Telstra Data Normalisation

Team Koala

Luke Hawkins

Luke
Joel
Ishan
Nikunj
Daniel

The Problem: Data Normalisation

Thousands of devices

Variety of formats

Security analysis requires data normalization

Must Normalize to standard schema

The Process: Data Normalisation

1

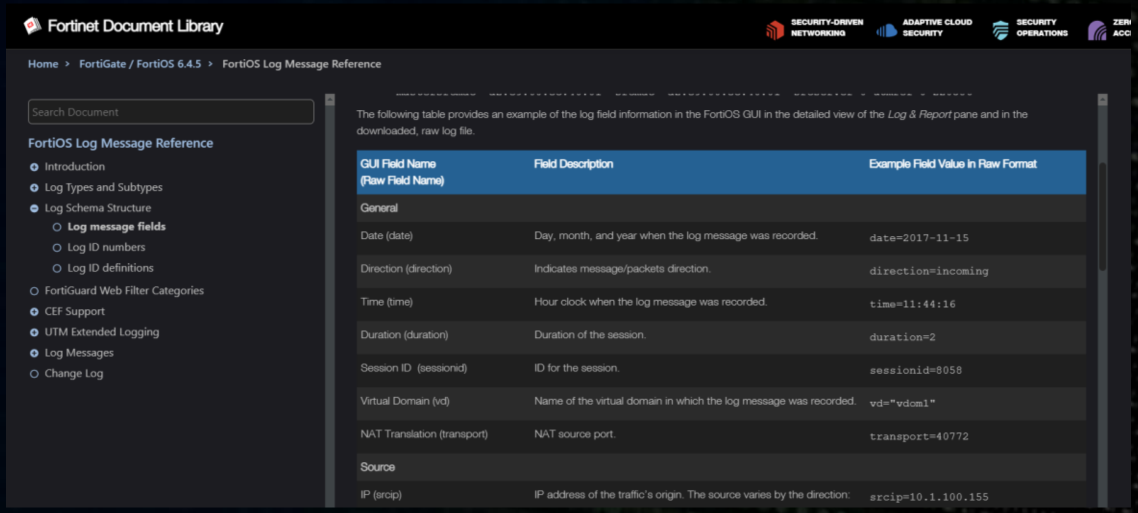
Constructing Mappings

2

Normalising Live Data

1. Current Constructing Mappings

Online Documentation



Fortinet Document Library

Home > FortiGate / FortiOS 6.4.5 > FortiOS Log Message Reference

Search Document

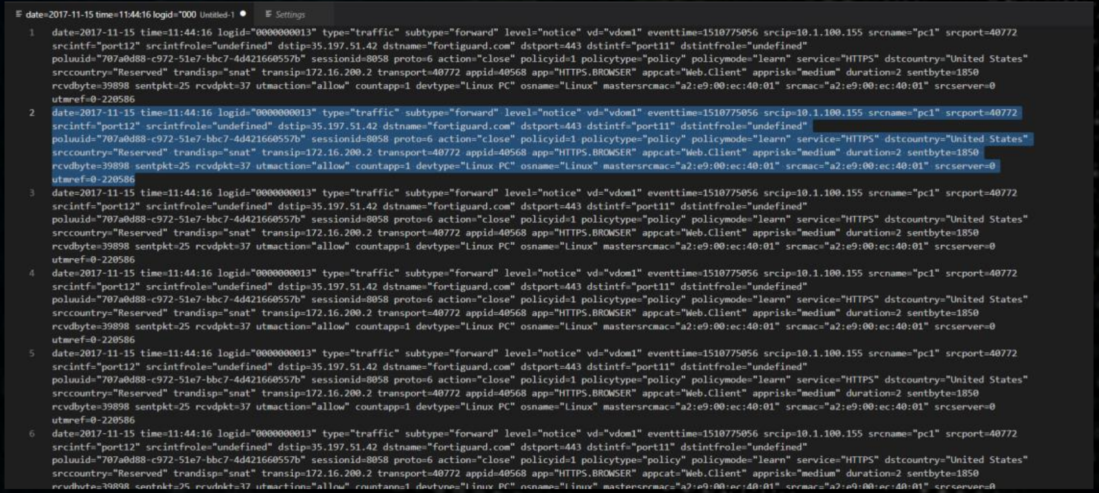
FortiOS Log Message Reference

- Introduction
- Log Types and Subtypes
- Log Schema Structure
  - **Log message fields**
  - Log ID numbers
  - Log ID definitions
- FortiGuard Web Filter Categories
- CEF Support
- UTM Extended Logging
- Log Messages
- Change Log

The following table provides an example of the log field information in the FortiOS GUI in the detailed view of the *Log & Report* pane and in the downloaded, raw log file.

