

Statistics

- INTRODUCTION
- THE DATA SET
- DATA COLLECTION METHODS
 - FROM SYSLOG STREAM
 - GOOGLE ANALYTICS TRACKER
 - CONTAINER DATA
 - SIGNATURE(S) DATA
 - EXAMPLE PAYLOAD
- REFERENCES

INTRODUCTION

A statistical data set is composed after every container validation.

Currently, there are two means to collect and process these data sets. One way is to extract this data set as a JSON object from the Syslog stream produced by Siva and forward it to a provider-specific central log processing solution (like ELK, Splunk, etc). Another way is to use the embedded Google Analytics tracker (disabled by default) and let Siva push the data set itself to Google Analytics (hereinafter GA).

THE DATA SET

For every container, the following data set will be logged by Siva for generic business statistics needs.

	Data	Cardinality	Description
1.	Validation duration	1	The time it takes to process an incoming request - measured in milliseconds
2.	Container type	1	Container type (text value that identifies the signature type of the incoming document: ASiC-E, XAdES, PAdES or ASiC-E (BatchSignature))
3.	Siva User ID	1	String (Text data that contains the SiVa user identifier for reports (from the HTTP x-authenticated-user header) or "N/A")
4.	Total signatures count	1	The value of the "signaturesCount" element in the validation report
5.	Valid signatures count	1	The value of the "validSignaturesCount" element in the validation report
6.	Signature validation indication(s)	1...*	Values of elements signatures/indication and signatures/subindication from the validation report. <indication[/subindication]>
7.	Signature country	1...*	Country code extracted from the signer cert. The ISO-3166-1 alpha-2 country code that is associated with signature (the signing certificate). Or constant string "XX" if the country cannot be determined.
8.	Signature format	1...*	Values of element signatures/signatureFormat from the validation report. <signatureFormat>

DATA COLLECTION METHODS

There are two methods to collect and analyze the container/signature data within Siva.

FROM SYSLOG STREAM

One way to collect statistical dataset's from Siva is to process Syslog data streams that web service nodes produce (among other technical monitoring data). For example, one could use Logstash to forward relevant statistical information from Syslog log files to a log index cluster like ElasticSearch to be visualized later by reporting tools like Kibana.

Statistical data set is sent over Syslog protocol in the message field as a JSON object with following properties:

Property	Description	Type
stats		
.type	Container type	String (one of the following constants "ASiC-E TS", "ASiC-E TM", "XAdES", "PAdES", "ASiC-E (BatchSignature)")

.usrId	Siva User ID	String (Text data that contains the SiVa user identifier for reports (from the HTTP x-authenticated-user header) or "N/A")
.dur	Validation duration	Number
.sigCt	Total signatures count	Number
.vSigCt	Valid signatures count	Number
.sigRslt [1 ...n]		
.i	Indication	String
.si	Subindication	String
.cc	Country code	String
.sf	Signature format	String

Example JSON message:

```

{
  "stats": {
    "type": "PADES",
    "usrId": "sample_user1",
    "dur": 4021,
    "sigCt": 2,
    "vSigCt": 1,
    "sigRslt": [
      {"i": "TOTAL-PASSED", "cc": "EE", "sf": "PADES_BASELINE_LT"},
      {"i": "INDETERMINATE", "si": "NO_CERTIFICATE_CHAIN_FOUND",
"cc": "EE", "sf": "PADES_BASELINE_LT"}
    ]
  }
}

```

GOOGLE ANALYTICS TRACKER

Another method, that does not require setting up infrastructure on service provider side, is to use Siva's built-in GA tracker to collect and analyze service statistics. The tracker is embedded on the server-side of Siva web service and is disabled by default. When enabled and configured with proper GA account, the statistics tracker component will capture validation events and forward statistical event related data to GA. Data is sent after the validation report has been composed but before it is returned to the user.

Data is sent to GA using the **Google Measurement Protocol API [GMP]** using batch protocol.

When SiVa sends any data to Google Analytics it's called sending a hit. All events in SiVa are communicated to GA as so call Hits. Every hit must have a predetermined hit type to fit into the GA reporting model. A hit is described with a predetermined list of key and value pairs (some of which are mandatory or optional depending on the hit type).

SiVa uses GA **event** types to fit the validation process results into the GA data model. For every container, the data set will be sent as GA events (for detailed event data see next section CONTAINER DATA and SIGNATURE(S) DATA):

CONTAINER DATA

Following key and value pairs are forwarded to GA about the validation report:

Parameter Name	Key	Value	Description
----------------	-----	-------	-------------

