Expedition

Log Analysis Feature Guide

Version 1.0.5

Palo Alto Networks, Inc.
www.paloaltonetworks.com

Revision Date: Apr 11, 2022

Contents

What are the Log Analysis Features in Expedition?

The Expedition provided below log analysis features that can help you to either migrate your service-port based security policies to APP-ID based security policies or to refine your existing security policies based on traffic logs.

- **APP-ID Adoption:** This feature is used when you have a set of service-port based security policies and you would like to migrate those legacy security policies to APP-ID based security policies. Expedition will retrieve APP-ID info from the active log connector, this feature does not require Firewall logs stored in Expedition.

- **Rule Enrichment (RE)**: This feature is used when you would like to tighten your existing security policies to remove "any" in security policies. For example, you have a security policy that contains "any" in either applications, users, zones, or services fields. This feature will auto discover APP-ID and service port info in the firewall traffic logs to see if it matched the application-default ports and help you to tighten your security policies to replace "any '' with correct APP-ID and service ports in the security policy. Expedition will process firewall log in csv format , so this feature does require Firewall logs stored in Expedition.

- **Machine Learning (ML)**: This feature will suggest new sets of security policies based on analysis of the firewall traffic logs. It is often used in Greenfield deployment or when you have a set of rules that's more permissive than required and you don't know what security policies are required. The ML process will identify servers, consumers and provide all the security policies including source, destinations, APP-ID, and service-ports . Expedition will process firewall log in csv format , so this feature does require Firewall logs stored in Expedition.

Which features should you choose?

Depending on your use case , you will choose different log analysis features. Below are some use cases and what features you can use.

- ***Use Case #1***

  I am clear on what security policies are required in my environment. I want to migrate all my service-port based (Layer 4)security policies to APP-ID (Layer 7) based security policies.

  Feature: **If you are clear on the security policy requirements, you will choose APP-ID Adoption, with this feature, you do not need to store firewall logs in Expedition. (if the PAN-OS device is running PANOS 9.1 or later, it is recommended to directly use Policy Optimizer feature in PAN-OS)**

- ***Use Case #2***
  I want to migrate all my service-port based (Layer 4)security policies to APP-ID (Layer 7) based security policies and would like to see if Expedition can auto suggest new APP-ID based rules based on live traffic logs.

Feature: **If you are looking for security rules suggestions based on live traffic logs, you will choose RE or ML , however these features require you to store firewall logs in Expedition. If you can't send traffic logs to Expedition, you will choose "APP-ID adoption" instead.**

- ***Use Case #3***

  I am not clear on what security policies are required in my environment , also I want to migrate all my service-port based (Layer 4)security policies to APP-ID (Layer 7) based security policies.

  Feature: Machine Learning (ML). **This feature will help you by suggesting a new set of security policies per APP-ID with detailed source and destination info.**

- ***Use Case #4***

  Most of my security policies are APP-ID based security policies and some of them contain "Any" in the service field, I would like to see if I can tighten the security policies to only use "application-default" or the ports that's needed.

  Feature: Rules Enrichment (RE). **This feature will help you refine the existing APP-ID based policy.**

- ***Use Case #5***

  We have security policies that are more permissive than what we really required and I would like to know what security policies are actually required.

  Feature: Machine Learning (ML). **This feature will help you by suggesting a new set of security policies per APP-ID with detailed source and destination info.**

- ***Use Case #6***

  This is a greenfield firewall deployment and we are not clear on what security policies are required , most of the security policies are too permissive.

  Feature: Machine Learning (ML). **This feature will help you by suggesting a new set of security policies per APP-ID with detailed source and destination info.**

Steps for Expedition Log Analysis Features

Here is the list of the steps for the log analysis features:

1. Adding a PAN-OS device in the device tab and retrieve latest contents
2. Creating a new Expedition project
3. Adding Panorama or firewall as Log-Connector
4. Importing the latest running configuration from PAN-OS device (Panorama or firewall)
5. Selecting Security Policies
6. Required Steps for ML and RE (For APP-ID Adoption, please jump to step 7)
   a. Configure Machine Learning settings
   b. Importing traffic logs to Expedition
   c. Configure M.Learning on the device
   d. Processing the logs from the selected PAN-OS device
7. Please refer below individual sections for the different subsequent steps
   a. APP-ID Adoption
   b. Rule Enrichment (RE)
   c. Machine Learning (ML)
8. Push Modified Security Policies back to PAN-OS Devices

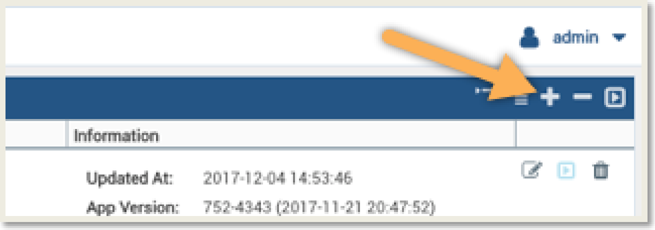
Adding a PAN-OS device in the device tab

Please follow the steps below to add the PAN-OS device in the device tab and retrieve the latest contents form the device.

Adding a Next-gen firewall and retrieve its contents

Expedition supports all the PanOS versions since version 7.0 up to 10.x. Let's follow an example on how to create a new Device and import the configuration and securely store it on Expedition.

a) Navigate to the DEVICES tab
b) Add a new Device by clicking on the *plus* button located on the top-right from the panel.



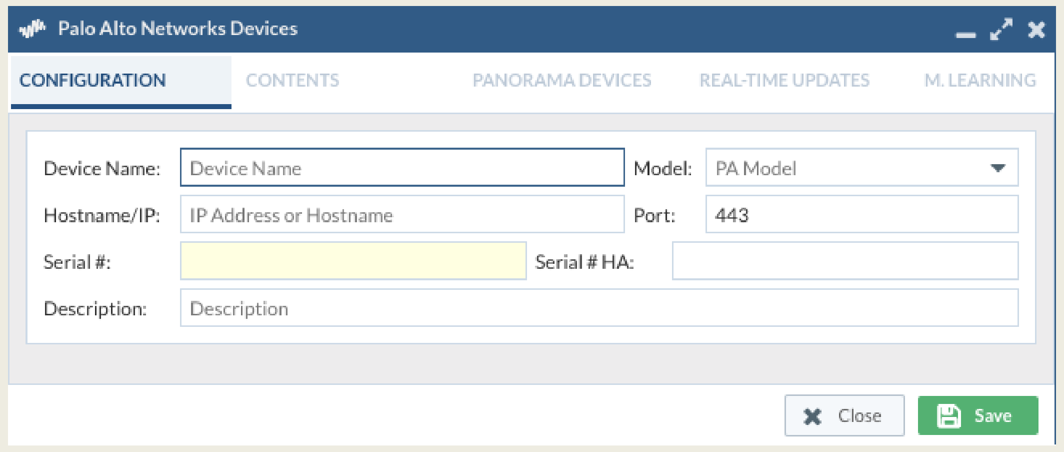
Updated At: 2017-12-04 14:53:46
App Version: 752-4343 (2017-11-21 20:47:52)

c) A new window will be shown to fill with all the information required.

- Device Name: It's the name you want to call your firewall
- Model: Palo Alto Networks device model
- Hostname/IP: IP or name used to connect to your firewall, if it's a name Expedition needs to know how to resolve it, check the DNS used by Expedition it's the right one. You can check from the CLI

```
# sudo cat /etc/resolv.conf
```

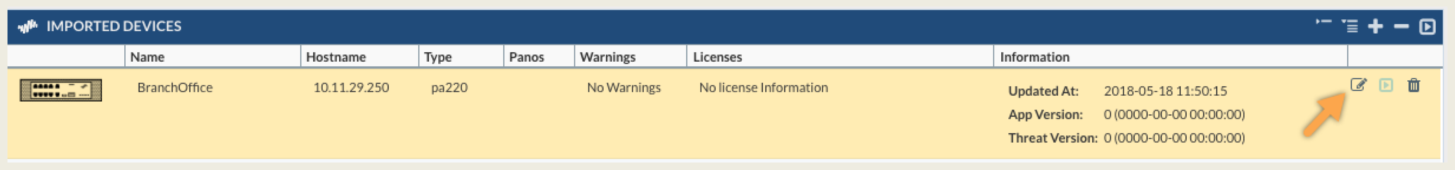
- Port: where the management is running, by default 443
- Serial #: This field is required and will be used as an Index to use the right one.
- Serial # HA: In case this firewall is part of a Cluster you can set the HA serial. This will matter for the Machine Learning module which will be explained in another chapter of this document.



Device Name: Device Name
Model: PA Model
Hostname/IP: IP Address or Hostname
Port: 443
Serial #:
Serial # HA:
Description: Description

- Click on Save to add the Device to Expedition

d) Once the Device has been created and listed from the Devices view we have to edit and add the credentials to retrieve the contents like applications database, system information and the configuration. Select the device and double click to edit it or by clicking on the Edit button (pencil).



