Splunk for Cisco IronPort WSA: Troubleshooting Guide

Table of Contents

Introduction:

This manual covers the Splunk Cisco IronPort WSA Product. This application is made up of a customized Splunk app and a Splunk server polling log data collected from an IronPort Web Security Appliance. The Splunk for Cisco IronPort WSA Reporting Application provides reports and dashboards designed to give insight into data from the IronPort Web Security Appliance (WSA).

Supplement to existing Splunk documentation:

This guide serves a supplement to the existing body of documentation existing at http://docs.splunk.com.

It is not intended to replace existing Splunk troubleshooting materials or techniques. This guide will serve as a supplement by reviewing troubleshooting as it relates to this application.

Architecture Overview:

This application receives data from a Cisco IronPort WSA appliance. This data is stored in the default index. Summaries are generated and stored in the summary index. These data are then presented via reports available via the menu. Ad hoc searches are also available via the flashtimeline and the web tracking forms (simple & advanced). Performance will be improved by choosing smaller time ranges and crafting searches to be as precise as possible (sparse vs. dense searches).

Summary Overview:

There exist a number of summary indexes to speed up out of box reports. Summaries are generated once / hour. Each night, those hourly summaries are summarized into daily summaries. Each week, those daily summaries are summarized into weekly summaries.

Troubleshooting Tip:

You may use the “jobs” menu to ensure there are no scheduled searches running too long. *Too long* is any search that exceeds its frequency (e.g. an hourly search running more than an hour is too long).

Table 1 contains a schedule of saved searches that populate the summaries.

Table 1

Summary Search	Frequency
[_dashboard_overview_base-sum-search-top-1h]	Hourly at 5 minutes past
[_dashboard_overview_base-sum-search-top-1d]	Daily at 3:30 AM
[_dashboard_overview_base-sum-search-top-1w]	Weekly at 4:30 AM
[_dashboard_overview_base-sum-search-bottom-1h]	Hourly at 15 minutes past
[_dashboard_overview_base-sum-search-bottom-1d]	Daily at 3:30 AM

[_dashboard_overview_base-sum-search-bottom-1w]	Weekly at 4:30 AM
[_dashboard_overview_base-sum-search-uid-1h]	Hourly at 25 minutes past
[_dashboard_overview_base-sum-search-uid-1d]	Daily at 3:30 AM
[_dashboard_overview_base-sum-search-uid-1w]	Weekly at 4:30 AM
[_dashboard_url-categories_base-sum-search-1h]	Hourly at 35 minutes past
[_dashboard_url-categories_base-sum-search-1d]	Daily at 3:30 AM
[_dashboard_url-categories_base-sum-search-1w]	Weekly at 4:30 AM
[_dashboard_users_base-sum-search-1h]	Hourly at 45 minutes past
[_dashboard_users_base-sum-search-1d]	Daily at 3:30 AM
[_dashboard_users_base-sum-search-1w]	Weekly at 4:30 AM
[_dashboard_web-sites_base-sum-search-1h]	Hourly at 55 minutes past
[_dashboard_web-sites_base-sum-search-1d]	Daily at 3:30 AM
[_dashboard_web-sites_base-sum-search-1w]	Sunday at 4:30 AM
[_dashboard_application-visibility_base-sum-search-1h]	Hourly at 10 minutes past
[_dashboard_application-visibility_base-sum-search-1d]	Daily at 3:30 AM
[_dashboard_application-visibility_base-sum-search-1w]	Sunday at 4:30 AM
[_dashboard_web-reputation-filters_base-sum-search-1h]	Hourly at 20 minutes past
[_dashboard_web-reputation-filters_base-sum-search-1d]	Daily at 3:30 AM
[_dashboard_web-reputation-filters_base-sum-search-1w]	Sunday at 4:30 AM
[_dashboard_anti-malware_base-sum-search-1h]	Hourly at 40 minutes past

[_dashboard_anti-malware_base-sum-search-1d]	Daily 3:30 AM
[_dashboard_anti-malware_base-sum-search-1w]	Sunday 4:30 AM

Troubleshooting Tip:

Ensure field extractions are correct before backfilling summaries as explained in the installation manual

Custom Modules:

SearchRangeIntention

The SearchRangeIntention module is a custom module determined which summary level should be used for any given report based on the time range chosen.

Troubleshooting Tip:

You may see its decision by clicking on the jobs menu. Each report's search(es) will have a marker denoting search description & interval summary. (e.g. search `_dashboard_users_base-search(*,1d)` is leveraging the user's 1 day summary)

FieldPicker

The FieldPicker module is a custom module that allows fields to be chosen by individual elements after post processing from one search. The UI was designed to meet the look and feel of existing 'on box' reports with which customers were familiar.

Troubleshooting Tip:

Check the jobs menu to see what search was executing
- Copy and paste that search into the Search view
- Expand macros as needed from Manager or macros.conf

Performance:

Troubleshooting Tip:

Ensure the saved searches responsible for populating the indexes have been successfully completing within their allocation window (per above *Summary Overview*)
Performance may be slow for any number of reasons relating to CPU, memory, or disk utilization stemming from indexing, searching, and concurrency. You may identify some of these challenges using built-in searches provided in the Search app under "Status" (e.g. Status -> Index activity-> Index health). Please consult splunk.com for advice on general index heatlh, bucket spread, and other ways to identify and resolve performance challenges
Craft ad hoc searches to be as specific as possible so that they return only information needed. Returning superfluous event data is more costly in terms of searching and retrieving all the rows

Lookup Files:

Host Selection

The lookup file that populates host selection drop box is generated once per hour, 22 minutes after each hour. It is generated from metadata. The search that runs it will honor certain 'edits' to manipulate what is presented to the user.