GUI Field Name (Raw Field Name)	Field Description	Example Field Value in Raw Format
General		
Date (date)	Day, month, and year when the log message was recorded.	date=2017-11-15
Direction (direction)	Indicates message/packets direction.	direction=incoming
Time (time)	Hour clock when the log message was recorded.	time=11:44:16
Duration (duration)	Duration of the session.	duration=2
Session ID (sessionid)	ID for the session.	sessionid=8058
Virtual Domain (vd)	Name of the virtual domain in which the log message was recorded.	vd="vdom1"
NAT Translation (transport)	NAT source port.	transport=40772
Source		
IP (srcip)	IP address of the traffic's origin. The source varies by the direction:	srcip=10.1.100.155

Sample Log



date=2017-11-15 time=11:44:16 logid="000 Untitled-1 ● Settings

```
date=2017-11-15 time=11:44:16 logid="0000000013" type="traffic" subtype="forward" level="notice" vd="vdom1" eventtime=1510775056 srcip=10.1.100.155 srcname="pc1" srcport=40772
srcintf="port12" srcintfrole="undefined" dstip=35.197.51.42 dstname="fortiguard.com" dstport=443 dstintf="port11" dstintfrole="undefined"
poluuid="707a0d88-c972-51e7-bbc7-4d421660557b" sessionid=8058 proto=6 action="close" policyid=1 policytype="policy" policymode="learn" service="HTTPS" dstcountry="United States"
srccountry="Reserved" trandisp="snat" transip=172.16.200.2 transport=40772 appid=40568 app="HTTPS.BROWSER" appcat="Web.Client" apprisk="medium" duration=2 sentbyte=1850
rcvdbyte=39898 sentpkt=25 rcvdpkt=37 utmaction="allow" countapp=1 devtype="Linux PC" osname="Linux" mastersrcmac="a2:e9:00:ec:40:01" srcmac="a2:e9:00:ec:40:01" srcserver=0
utmref=0-220586
date=2017-11-15 time=11:44:16 logid="0000000013" type="traffic" subtype="forward" level="notice" vd="vdom1" eventtime=1510775056 srcip=10.1.100.155 srcname="pc1" srcport=40772
srcintf="port12" srcintfrole="undefined" dstip=35.197.51.42 dstname="fortiguard.com" dstport=443 dstintf="port11" dstintfrole="undefined"
poluuid="707a0d88-c972-51e7-bbc7-4d421660557b" sessionid=8058 proto=6 action="close" policyid=1 policytype="policy" policymode="learn" service="HTTPS" dstcountry="United States"
srccountry="Reserved" trandisp="snat" transip=172.16.200.2 transport=40772 appid=40568 app="HTTPS.BROWSER" appcat="Web.Client" apprisk="medium" duration=2 sentbyte=1850
rcvdbyte=39898 sentpkt=25 rcvdpkt=37 utmaction="allow" countapp=1 devtype="Linux PC" osname="Linux" mastersrcmac="a2:e9:00:ec:40:01" srcmac="a2:e9:00:ec:40:01" srcserver=0
utmref=0-220586
date=2017-11-15 time=11:44:16 logid="0000000013" type="traffic" subtype="forward" level="notice" vd="vdom1" eventtime=1510775056 srcip=10.1.100.155 srcname="pc1" srcport=40772
srcintf="port12" srcintfrole="undefined" dstip=35.197.51.42 dstname="fortiguard.com" dstport=443 dstintf="port11" dstintfrole="undefined"
poluuid="707a0d88-c972-51e7-bbc7-4d421660557b" sessionid=8058 proto=6 action="close" policyid=1 policytype="policy" policymode="learn" service="HTTPS" dstcountry="United States"