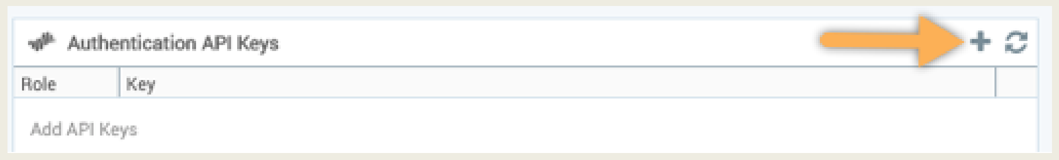
IMPORTED DEVICES

Name	Hostname	Type	Panos	Warnings	Licenses	Information
BranchOffice	10.11.29.250	pa220		No Warnings	No license Information	Updated At: 2018-05-18 11:50:15 App Version: 0 (0000-00-00 00:00:00) Threat Version: 0 (0000-00-00 00:00:00)

e) The Edit Device window now is displayed. From the Configuration Tab let's add our credentials to connect to the firewall and Expedition will request the firewall to generate a new API key.

notice the generated API Key will be valid as long as the user doesn't change the password from the firewall.

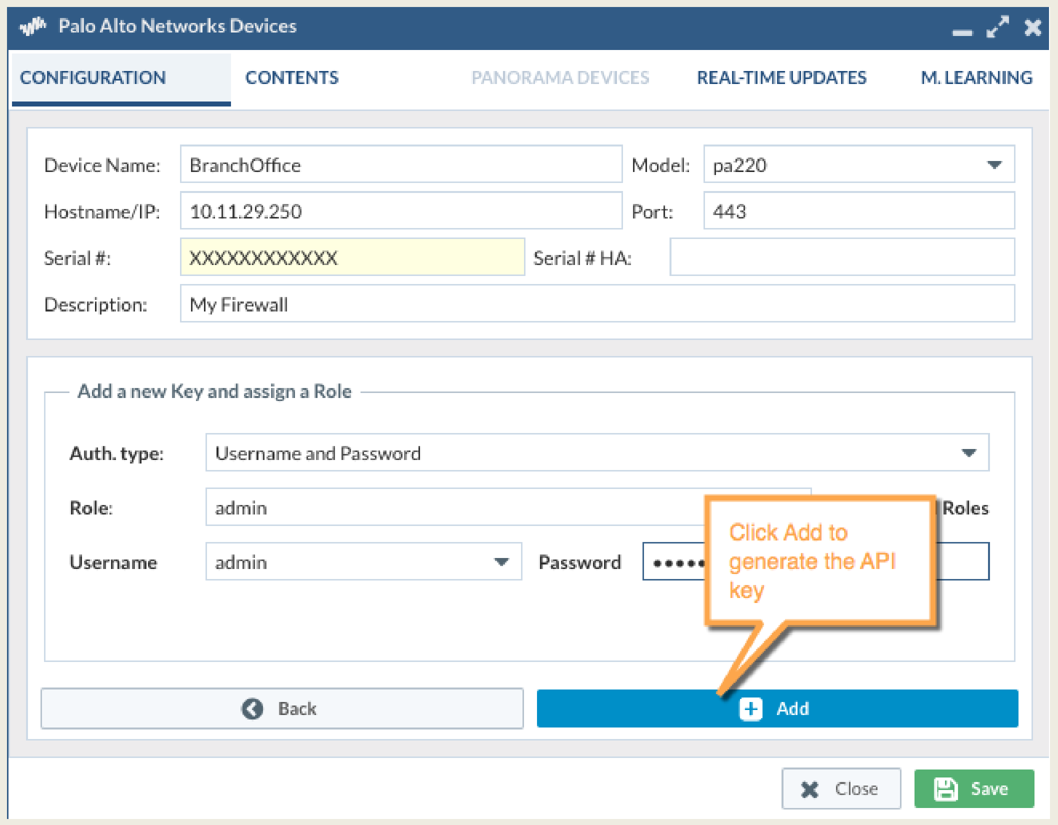
- Click on the plus icon to add a new API Keys



Authentication API Keys

Role	Key
Add API Keys	

- Auth. Type: How we want to authenticate against he firewalls, we can choose to provide username and password and let Expedition request the API key to your firewall or in case you already have the API key choose API KEY and paste your key in the text field



Device Name: BranchOffice

Model: pa220

Hostname/IP: 10.11.29.250

Port: 443

Serial #: XXXXXXXXXXXX

Serial # HA:

Description: My Firewall

Add a new Key and assign a Role

Auth. type: Username and Password

Role: admin

Roles

Username admin

Password •••••

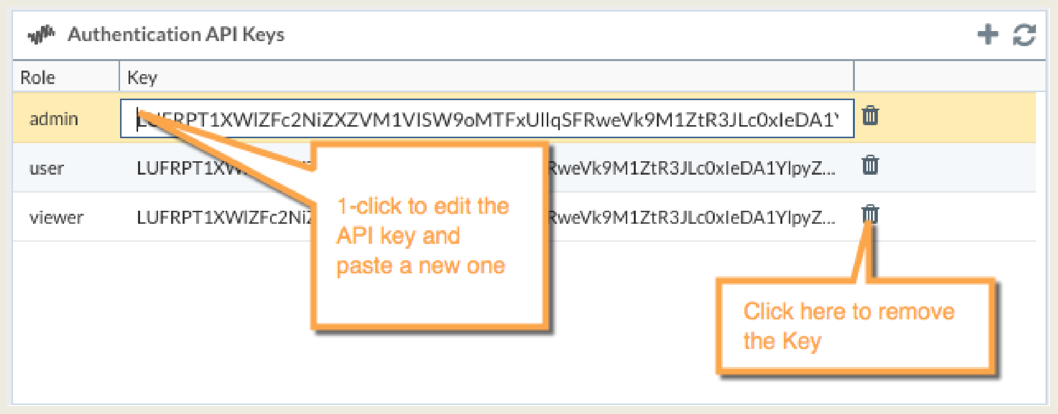
Click Add to generate the API key

In this example we are going to use Username and Password and provide them:

- Role and Apply all Roles: When you add a new API Key this can be attached to a Role inside Expedition, that means when you have a user from Expedition with Role admin inside one Project and that user tries to push changes using API Keys Expedition will use the API Key based on the user's Role in this example admin. If you didn't add an API key to the admin role that user will be unable to send any API Call out.

For small environments where you will have only one user and it will be admin there is no need to check the Apply all Roles and keep that key only attached to the admin Role.

- Click on the Add blue button to generate the Keys.



