploaded to DLP agents over a secure (TLS 1.2) connection and downloaded by endpoints also via TLS 1.2 from endpoint servers, as an encrypted file.	Dictionary data is stored in the SQL database and accessible though the DLP UI until the customer deletes the information. Deletion of custom dictionaries entries or the full dictionary can be performed via the DLP UI by a DLP admin with the correct permissions

How to Manage Subject Access Request (SAR)

SAR - Right to Access	Only an administrator with specific permissions can access the data in the dictionary. An end user wishing to access this information can always do so with an administrator.
SAR - Correction/Rectification	Correcting the information in the dictionary is done by the customer admin editing the dictionary entries. Upon saving a corrected/modified policy the admin can click "Deploy policy" button to publish the policy to the DLP agents.
SAR - Right to be Forgotten	A manual deletion of a dictionary entry via the DLP UI and the policy deployment to all DLP agents and endpoints, will result in the dictionary entry data to be completely removed from the system.
Data Storage / Localization	The storage of the information in the custom dictionary is on the customer's SQL DB. Copies of this data are stored in an encrypted file on all DLP agents and endpoints.

DLP Incidents

Data Set	What Personal Data is Used?	Purpose	Is Pseudonymization Possible?	Storage, Flow & Protection	Retention
Any event captured by a DLP agent or endpoint due to data being matched by a DLP rule.	DLP rules can match a variety of data types, including PII. The data that is matched depends on the policy configuration. PII in an incident can be in the details of the matched data in case the rule is looking for PII. The incident also includes information on the source of the incident, and that information may also reflect PII (i.e., username, email).	The incidents are stored for a security admin to be able to investigation intentional or accidental data leakage.	The usernames/incident sources in the database are not pseudo anonymized. Only admins with incident manager role can view the data. In the FSM UI, the source (username) is pseudo anonymized, and it is possible to configure for specific admin roles to view only the pseudo anonymized usernames, while leaving the full data only to other specific admins. Some data types such as credit card numbers are also masked, therefore an administrator that is allows to view violations but not allowed to view forensics will only view masked data. For example: a credit card 4111-1111-1111-1111 will be visible as 4111-xxxx-xxxx-1111.	Incidents are created by DLP agents and endpoints upon matching DLP rules. The information of the incident is represented as an XML file which is sent to the DLP manager. In the DLP manager it is processed and inserted into the SQL DB. Certain incident details are not originally created by the policy engine, those are pulled from the Active Directory per user according to the incident source. For example, contact details of the user who is the source of the incident. Incidents are saved in the customer's SQL server. It is the customer's responsibility to keep that server secure and limit the access to the SQL DB. Incidents are stored on DLP agents only until they are sent to the FSM. In a normal connected DLP agent this should be a matter of milliseconds to seconds. In the rare case of network disconnection for a long time, these incidents are cleaned up periodically.	Incidents are deleted after 12 partitions are full (by default this is 3 years). It is possible to change the default retention of a single partition from 90 days to a lower number of days. The maximum retention time will change accordingly (i.e.,30 days per partition will translate to 12 months maximum retention time). An incident can be deleted by a DLP admin with the correct permissions. An archived partition that holds 90 days' worth of incidents (90 days by default – configurable) can also be fully deleted by a DLP with the correct permissions. Deleting an incident is audited.

How to Manage Subject Access Request (SAR)

SAR - Right to Access	Only an administrator with specific permissions can access incidents. An end user wishing to access this information can only do so together with such an administrator.
SAR - Correction/Rectification	Information in incidents can’t be modified. The meta data part of the incident is reflected from the customer’s active directory, and therefore any modifications need to be performed there. Matched information can’t be modified as it reflects the event as it happened. If the information in the incident is not correct the incident can be marked as “false positive” or “false negative.”
SAR - Right to be Forgotten	*A manual deletion of a dictionary entry via the DLP UI and the policy deployment to all DLP agents and endpoints, will result in the dictionary entry data to be completely removed from the system.*
Data Storage / Localization	*The storage of the information in the custom dictionary is on the customer’s SQL DB. Copies of this data are stored in an encrypted file on all DLP agents and endpoints.*