srccountry="Reserved" trandisp="snat" transip=172.16.200.2 transport=40772 appid=40568 app="HTTPS.BROWSER" appcat="Web.Client" apprisk="medium" duration=2 sentbyte=1850
rcvdbyte=39898 sentpkt=25 rcvdpkt=37 utmaction="allow" countapp=1 devtype="Linux PC" osname="Linux" mastersrcmac="a2:e9:00:ec:40:01" srcmac="a2:e9:00:ec:40:01" srcserver=0
utmref=0-220586
date=2017-11-15 time=11:44:16 logid="0000000013" type="traffic" subtype="forward" level="notice" vd="vdom1" eventtime=1510775056 srcip=10.1.100.155 srcname="pc1" srcport=40772
srcintf="port12" srcintfrole="undefined" dstip=35.197.51.42 dstname="fortiguard.com" dstport=443 dstintf="port11" dstintfrole="undefined"
poluuid="707a0d88-c972-51e7-bbc7-4d421660557b" sessionid=8058 proto=6 action="close" policyid=1 policytype="policy" policymode="learn" service="HTTPS" dstcountry="United States"
srccountry="Reserved" trandisp="snat" transip=172.16.200.2 transport=40772 appid=40568 app="HTTPS.BROWSER" appcat="Web.Client" apprisk="medium" duration=2 sentbyte=1850
rcvdbyte=39898 sentpkt=25 rcvdpkt=37 utmaction="allow" countapp=1 devtype="Linux PC" osname="Linux" mastersrcmac="a2:e9:00:ec:40:01" srcmac="a2:e9:00:ec:40:01" srcserver=0
utmref=0-220586
date=2017-11-15 time=11:44:16 logid="0000000013" type="traffic" subtype="forward" level="notice" vd="vdom1" eventtime=1510775056 srcip=10.1.100.155 srcname="pc1" srcport=40772
srcintf="port12" srcintfrole="undefined" dstip=35.197.51.42 dstname="fortiguard.com" dstport=443 dstintf="port11" dstintfrole="undefined"
poluuid="707a0d88-c972-51e7-bbc7-4d421660557b" sessionid=8058 proto=6 action="close" policyid=1 policytype="policy" policymode="learn" service="HTTPS" dstcountry="United States"
srccountry="Reserved" trandisp="snat" transip=172.16.200.2 transport=40772 appid=40568 app="HTTPS.BROWSER" appcat="Web.Client" apprisk="medium" duration=2 sentbyte=1850
rcvdbyte=39898 sentpkt=25 rcvdpkt=37 utmaction="allow" countapp=1 devtype="Linux PC" osname="Linux" mastersrcmac="a2:e9:00:ec:40:01" srcmac="a2:e9:00:ec:40:01" srcserver=0
utmref=0-220586
date=2017-11-15 time=11:44:16 logid="0000000013" type="traffic" subtype="forward" level="notice" vd="vdom1" eventtime=1510775056 srcip=10.1.100.155 srcname="pc1" srcport=40772
srcintf="port12" srcintfrole="undefined" dstip=35.197.51.42 dstname="fortiguard.com" dstport=443 dstintf="port11" dstintfrole="undefined"
poluuid="707a0d88-c972-51e7-bbc7-4d421660557b" sessionid=8058 proto=6 action="close" policyid=1 policytype="policy" policymode="learn" service="HTTPS" dstcountry="United States"
srccountry="Reserved" trandisp="snat" transip=172.16.200.2 transport=40772 appid=40568 app="HTTPS.BROWSER" appcat="Web.Client" apprisk="medium" duration=2 sentbyte=1850
rcvdbyte=39898 sentpkt=25 rcvdpkt=37 utmaction="allow" countapp=1 devtype="Linux PC" osname="Linux" mastersrcmac="a2:e9:00:ec:40:01" srcmac="a2:e9:00:ec:40:01" srcserver=0
```