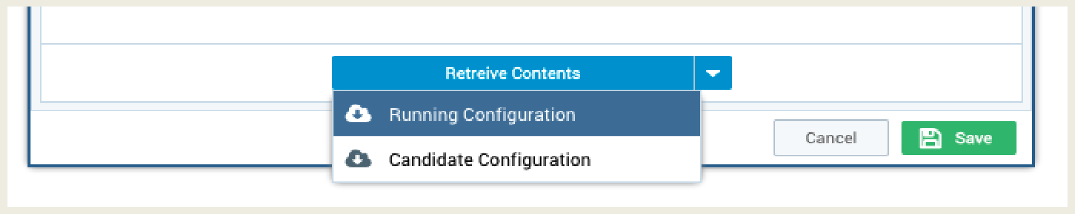
Authentication API Keys

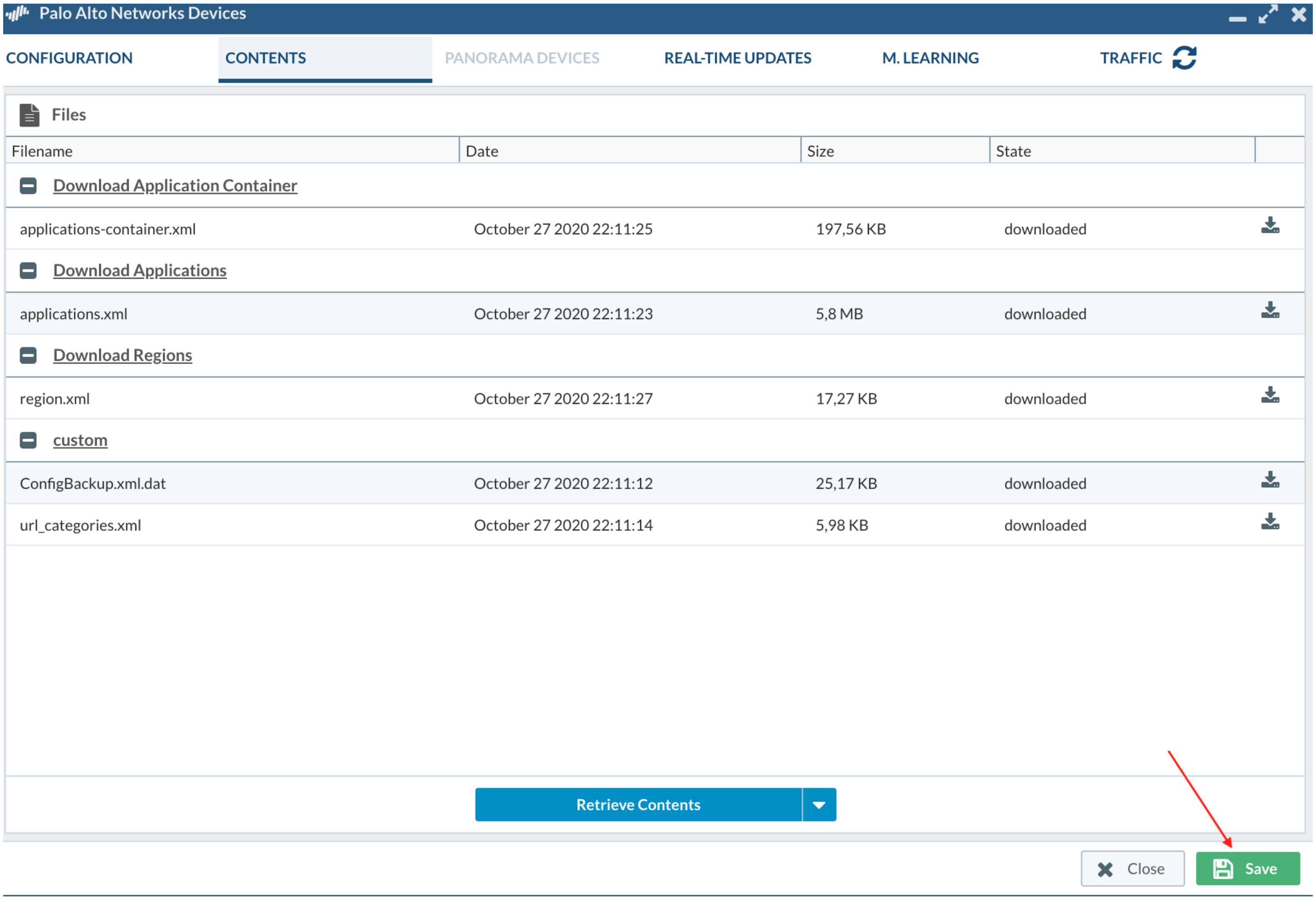
Role	Key
admin	L FRPT1XWIZFc2NiZXZVM1VISW9oMTFxUllqSFRweVk9M1ZtR3JLc0xleDA1'
user	LUFRPT1XW RweVk9M1ZtR3JLc0xleDA1YlpyZ...
viewer	LUFRPT1XWIZFc2NiZ RweVk9M1ZtR3JLc0xleDA1YlpyZ...

1-click to edit the API key and paste a new one

Click here to remove the Key

- Navigate to CONTENTS to retrieve the Running configuration.





Filename	Date	Size	State
Download Application Container			
applications-container.xml	October 27 2020 22:11:25	197,56 KB	downloaded
Download Applications			
applications.xml	October 27 2020 22:11:23	5,8 MB	downloaded
Download Regions			
region.xml	October 27 2020 22:11:27	17,27 KB	downloaded
custom			
ConfigBackup.xml.dat	October 27 2020 22:11:12	25,17 KB	downloaded
url_categories.xml	October 27 2020 22:11:14	5,98 KB	downloaded

- Once the download process is completed, you will see the “State” column changed to “Downloaded” , then click on “*Save*” .

Expedition downloaded the device configuration and stored it in the hard-drive encrypted. You can check the file from the CLI by entering in the following folder:

```
$ cd /home/userSpace
$ cd  devices
$ cd <the device serial number >
$ ls -la
```

For debugging purposes there is a file on the devices folder called “debug.txt”. The content of this file comes from the daemon who controls the access to the firewalls so expect to find the API requests made to retrieve the keys or dynamic reports in some cases, for security we stripped out the parameter key from the request. So, in case you want to re-use an API call you will have to add at the end the &key=<and your API key>

At this moment Expedition keeps a snapshot of your running configuration. In case you make any change on the device and you want to update your snapshot you have to edit the Device again and retrieve the running configuration again.

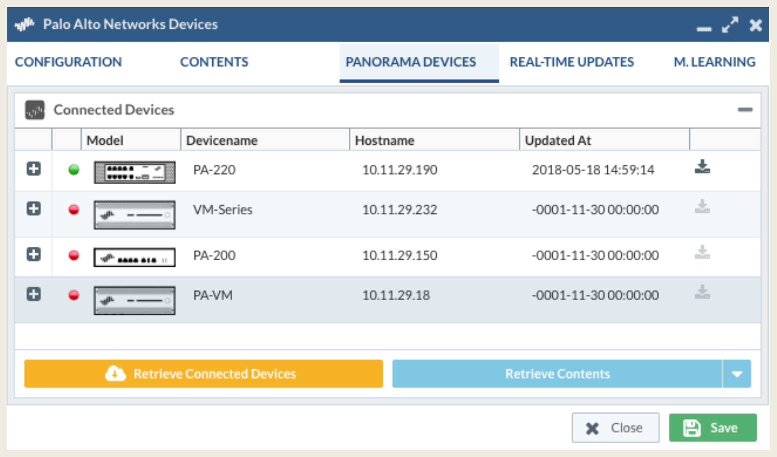
If you were already using that config on a project and you want to import the changes there is no other way than remove the current config from the project and import the device config again, all changes you made on that config and not exported will be lost.

Adding a Panorama and retrieve it's contents

Importing a Panorama is really similar to import a Next-gen Firewall and here we will show only what's is different

You can do the same steps used to create a new Next-gen Firewall, generate the API keys and retrieve the Contents, now a new tab called PANORAMA DEVICES will be active, so click on that tab.

1. If we need to play with the firewalls connected to that Panorama we need to know them first, so click on Retrieve Connected Devices. This will request that information to Panorama and create the devices on Expedition but referencing them under this Panorama device. That means all the requests we will generate to talk with that Firewall will be done using Panorama as a Proxy.

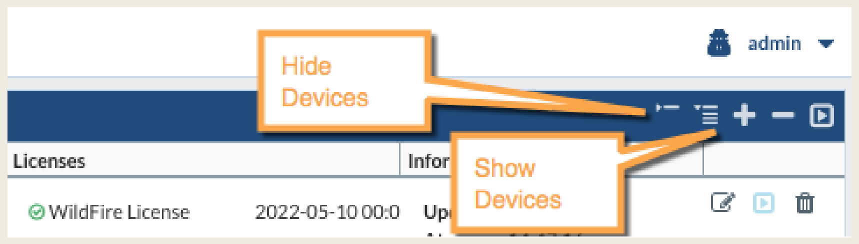


		Model	Devicename	Hostname	Updated At
⊞	●		PA-220	10.11.29.190	2018-05-18 14:59:14
⊞	●		VM-Series	10.11.29.232	-0001-11-30 00:00:00
⊞	●		PA-200	10.11.29.150	-0001-11-30 00:00:00
⊞	●		PA-VM	10.11.29.18	-0001-11-30 00:00:00

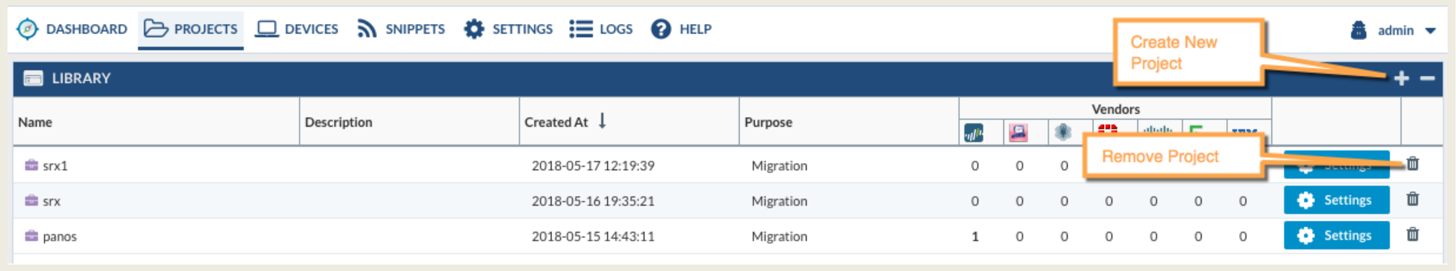
2. In case we need to download the configuration of one of these devices just select it from the list and click on Retrieve Contents. This will allow you to import the configuration of that firewall into your project once we create the project.

If you have one firewall created in Expedition and then you import Panorama and the that firewall it's listed as one of the connected devices to that panorama (Serial Number is used as index here) the existing firewall in Expedition will go under the Panorama management meaning from that moment the API requests against it will pass through Panorama.

