Reducing noise in Splunk ITSI

You are swamped with alerts in your ITSI instance. Even though you use event aggregation policies and are able to group alerts by well understood factors, you still need help identifying the alerts or groups of alerts that appear to be unusual compared to what you normally see.

Recognize when alert volumes are high

There are a couple of factors that can be examined when deciding whether alert volumes are high:

- Is this an unusual volume of alerts for this time of day?
- Is this an unusual volume of alerts for this particular service?
- Is this an unusual volume of alerts for this community of services?

The following search uses the Probability Density Function in the Splunk Machine Learning Toolkit. It generates an anomaly score for each service by looking at the expected number of alerts at any given time for any service, for each individual service, and for each community label as well. It allows you to identify the services that have the most unusual volumes of alerts.

```
index=itsi_tracked_alerts
| bin _time span=5m
| stats count as alerts by _time service_name
| join service_name type=outer [inputlookup service_community_labels.csv | table src labeled_community | dedup src labeled_community | rename src as service_name]
| eval hour=strftime(_time,"%H")
| fit DensityFunction alerts by "hour" into df_itsi_tracked_alerts_volume as alert_volume_outlier
| fit DensityFunction alerts by "service_name" into df_itsi_tracked_alerts_service as service_alert_outlier
| fit DensityFunction alerts by "labeled_community" into df_itsi_tracked_alerts_community as community_alert_outlier
| eval anomaly_score=0
| foreach *_outlier [| eval anomaly_score=anomaly_score+<<FIELD>>]
| table _time alerts service_name anomaly_score *_outlier
| xyseries _time service_name anomaly_score
| fillnull value=0
```

You could even plot the results over time to see what the anomaly scores are – perhaps comparing this to a service health score or another metric that provides more context about the health of your environment.

If you have over 1000 services or community labels in your data, you may want to consider training separate density function models (perhaps by service or community) to manage the load on your Splunk instance.

These results only tell you about unusual volumes of alerts. They don’t have much context about what has made up those alerts.
Recognize unusual sourcetype combinations

An unusual sourcetype combination can indicate activity you want to investigate. This search looks for a count of the different sourcetypes seen in the tracked alerts index every five minutes for each service. The ratio of the count to the total will tell you a bit about how likely this sourcetype combination is for each service.

```
index=itsi_tracked_alerts
  | bin _time span=5m
  | stats values(orig_sourcetype) as sourcetypes by _time service_name
  | eval sourcetypes=mvjoin(mvsort(sourcetypes),"")
  | stats count by sourcetypes service_name
  | eventstats sum(count) as total by service_name
```

Recognize unusual alert descriptions

Here we’re going to use the Smart Ticket Insights app for Splunk to determine the likely alert descriptions in our data. Start with a search that returns the alert descriptions, IDs, and services from your correlation search index.

This app is not Splunk supported.

```
index=itsi_tracked_alerts
  | table event_id, service_name, description
  | eval type="everything"
```

Then, select the following fields from the dropdown to get a report on the data.

- Incident ID Field: event_id
- Category Field: type
- Subcategory Field: service_name

After the report panels have populated, select Single threshold from the dropdown and click Identify frequently occurring types of tickets. On the next dashboard, you may want to modify some of the selections depending on the groups that are identified.

After you have checked over the groups, save the model and move onto the Manage smart groups dashboard. Selecting a group gives you an Open in search button. Click it to run a search that provides information about expected alert descriptions for each service that we have some correlation searches against. You may want to make a few changes to the provided search, such as adding in the _time field throughout the search or replacing the last few lines to calculate some statistics.

Next steps

The content in this article comes from a previously published blog, one of the thousands of Splunk resources available to help users succeed. These additional Splunk resources might help you understand and implement the recommendations in this article:
• .Conf Talk: A prescriptive design for enterprise-wide alerts in IT Service Intelligence
• Docs: Splunk ITSI Content Packs