Protocol version	v	1	Constant value
Hit type	t	"event"	Constant value
Tracking ID	tid	UA-XXXXX-Y	Siva operator specific value (has to be created and obtained from the SiVa operators GA account)
Client ID	cid	555 or <random UUID>	This anonymously identifies a particular user. The value of this field should be a random UUID (version 4) as described in http://www.ietf.org/rfc/rfc4122.txt or 555 for anonymous user.
Data source	ds	"SiVa"	Text data used to distinguish different sources of data. Initially, only SiVa service will be used as the source for GA data.
Siva User ID	cd1	<Siva User Identifier> or "N/A"	Text data that contains the SiVa user identifier for reports (from the HTTP x-authenticated-user header). Although GA provides it's own user distinction mechanism in the Measurement Protocol (CID – client id), it cannot be used in custom reports. NB! Requires additional configuration in SiVa operators GA account NB! Since the value is taken from user provided authentication mechanism, it must be ensured that the x-authenticated-user header does not contain any information about the actual user
Event category	ec	<Signature container type>	A text value that identifies the signature type of the incoming document (ie ASiC-E, XAdES, PAdES, ASiC-E (BatchSignature))
Event action	ea	"Container%20validation"	A constant text value that describes the event action
Event label	el	?	A constant text value that describes the event operation. See the next table for details.
Event value	ev	?	An integer value that describes the event. See the next table for details.

Specific event labels and values are as follows:

Event label	Event value	Description
"validationDuration"	<Integer>	A time value that represents the time it took SiVa to complete the container validation and return the validation report (in milliseconds)
"signaturesCount"	<Integer>	The value of the "signaturesCount" element in the validation report
"validSignaturesCount"	<Integer>	The value of the "validSignaturesCount" element in the validation report

SIGNATURE(S) DATA

One event is sent for every signature.

Following key and value pairs are forwarded to GA about the validation report:

Parameter Name	Key	Value	Description
Protocol version	v	1	Constant value
Hit type	t	"event"	Constant value
Tracking ID	tid	UA-XXXXX-Y	Siva operator specific value (has to be created and obtained from the SiVa operators GA account)
Client ID	cid	<random UUID>	This anonymously identifies a particular user. The value of this field should be a random UUID (version 4) as described in http://www.ietf.org/rfc/rfc4122.txt
Data source	ds	"SiVa"	Text data used to distinguish different sources of data. Initially, only SiVa service will be used as the source for GA data.
Siva User ID	cd1	<Siva User Identifier> or "N/A"	Text data that contains the SiVa user identifier for reports (from the HTTP x-authenticated-user header). Although GA provides it's own user distinction mechanism in the Measurement Protocol (CID – client id), it cannot be used in custom reports. NB! Requires additional configuration in SiVa operators GA account NB! Since the value is taken from user provided authentication mechanism, it must be ensured that the x-authenticated-user header does not contain any information about the actual user

Event category	ec	<Signature container type/Signature format>	A text value that identifies the signature type of the incoming document by combining container type (ie ASiC-E, XAdES, PAdES, ASiC-E (BatchSignature)) and signature format taken from the validation report
Event action	ea	"Signature%20validation"	A constant text value that describes the event action
Event label	el	<indication[/subindication]>	Values of elements signatures/indication and signatures/subindication from the validation report
Geographical override	geoid	<country ISO code> or "XX"	The ISO-3166-1 alpha-2 country code that is associated with signature (the signing certificate). Or constant string "XX" if the country cannot be determined.

EXAMPLE PAYLOAD

```
POST http://www.google-analytics.com/batch HTTP/1.1
Accept-Encoding: gzip,deflate
Content-Type: text/xml
Content-Length: 1053
Host: www.google-analytics.com
Connection: Keep-Alive
User-Agent: Apache-HttpClient/4.1.1 (java 1.5)
v=1&t=event&tid=UA-80320579-4&cid=22a40135-3a4c-4267-76f0-a18701011301&d
s=siva&cdl=sample_user1&ec=ASiC-E%20(BatchSignature)&ea=Container%20vali
dation&el=duration&ev=388
v=1&t=event&tid=UA-80320579-4&cid=11a40235-1a3a-4267-6777-40720002e302&d
s=siva&cdl=sample_user1&ec=ASiC-E%20(BatchSignature)&ea=Container%20vali
dation&el=signaturesCount&ev=2
v=1&t=event&tid=UA-80320579-4&cid=33a40135-3341-4267-9fff-20872302b30d&d
s=siva&cdl=sample_user1&ec=ASiC-E%20(BatchSignature)&ea=Container%20vali
dation&el=validSignaturesCount&ev=1
v=1&t=event&tid=UA-80320579-4&cid=56a40a35-1a4c-4267-86f0-a4870002b30f&d
s=siva&cdl=sample_user1&geoid=EE&ec=ASiC-E%20(BatchSignature)/XAdES_BASE
LINE_LT&ea=Signature%20validation&el=TOTAL-PASSED
v=1&t=event&tid=UA-80320579-4&cid=ccfbc7d4-3ff7-4bf1-89e6-2b62a02120ec&d
s=siva&cdl=sample_user1&geoid=FI&ec=ASiC-E%20(BatchSignature)/XAdES_BASE
LINE_LT&ea=Signature%20validation&el=INDETERMINATE/NO_CERTIFICATE_CHAIN_
FOUND
```

REFERENCES

[GMP] Google Measurement Protocol V1
<https://developers.google.com/analytics/devguides/collection/protocol/v1/reference>

[events] Events in Google Analytics

<https://support.google.com/analytics/answer/1033068?hl=en>