From the Devices view we can hide all the connected devices to its Panorama by using the following buttons from the Panel header



Creating a new Expedition project



Create New Project

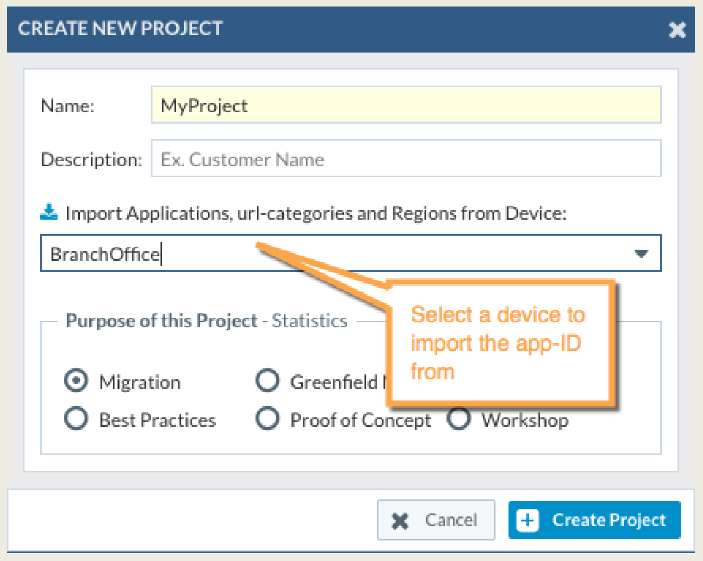
Remove Project

Name	Description	Created At	Purpose	Vendors								
srx1		2018-05-17 12:19:39	Migration	0	0	0					Settings	
srx		2018-05-16 19:35:21	Migration	0	0	0	0	0	0	0	Settings	
panos		2018-05-15 14:43:11	Migration	1	0	0	0	0	0	0	Settings	

In Expedition each project has its own database. You can create as many projects as you want.

Let's create a new project and see the workflow we have to follow:

a) Click on the plus button to create a new Project
b) A new window will pop up, assign a name to your Project and in case you already created a Device you can select it from the combo box, when you select a firewall or panorama on that combo box you are forcing Expedition to import the same Applications database to your project, that Applications database was downloaded to Expedition at the same time we were retrieving the configuration. Doing this your Project will have the same applications that your Firewall.
c) Purpose if this Project: this is used to provide some statistics but doesn't affect in nothing
d) Click on *Create Project*.



Name: MyProject

Description: Ex. Customer Name

Import Applications, url-categories and Regions from Device:

BranchOffice

Purpose of this Project - Statistics

● Migration ○ Greenfield I
○ Best Practices ○ Proof of Concept ○ Workshop

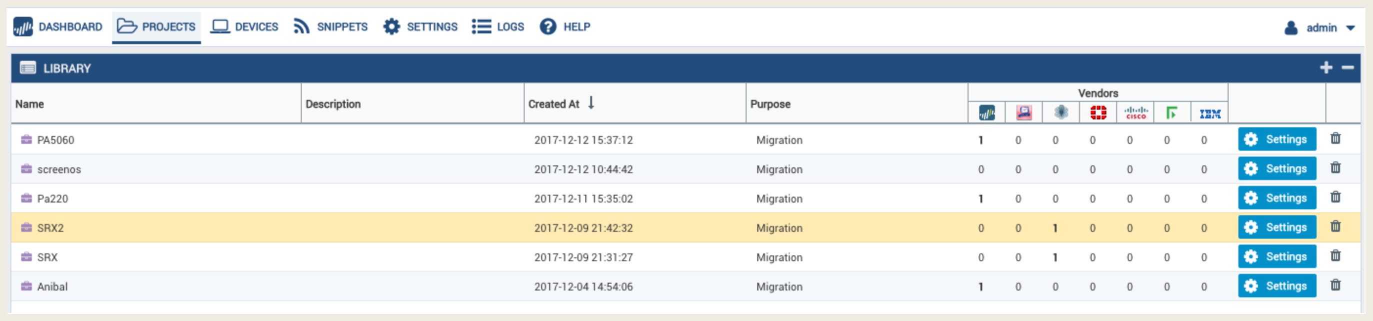
Select a device to import the app-ID from

In case we selected a device, this will automatically be added to your project, you can edit these settings to manage users and devices from within the Project.

Project Settings

After the project creation we can continue managing the project , we can add more users or devices to the project and import or export the project to be shared with other Expedition instances.

To Edit the Settings just select the Project and click on Settings

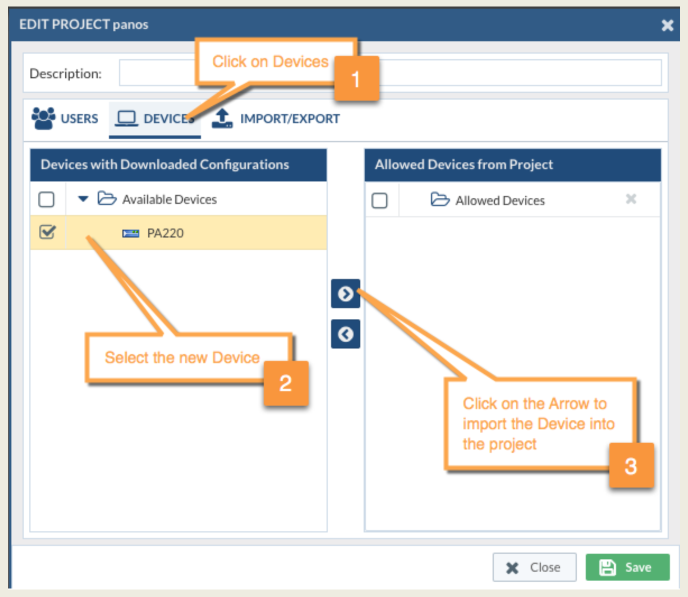


Name	Description	Created At ↓	Purpose	Vendors								
								cisco		IBM		
PA5060		2017-12-12 15:37:12	Migration	1	0	0	0	0	0	0	Settings	
screenos		2017-12-12 10:44:42	Migration	0	0	0	0	0	0	0	Settings	
Pa220		2017-12-11 15:35:02	Migration	1	0	0	0	0	0	0	Settings	
SRX2		2017-12-09 21:42:32	Migration	0	0	1	0	0	0	0	Settings	
SRX		2017-12-09 21:31:27	Migration	0	0	1	0	0	0	0	Settings	
Anibal		2017-12-04 14:54:06	Migration	1	0	0	0	0	0	0	Settings	

Manage Devices

If we want to allow access to import a configuration from an existing device or enable the users to generate API calls to be sent to the device you must allow the project to have access to that device.

To do it just click on Devices and select the firewalls you want to allow and move them to the right panel



Click on Devices 1

Select the new Device 2

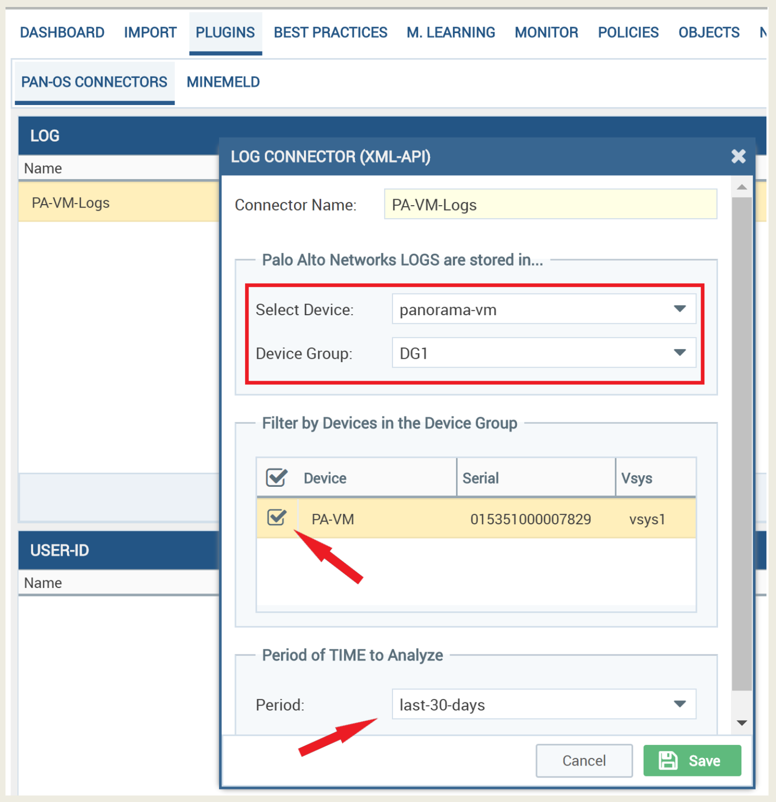
Click on the Arrow to import the Device into the project 3

Enter into the Project and Import the configuration from the device who has the configuration (Panorama or a firewall) and remember that device must be in your Allowed Devices list.

Adding a PAN-OS device as the Log Connector

The log connector is used by Expedition like a FILTER to know from all the information is stored as Parquet files from what serial numbers and vsys we must dig in only and skip other information coming from a different serial and vsys.

Under the plugins tab, add a log connector. Give it a name, select the device and the device group, the firewall, and choose the time period for log analysis. This view will show you Device-Groups or Vsys based on the configuration you were importing as a Base configuration in your project. The time period can be custom as well.



PA-VM-Logs

Connector Name: PA-VM-Logs

Palo Alto Networks LOGS are stored in...