Internal Documentation

ECS Field	**Syslog/Dark Trace Field Name**	**Example Output**	**Comments**
@timestamp	"@timestamp"		This should be the time the event was parsed into Elastic, in UTC
tenant.name	N/A		Manually set in the router, don't change after this
event.start	syslog header timestamp	<165>**Nov 22 09:02:32**10.107.250.1 {"creationTime":...	This should be the original device/event timestamp from the syslog header, in UTC
event.created	creationTime	1542877373000 (*requires appropriate conversion*)	This should be the time of the event from within the contents of the event, in UTC The timestamp that the record of the breach was created. This is distinct from the "time" field.
event.timezone	N/A	*specific to customer*	This should be the original device/event timezone eg "AEST" and may need to be manually specified at parsing time based on customer

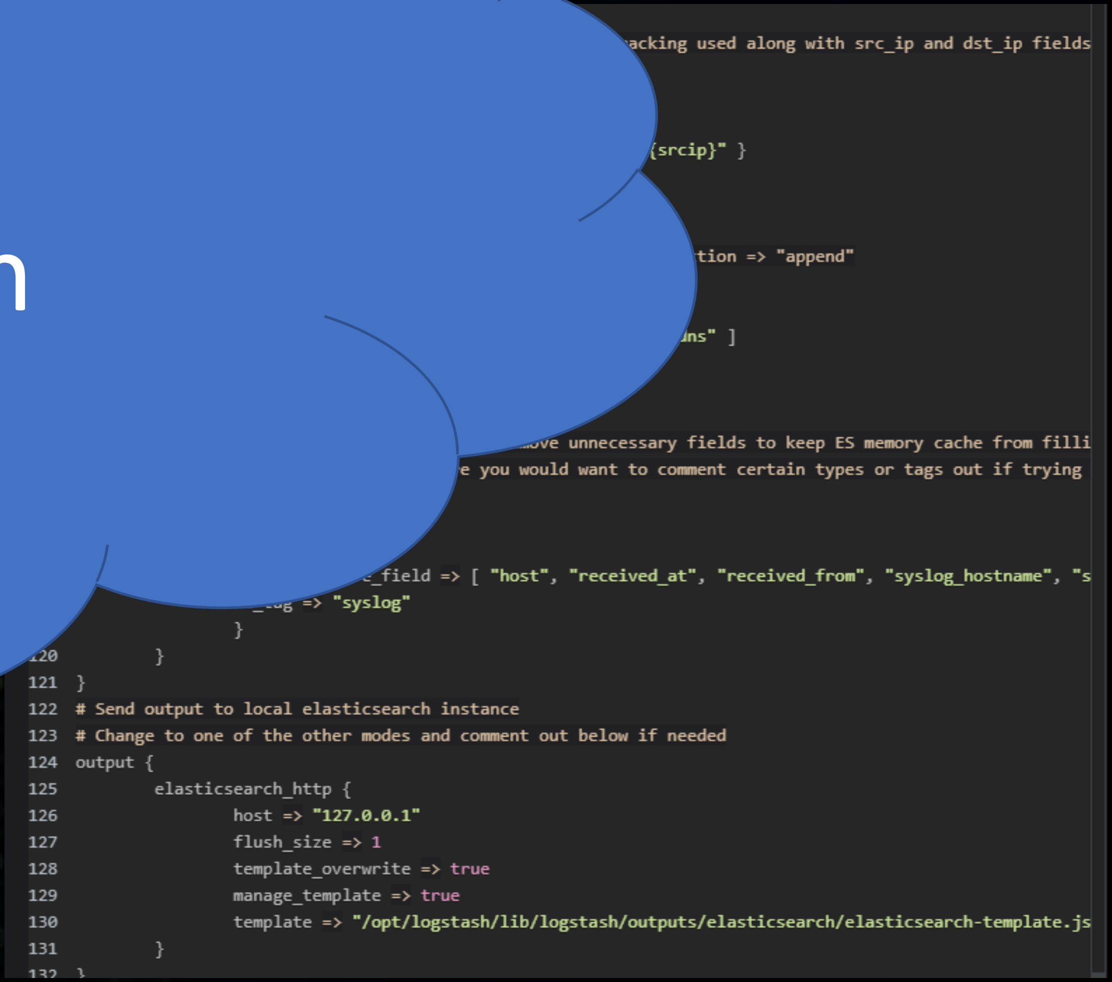
2. Normalising Live Data

Internal Documentation

ECS Field	Syslog/Dark Trace Field Name	Ex	
@timestamp	"@timestam "		
tenant.name			
event.start	sys he tin		
event.created	creationTime	(requ appropria conversion)	stamp record each was creat distinct from the "
event.timezone	N/A	*specific to customer*	This d be the original device/event timezone eg "AEST" and may need to be manually specified at parsing time based on customer

Automation

pping Configuration



```
acking used along with src_ip and dst_ip fields

{srcip}" }

tion => "append"

ns" ]

ove unnecessary fields to keep ES memory cache from filli
e you would want to comment certain types or tags out if trying

_field => [ "host", "received_at", "received_from", "syslog_hostname", "s
   => "syslog"
                }
        }
}
# Send output to local elasticsearch instance
# Change to one of the other modes and comment out below if needed
output {
        elasticsearch_http {
                host => "127.0.0.1"
                flush_size => 1
                template_overwrite => true
                manage_template => true
                template => "/opt/logstash/lib/logstash/outputs/elasticsearch/elasticsearch-template.js
        }
}
```

Motivation

Simplify the process

Reduce time and effort

Improve accuracy

Easier Automation

The Process: Data Normalisation

1

Our Solution

Constructing Mappings

2

Extension

Normalising Live Data

Our Solution: Core Functionality

Normalising input and
output fields

source_ip → source_destination

UINT32

Typing of Fields

dateTime as EPOCH
dateTime as UTC

(.*)

Types represent
regular expressions

IPv4
\b((25[0-5]|2[0-4][0-9]|[01
]?[0-9][0-9]?)(\.|$)){4}\b

149.167.143.3: Match
Bananas: No Match

Our Solution: UX Functionality

Suggestion of mappings

Cloning existing mappings and formats

Validation against sample data

Extensions

- Generating Splunk/Elastic config files automatically
- Read and detect different log formats
- Integrated Ingestion of Log files
- API access of documents
- Generating validation code for continually integration

Demo

1. Creating a Type
2. Creating an Input Format
3. Validating Input Format
4. Creating a Normalisation Format

Questions