The file may be edited to exclude hosts that exist in the metadata. To do this, simply open the file and change the corresponding "include" field of the host you which to exclude to "false". The next time the search runs it will discover

that you don't want to show that host and preserve the "false" while it updates the list from its newly found or existing hosts.

Troubleshooting Tip:

Ensure hosts.csv exists in lookups folder
Ensure the saved search which populates it runs properly
- omit the '| outputlookup lu-host' to see what it shows on screen

Ensure the host was not added since the saved search last completed successfully
Hosts may be renamed in this file as well: Be sure to verify the display name has not been edited in a way that the intended host cannot be selected from the dropdown list (i.e. ensure that host_a hasn't been accidentally renamed host_b in the display value column or both host_a and host_b will appear as host_b in the dropdown selection on reports where that feature is available

Malware Categories

The malware_categories.csv is a lookup file used by the web-tracking page to populate the dropdown list in the advanced view.

Troubleshooting Tip:

Ensure the file exists in the application's lookups folder

Transaction Types

The transaction_types.csv is a lookup file used by the web-tracking page to populate the dropdown list in the advanced and simple views.

Troubleshooting Tip:

Ensure the file exists in the application's lookups folder

Department

The departments.csv is a file used as part of the role based security functionality. This file may be edited manually or by configuring one of the role discovery scripts (available in the application's bin folder) as a scripted input. There is a script for both Linux and Windows.

Troubleshooting Tip:

Ensure the file exists in the application's lookup folder
If the Linux version is used, ensure the CLI *ldapsearch* is installed and in the Splunk user's path
If the Windows version is used "option explicit" may be commented out to reveal more specific information regarding from where and why an error may have originated.
Verify the LDAP paths are syntactically correct
Verify the bind service account name is correct
Verify the correct bind password is entered
Test connection to the remote machine over port 389
Verify the correct attribute was configured for the member name
Verify the correct attribute was used for group membership
Verify the correct attribute was configured for group name

URL Categories

The url_categories.csv is a lookup file used by the web-tracking page to populate the dropdown list in the advanced view.

Troubleshooting Tip:

Ensure the file exists in the application's lookups folder

Field Extractions

This application relies heavily on field extractions. As most reports are generated from summary data – it is important to ensure fields are being extracted correctly to enable successful and accurate reporting.

AccessLogs

Troubleshooting Tip:

Ensure timestamps are correctly being indexed

Search for "*" and ensure app-specific fields are populated in the field picker. The next bullet item contains a more thorough examination of extracted fields

Copy and paste the below search. You should not see any results and especially not very many results. If 1000 results are returned – the transforms.conf will need to be adjusted for the unique log format being indexed…

```
sourcetype=wsa_accesslogs | head 1000 | fillnull value="!!!!"
x_webcat_code_abbr x_wbrs_score x_webroot_scanverdict
x_webroot_threat_name x_webroot_trr x_webroot_spyid x_webroot_trace_id
x_mcaffe_scanverdict x_mcafee_filename x_mcafee_scan_error
x_mcafee_detecttype x_mcafee_av_virustype x_mcafee_virus_name
x_sophos_scanverdict x x_sophos_filename x_sophos_virus_name
x_ids_verdict x_icap_verdict x_webcat_req_code_abbr
x_webcat_resp_code_abbr x_resp_dvs_threat_name x_wbrs_threat_type
x_avc_app x_avc_type x_avc_behavior x_request_rewrite x_avg_bw
x_bw_throttled x_user_type
x_resp_dvs_verdictnamex_req_dvs_threat_namex_suspect_user_agent
x_wbrs_threat_reason dvc_time duration dvc_ip result http_status
bytes_in http_method dest_url user_id_dom hierarchy hierarchy_domain
mime_type acl_tag user_id user_domain dest_domain | stats count by
x_webcat_code_abbr x_wbrs_score x_webroot_scanverdict
x_webroot_threat_name x_webroot_trr x_webroot_spyid x_webroot_trace_id
x_mcaffe_scanverdict x_mcafee_filename x_mcafee_scan_error
x_mcafee_detecttype x_mcafee_av_virustype x_mcafee_virus_name
x_sophos_scanverdict x x_sophos_filename x_sophos_virus_name
x_ids_verdict x_icap_verdict x_webcat_req_code_abbr
x_webcat_resp_code_abbr x_resp_dvs_threat_name x_wbrs_threat_type
x_avc_app x_avc_type x_avc_behavior x_request_rewrite x_avg_bw
x_bw_throttled x_user_type
x_resp_dvs_verdictnamex_req_dvs_threat_namex_suspect_user_agent
x_wbrs_threat_reason dvc_time duration dvc_ip result http_status
bytes_in http_method dest_url user_id_dom hierarchy hierarchy_domain
mime_type acl_tag user_id user_domain dest_domain | convert
ctime(dvc_time) | search user_id="!!!!" AND host="!!!!" AND
src_ip="!!!!" AND cause="!!!!" AND action="!!!!" AND dest_domain="!!!!"
```

Verify the host extractions are correct. This is part of the inputs strategy discussed in the installation guide. The folder structure should be appropriately established to allow proper host extractions to occur.

- Hosts may be renamed per the section of this guide that discusses the host lookup file

Trafmonlogs

The L4TM reports are generated from L4TM data (not summary data). Field extractions will still need to be operable for those reports to function though the format is not as versatile as accesslogs, they may still be verified with the same technique.

Troubleshooting Tip:

Use this search and verify there are little to no results:

```
sourcetype=wsa_trafmonlogs | head 1000 | fillnull value="!!!!" dvc_time
log_level action proto src_ip src_port dest_ip dest_host dest_port |
stats count by dvc_time log_level action proto src_ip src_port dest_ip
dest_host dest_port | search src_ip="!!!!"
```