Select Device: panorama-vm

Device Group: DG1

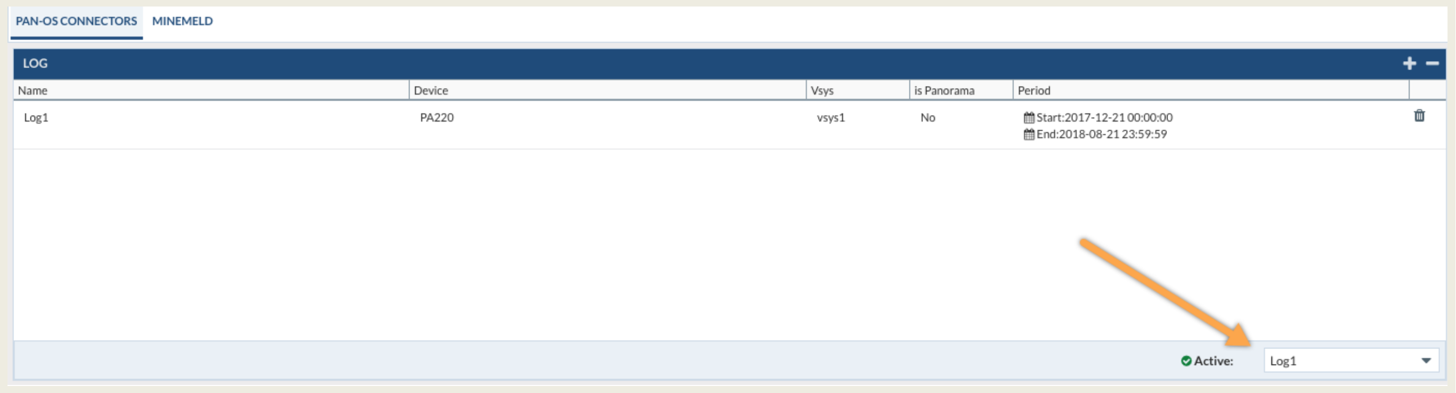
Filter by Devices in the Device Group

☑	Device	Serial	Vsys
☑	PA-VM	015351000007829	vsys1

Period of TIME to Analyze

Period: last-30-days

Ensure the Log connector is Active.

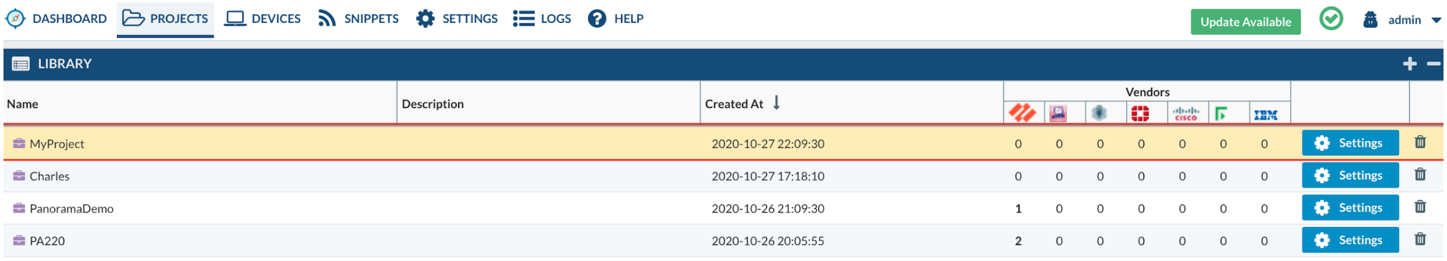


Name	Device	Vsys	is Panorama	Period
Log1	PA220	vsys1	No	Start:2017-12-21 00:00:00 End:2018-08-21 23:59:59

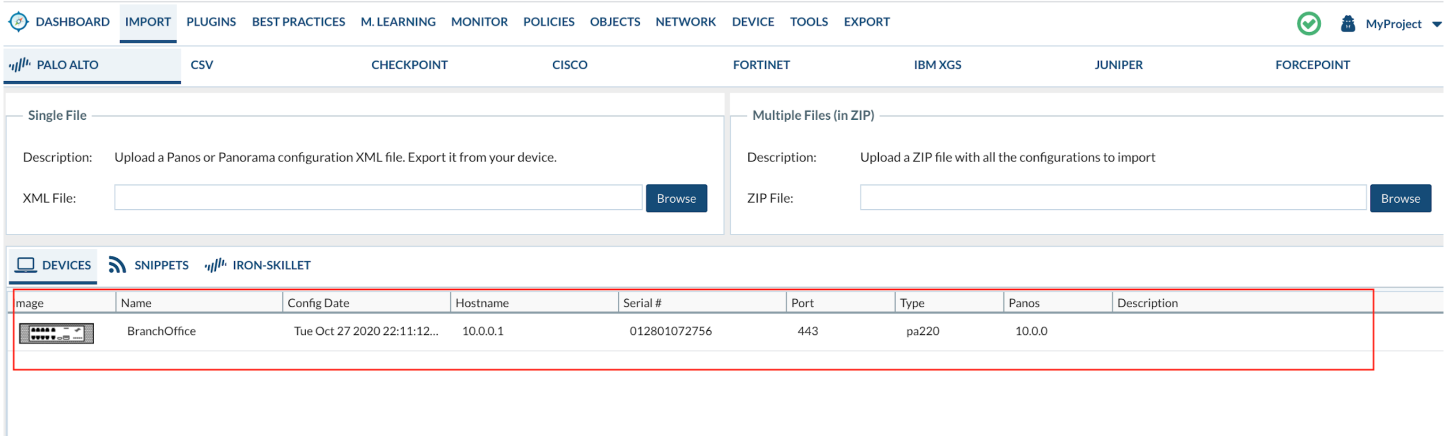
Active: Log1

Importing the latest running configuration from PAN-OS device (Panorama or firewall)

In the Projects tab, Double Click on the project you created in the previous step, go to "**Import**" and double click on the device to import the running configuration into the project.



Name	Description	Created At	Vendors								
MyProject		2020-10-27 22:09:30	0	0	0	0	0	0	0	Settings	
Charles		2020-10-27 17:18:10	0	0	0	0	0	0	0	Settings	
PanoramaDemo		2020-10-26 21:09:30	1	0	0	0	0	0	0	Settings	
PA220		2020-10-26 20:05:55	2	0	0	0	0	0	0	Settings	



Single File

Description: Upload a Panos or Panorama configuration XML file. Export it from your device.

XML File:

Multiple Files (in ZIP)

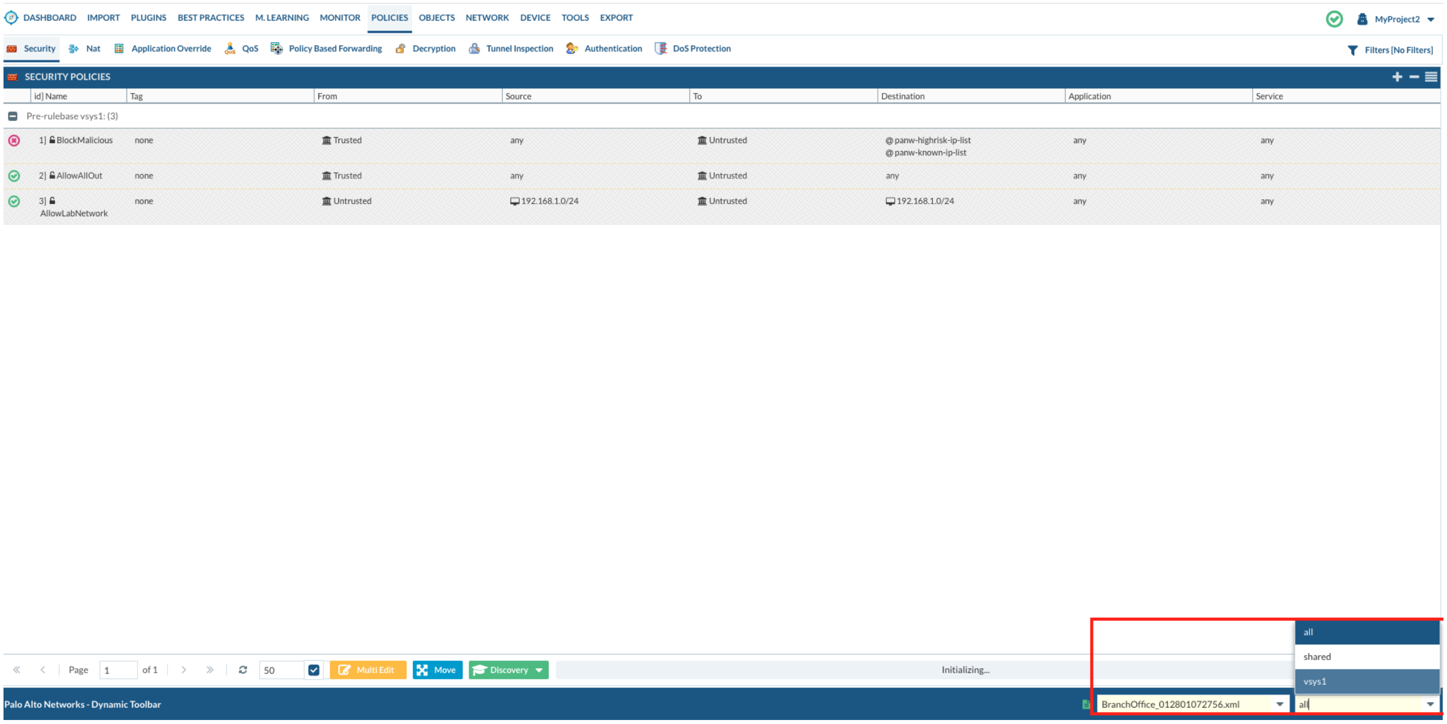
Description: Upload a ZIP file with all the configurations to import