Structured DLP Fingerprinting

Data Set	What Personal Data is Used?	Purpose	Is Pseudonymization Possible?	Storage, Flow & Protection	Retention
Data structures representing any type of tables such as CSV or databases.	The fingerprinting structures reflect any database table data that the customer sees as sensitive and needs protecting.	The fingerprinting data structures are used represent the protected data. That representation is used to identify sensitive data in emails/HTTP/other data in an efficient manner.	The data is hashed and saved only in the hashed format	The crawler component retrieves data from databases. Depending on the source of the data, the crawler processes those into a canonical format (converting to lower case, reducing white spaces). The crawler then passes pieces of data through a one-way hash function which are stored in the fingerprint repository. All data that resides in the FPR has passed through a one-way hash function and is stored in the FPR as such.	The data that the fingerprinting classifier represents is stored in the FPR (fingerprint repository) and will be deleted when one of the following occurs: 1. The FPR is manually reset by the DLP administrator 2. The fingerprinting classifier is deleted 3. The data is deleted from the original database and the fingerprinting job is run again

How to Manage Subject Access Request (SAR)

SAR - Right to Access	There is no direct way to access the information from DLP as it is only saved after it was hashed. It is possible to request access to the customer database table that was fingerprinted.
SAR - Correction/Rectification	The customer administrator must change the data at the source and then rerun the fingerprinting job.
SAR - Right to be Forgotten	The customer administrator will need to remove the relevant database entry and rerun the fingerprinting job
Data Storage / Localization	The data is stored in the FPR which a proprietary data structure and is represented as hashes.

Unstructured DLP Fingerprinting

Data Set	What Personal Data is Used?	Purpose	Is Pseudonymization Possible?	Storage, Flow & Protection	Retention
Data structures representing the files with sensitive data that the customer would like to protect	The fingerprinting structures reflect any files that the customer sees as sensitive and needs protecting. Those may or may not include PII.	The fingerprinting data structures are used represent the protected data. That representation is used to identify sensitive data in emails/HTTP/other data in an efficient manner.	As the PII is mixed with other words and then hashed then it is completely obscure and can't be retrieved from the FPR	The crawler component retrieves data from file systems. Depending on the source of the data, the crawler processes those into a canonical format (converting to lower case, reducing white spaces, removing stop words). The crawler then passes short phrases (few words) through a one-way hash function which are stored in the fingerprint repository. As the FPR does not store single words rather than only combinations reflecting sentences, there is no PII stored in the FPR and no PII can be retrieved from the FPR. All data that resides in the FPR has passed through a one-way hash function and is stored in the FPR as such	The data that the fingerprinting classifier represents is stored in the FPR (fingerprint repository) and will be deleted when one of the following occurs: 1. The FPR is manually reset by the DLP administrator

How to Manage Subject Access Request (SAR)

SAR - Right to Access	As any possible PII information is hashed together with other words and reflects content of a fingerprinted file, it is not possible for the admin or any other entity to retrieve that specific PII data.
SAR - Correction/Rectification	As any possible PII information is hashed together with other words and reflects content of a fingerprinted file, it is not possible for the admin or any other entity to retrieve that specific PII data, there is nothing to correct.
SAR - Right to be Forgotten	As any possible PII information is hashed together with other words and reflects content of a fingerprinted file, it is not possible for the admin or any other entity to retrieve that specific PII data, there is nothing to correct.
Data Storage / Localization	The data is stored in the FPR which a proprietary data structure and is represented as hashes.

Incident Forensics

Data Set	What Personal Data is Used?	Purpose	Is Pseudonymization Possible?	Storage, Flow & Protection	Retention
The forensics data related to incident captured by a DLP agent or endpoint due to data being matched by a DLP rule. Forensics data is kept only for networking or mobile incidents but not for discovery incidents.	Forensics can be a copy of a real file/email/key phrase of the user that was matched by a DLP rule. This forensics file can include personal data of the user and reveal the source and destination.	The forensics data is the evidence for a data breach. When a security admin needs to investigate an incident, the forensics is one of his tools and evidence even for court	*While incident data is synonymized the forensics data must stay in the original state for the following reasons: the forensics file is needed as evidence and must keep its original value. Viewing the forensics file is protected by an admin role as configured by the customer admin.*	*Incident Forensics files are created by DLP agents and endpoints together with the incident upon matching DLP rules. The incident forensics is created only if requested in action plan configuration. The incident forensics is sent to the DLP manager after the incident is sent to the DLP manager. The incident forensics is connected to the incident using the event id. The incident forensics is kept as encrypted file on the forensics repository.* *Incidents are saved encrypted to the forensics repository disk.* *The customer has its own key and managed only by the customer.* *The encryption key is AES 256 and backed up and encrypted in the FSM PostgreSQL DB which is open to local host only. The key is also kept in the forensics repository control record which is also encrypted. It is the customer's responsibility to keep the server of the*	*Incidents forensics are kept on the forensics repository until archived or deleted.* *Incident forensics are deleted in the following cases:* 1. *The incident is deleted. (After 3 years when incident is deleted, its forensics file will also be deleted)* 2. *The incident status is changed to close, and it is configured by the admin to delete forensics file for a closed incident* 3. *No disk space in archive disk – oldest forensics files are deleted* 4. *Manually delete partition by an admin of the DLP manager.* *Incident forensics are archived in the following cases:* 1. *No disk space for new forensics – 15% of oldest forensics files are automatically archived (15% is configurable value)* 2. *Manually archived by an admin of the DLP manager.*