ZIP File:

mage	Name	Config Date	Hostname	Serial #	Port	Type	Panos	Description
	BranchOffice	Tue Oct 27 2020 22:11:12...	10.0.0.1	012801072756	443	pa220	10.0.0	

Selecting Security Policies

Navigate to **“POLICIES”-> “Security”** , you will select from the right lower bottom drop down menu for the vsys or device group policy you would like to work on:

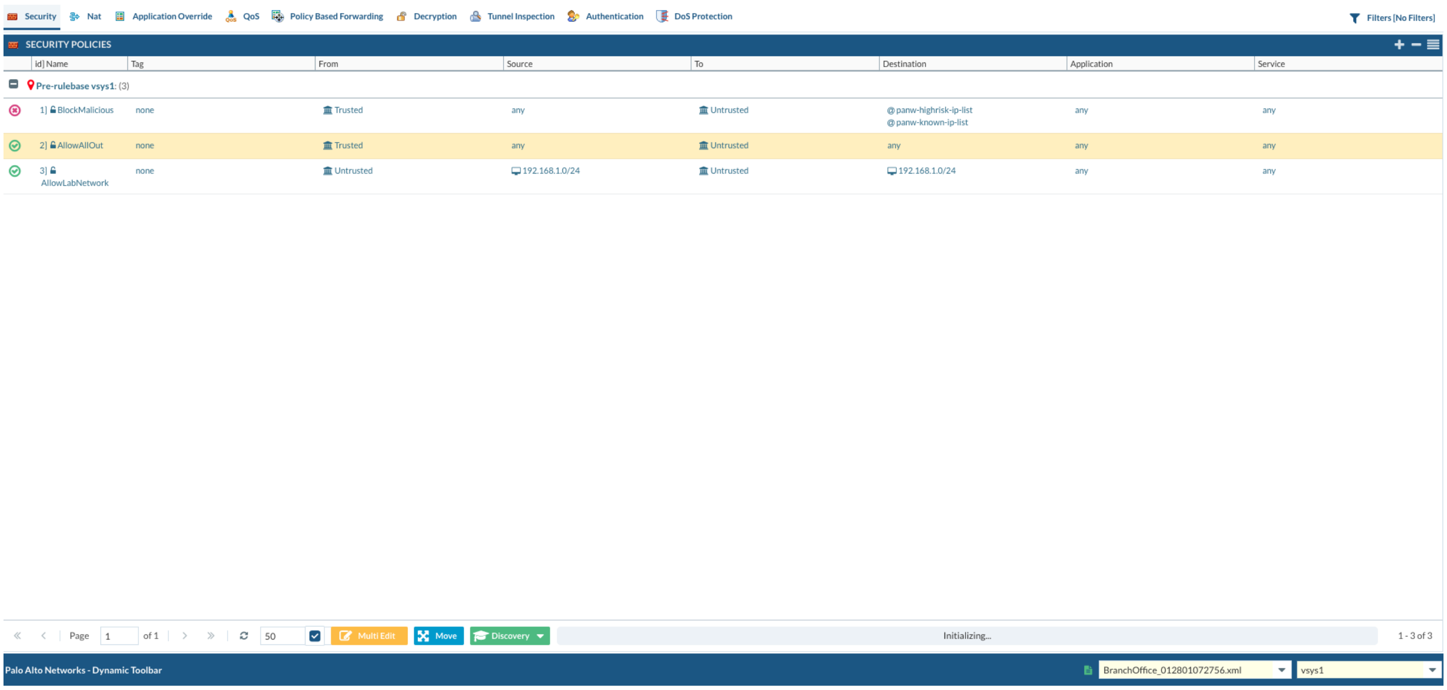


id] Name	Tag	From	Source	To	Destination	Application	Service
Pre-rulebase vsys1: (3)							
1] BlockMalicious	none	Trusted	any	Untrusted	@panw-highrisk-ip-list @panw-known-ip-list	any	any
2] AllowAllOut	none	Trusted	any	Untrusted	any	any	any
3] AllowLabNetwork	none	Untrusted	192.168.1.0/24	Untrusted	192.168.1.0/24	any	any

BranchOffice_012801072756.xml

all

Once you select the correct vsys or device group, you will see the background color of policies turn from gray to white , that means you can now edit the policies, in below example, I have a firewall configuration and the security policy is in vsys1 , so I will select vsys1.



id] Name	Tag	From	Source	To	Destination	Application	Service
Pre-rulebase vsys1: (3)							
1] BlockMalicious	none	Trusted	any	Untrusted	@panw-highrisk-ip-list @panw-known-ip-list	any	any
2] AllowAllOut	none	Trusted	any	Untrusted	any	any	any
3] AllowLabNetwork	none	Untrusted	192.168.1.0/24	Untrusted	192.168.1.0/24	any	any

BranchOffice_012801072756.xml

vsys1

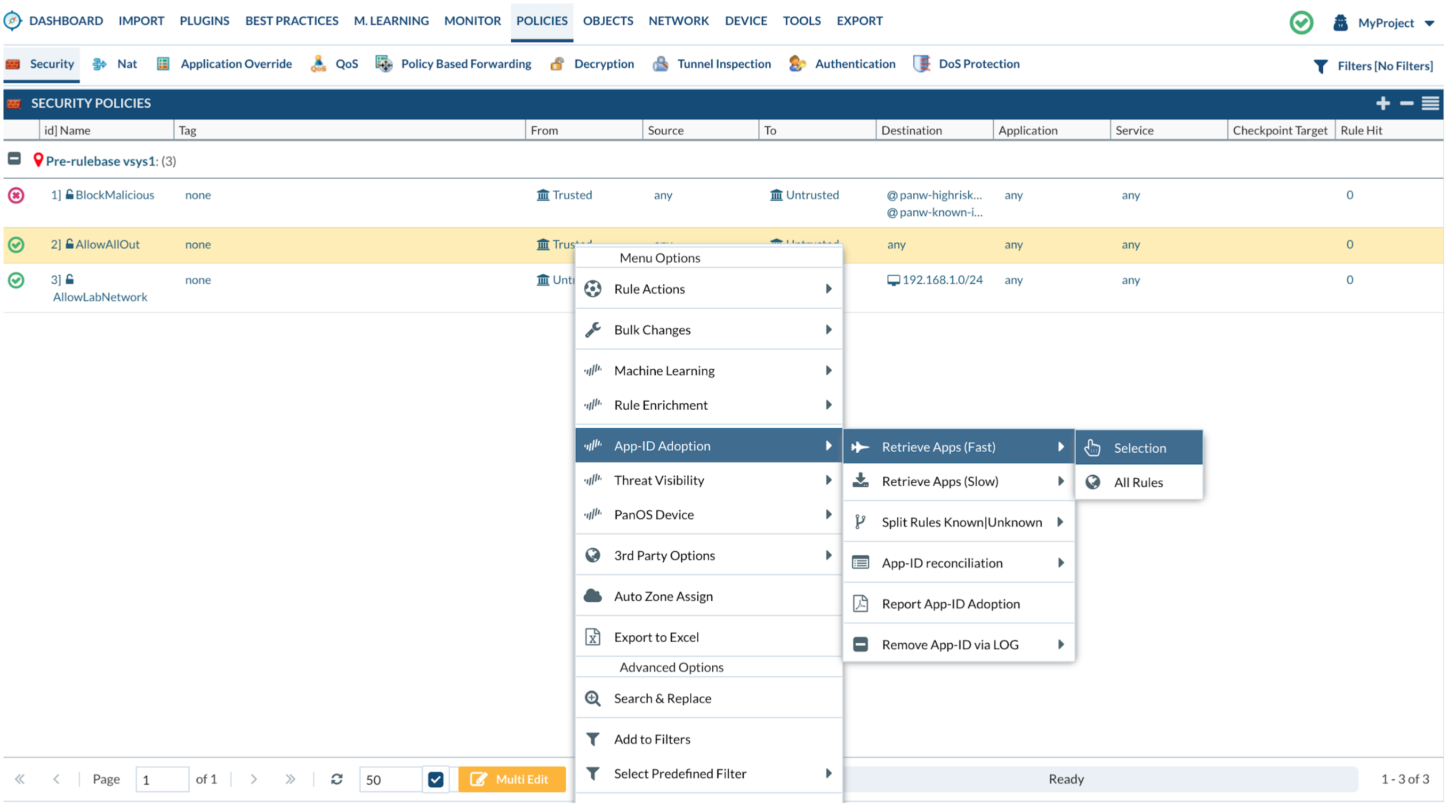
APP-ID Adoption

Expedition APP-ID Adoption feature can help you to convert port-based rules to application-based rules enables you to include the applications you want to allow in an allow list and deny access to all other applications, which improves your security posture. Restricting application traffic to its default ports prevents evasive applications from running on non-standard ports. Removing unused applications from rules is a best practice that reduces the attack surface and keeps the rulebase clean.

- ❏ Retrieve APP-IDs
- ❏ Split Known/Unknown Rules
- ❏ Clone the existing rule
- ❏ APP-ID Reconciliation

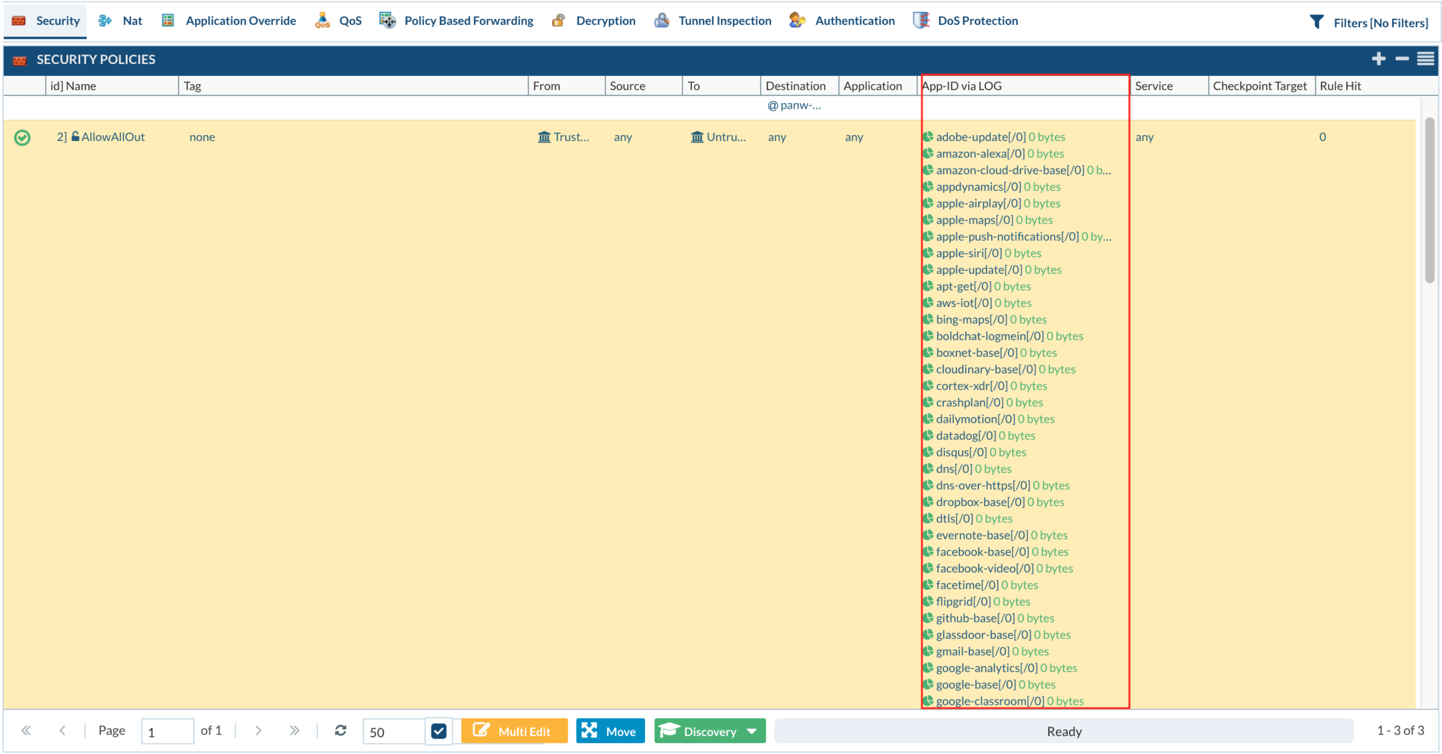
Retrieve APP-IDs

Go to **"POLICIES"** -> **"Security"** , select the Port-Based security policies , and right click **"Retrieve APPs(Fast)"** or **"Retrieve APPs(Slow)"** -> **"Selection"** **,** alternatively , you could also choose **"All Rules"** if most of your rules are port-based.



id] Name	Tag	From	Source	To	Destination	Application	Service	Checkpoint Target	Rule Hit
1] BlockMalicious	none	Trusted	any	Untrusted	@panw-highrisk... @panw-known-i...	any	any		0
2] AllowAllOut	none	Trusted	any	Untrusted	any	any	any		0
3] AllowLabNetwork	none	Unt			192.168.1.0/24	any	any		0

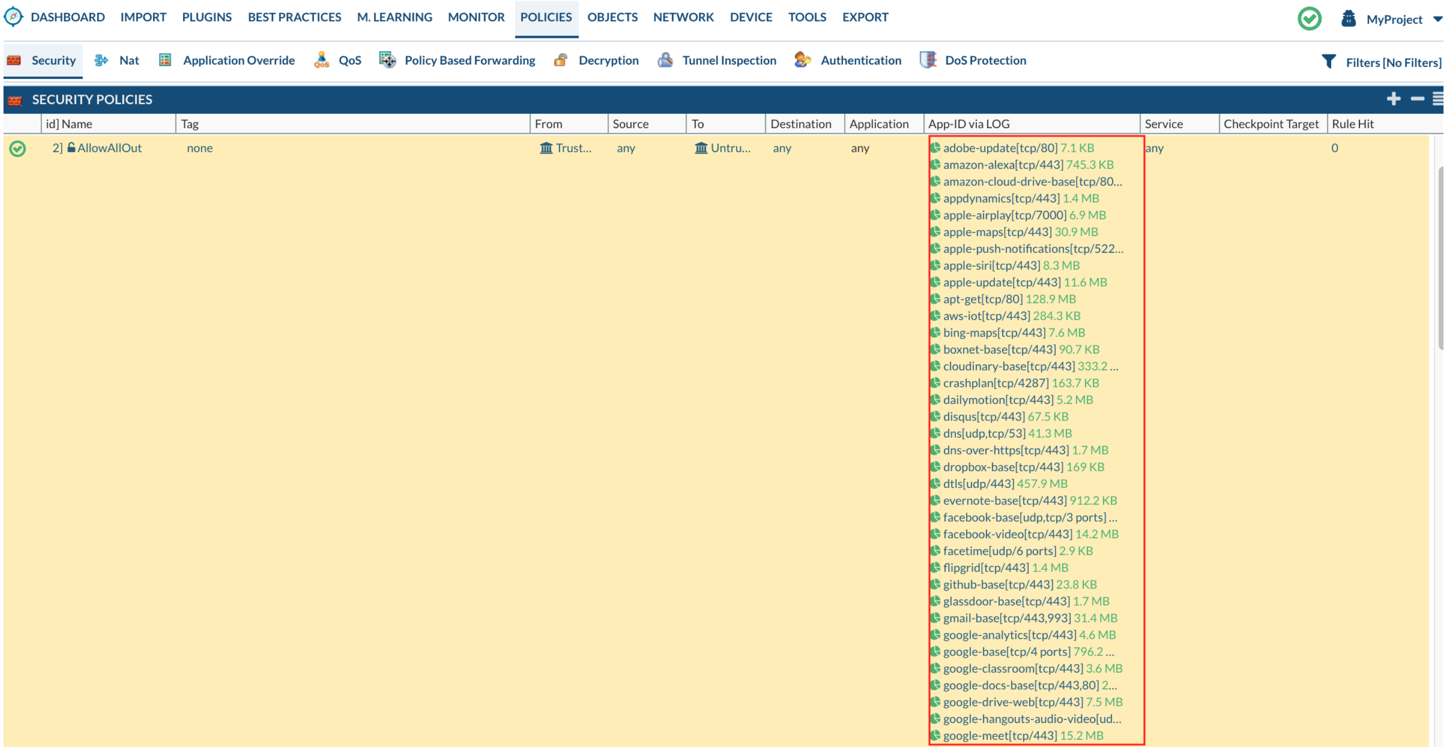
After the Retrieving APPs process finished , you will see a new column **"APP-ID via Log"** shown below and it listed all the APP-IDs found in the firewall traffic logs . Below is the result when you select "Retrieve Apps (Fast)" , the report will be generated faster without traffic size detailed, for an environment that has a lot of traffic logs, please select this option for faster processing.



id] Name	Tag	From	Source	To	Destination	Application	App-ID via LOG	Service	Checkpoint Target	Rule Hit
					@panw-...					
2] AllowAllOut	none	Trust...	any	Untru...	any	any	adobe-update[/0] 0 bytes amazon-alexa[/0] 0 bytes amazon-cloud-drive-base[/0] 0 b... appdynamics[/0] 0 bytes apple-airplay[/0] 0 bytes apple-maps[/0] 0 bytes apple-push-notifications[/0] 0 by... apple-siri[/0] 0 bytes apple-update[/0] 0 bytes apt-get[/0] 0 bytes aws-iot[/0] 0 bytes bing-maps[/0] 0 bytes boldchat-logmein[/0] 0 bytes boxnet-base[/0] 0 bytes cloudinary-base[/0] 0 bytes cortex-xdr[/0] 0 bytes crashplan[/0] 0 bytes dailymotion[/0] 0 bytes datadog[/0] 0 bytes disqus[/0] 0 bytes dns[/0] 0 bytes dns-over-https[/0] 0 bytes dropbox-base[/0] 0 bytes dtls[/0] 0 bytes evernote-base[/0] 0 bytes facebook-base[/0] 0 bytes facebook-video[/0] 0 bytes facetime[/0] 0 bytes flipgrid[/0] 0 bytes github-base[/0] 0 bytes glassdoor-base[/0] 0 bytes gmail-base[/0] 0 bytes google-analytics[/0] 0 bytes google-base[/0] 0 bytes google-classroom[/0] 0 bytes	any		0