				forensics repository secure and limit the access to it. *Incidents forensics are stored on DLP agents only until they are sent to the FSM. In a normal connected DLP agent this should be a matter of milliseconds to seconds. In the rare case of network disconnection for a long time, these incidents are cleaned up periodically. When network connection restored – incident forensics are sent immediately to the FSM.*	

How to Manage Subject Access Request (SAR)

SAR - Right to Access	*End users can access data with FSM administrator support.*
SAR - Correction/Rectification	Incidents can be marked as "false positive" or closed but incidents forensics cannot be modified since it precisely reflects what occurred on the network.
SAR - Right to be Forgotten	*A manual deletion of a forensics via the with the FSM application with the FSM administrator support.*
Data Storage / Localization	*The data is stored within the customer own and managed server/storage. The storage of the forensics on disk encoded and encrypted.*

Active Directory Data / Resource Resolver

Data Set	What Personal Data is Used?	Purpose	Is Pseudonymization Possible?	Storage, Flow & Protection	Retention
The customer's Active Directory	The RR includes names, emails, list of users in dist lists and other data available in the active directory.	DLP needs to be able to enforce rules on specific groups or users. Those users need to be identified as their data traverses through DLP agents.	*The policy engine must be able to recognize data transactions (such as email) according to personal data (email address or login name), so this data is not anonymized rather than stored per resource together with an identifier. This identifier is used wherever possible and translated by the resource resolver.*	*The active directory data is imported as is and stored in the SQL DB.* *The resource resolver process translates the SQL data into an xml and analyses which user belong to which active directory groups.*	*Every import from the active directory, overrides the previous data.* *Manual entries are kept until deleted manually.*

How to Manage Subject Access Request (SAR)

SAR - Right to Access	*A DLP administrator can search the DLP resources for a specific user.*
SAR - Correction/Rectification	An entry imported from the active directory needs to be edited in the active directory. Following the next daily import of data, it will be updated in the DLP database. A manual entry can be edited by an administrator deleting it from the resources table
SAR - Right to be Forgotten	An entry imported from the active directory needs to be deleted from the active directory. Following the next daily import of data, it will be updated in the DLP database. A manual entry can be removed by an administrator deleting it from the resources table
Data Storage / Localization	*SQL DB* *XML* *Only on customer's servers, at customer's discretion*

Appendix A

TERMINOLOGY

Term	Explanation
Event	An event is any transaction that traverses the Forcepoint DLP system. Not all events are stopped by the Forcepoint DLP sniffer and queued for analysis—for that to happen, something must look suspicious, meaning that something in the event seems to match with a Policy rule. • **Unmatched events** are events that pass through the system transparently, because they raise no suspicion. • **Policy matches** are events that are analyzed as they traverse the system, because something in the transaction is suspicious according to the policies. Policy matches are then either deemed **authorized incidents**—events that seemed to match a policy but are in fact allowed—or **incidents**, which are policy violations.
Incident	An incident is a transaction or set of transactions that violate a policy. Depending on how you configure a rule, incidents can be created for every policy breach, or for matches that occur within a defined period. Assigned/Unassigned Incident: Incidents can be tracked through the system by administrators. To give a single administrator the responsibility to handle the incident, assign the incident to a single administrator. Unassigned Incidents are those that have not been assigned and can therefore be handled by any administrator who has access to the incident.
Forcepoint DLP Administrator	A user who manages and maintains the Forcepoint DLP system.
Policy	The system can be set to include multiple policies. A policy is a list of criteria to be searched for over your channels. These criteria are set with a certain rule which defines what the system does when it comes across a transmission that meets the designated criteria.
Structured Fingerprinting	The process of identifying a unique set of characteristics for any type of tables such as CSV or databases.
Unstructured Fingerprinting	The process of identifying a unique set of characteristics for file or document contents.
User	The personnel within an organization who can distribute and receive information.