1 | 50 | 1 - 3 of 3

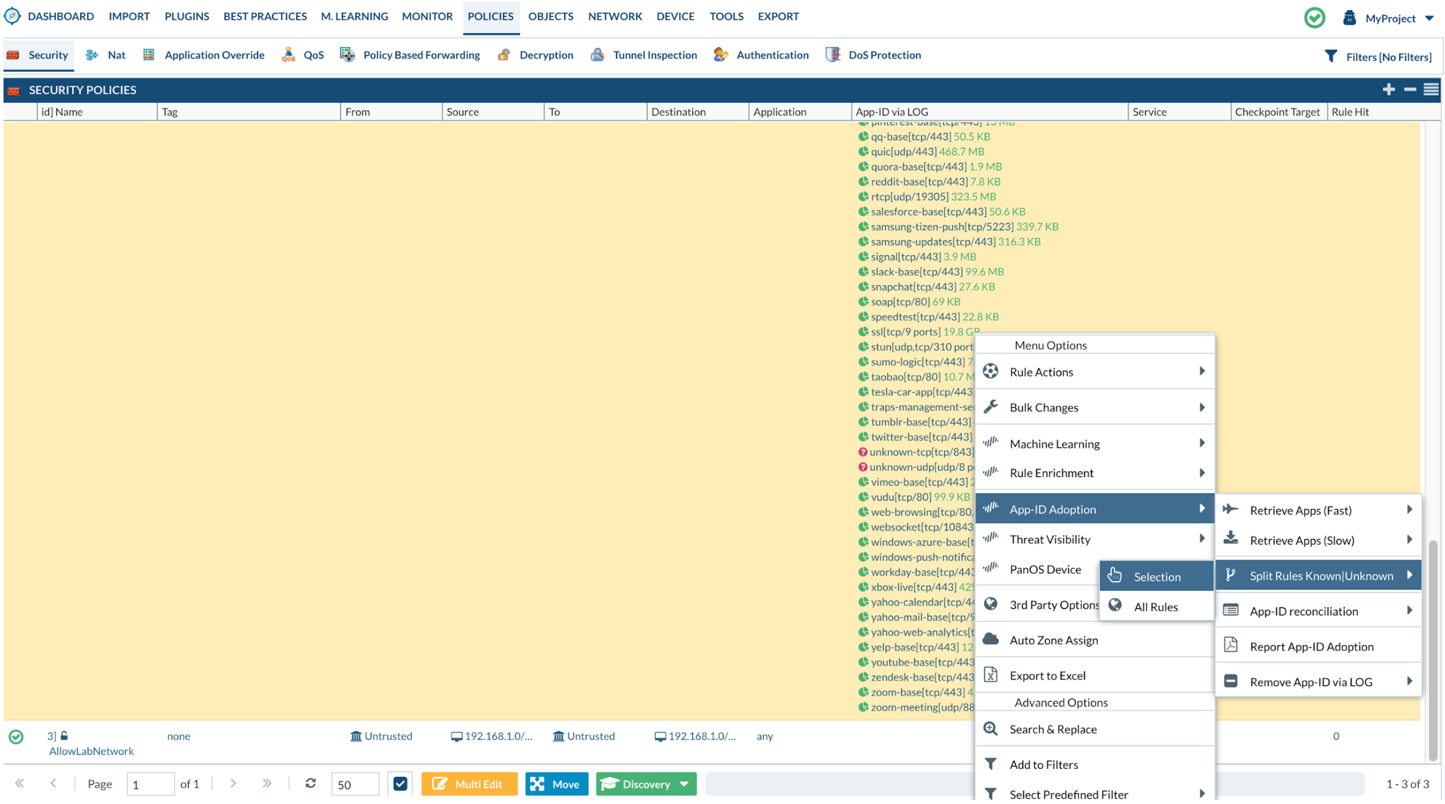
The difference between **"Retrieve APPs(Fast)"** and **"Retrieve APPs(Slow)"** are the **"Retrieve APPs(Slow)"** will take longer time to process and show you the detail traffic size based on APP-ID as shown in below screenshot:



id] Name	Tag	From	Source	To	Destination	Application	App-ID via LOG	Service	Checkpoint Target	Rule Hit
2] AllowAllOut	none	Trust...	any	Untru...	any	any	adobe-update[tcp/80] 7.1 KB amazon-alexa[tcp/443] 745.3 KB amazon-cloud-drive-base[tcp/80... appdynamics[tcp/443] 1.4 MB apple-airplay[tcp/7000] 6.9 MB apple-maps[tcp/443] 30.9 MB apple-push-notifications[tcp/522... apple-siri[tcp/443] 8.3 MB apple-update[tcp/443] 11.6 MB apt-get[tcp/80] 128.9 MB aws-iot[tcp/443] 284.3 KB bing-maps[tcp/443] 7.6 MB boxnet-base[tcp/443] 90.7 KB cloudinary-base[tcp/443] 333.2 ... crashplan[tcp/4287] 163.7 KB dailymotion[tcp/443] 5.2 MB disqus[tcp/443] 67.5 KB dns[udp,tcp/53] 41.3 MB dns-over-https[tcp/443] 1.7 MB dropbox-base[tcp/443] 169 KB dtls[udp/443] 457.9 MB evernote-base[tcp/443] 912.2 KB facebook-base[udp,tcp/3 ports] ... facebook-video[tcp/443] 14.2 MB facetime[udp/6 ports] 2.9 KB flipgrid[tcp/443] 1.4 MB github-base[tcp/443] 23.8 KB glassdoor-base[tcp/443] 1.7 MB gmail-base[tcp/443,993] 31.4 MB google-analytics[tcp/443] 4.6 MB google-base[tcp/4 ports] 796.2 ... google-classroom[tcp/443] 3.6 MB google-docs-base[tcp/443,80] 2... google-drive-web[tcp/443] 7.5 MB google-hangouts-audio-video[ud... google-meet[tcp/443] 15.2 MB	any		0

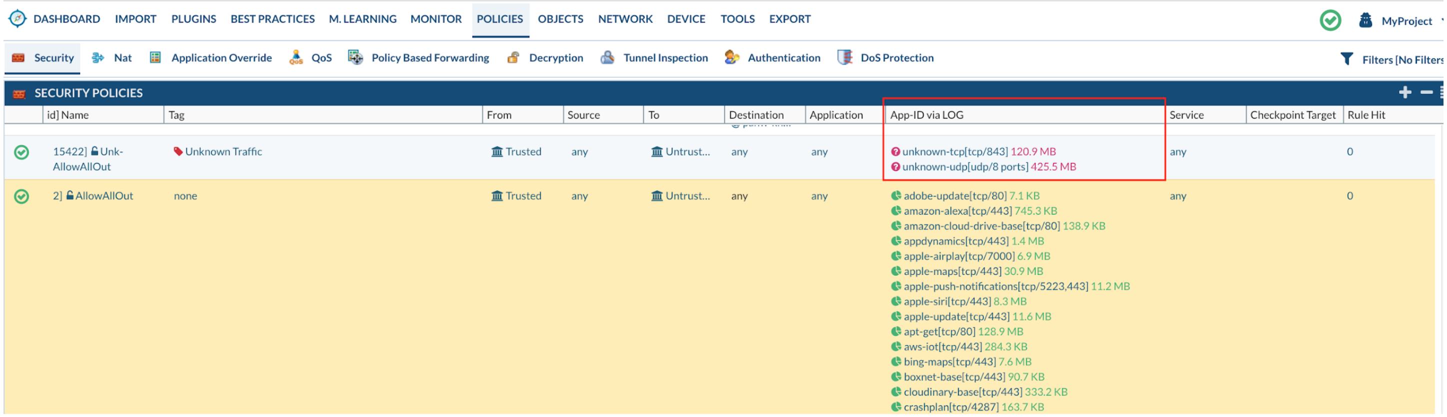
Split Rules Known/Unknown

When discovered APP-ID contains Unknown-TCP or Unknown-UDP, it's better to split them to different rules to further analyze the traffic logs later. To split the Known rules from Unknown rules , right click on the rules , select **"APP-ID Adoption" -> "Split Rules Known/Unknown" -> "Selection".**



	id] Name	Tag	From	Source	To	Destination	Application	App-ID via LOG	Service	Checkpoint Target	Rule Hit
								qq-base[tcp/443] 50.5 KB quic[udp/443] 468.7 MB quora-base[tcp/443] 1.9 MB reddit-base[tcp/443] 7.8 KB rtcp[udp/19305] 323.5 MB salesforce-base[tcp/443] 50.6 KB samsung-tizen-push[tcp/5223] 339.7 KB samsung-updates[tcp/443] 316.3 KB signal[tcp/443] 3.9 MB slack-base[tcp/443] 99.6 MB snapchat[tcp/443] 27.6 KB soap[tcp/80] 69 KB speedtest[tcp/443] 22.8 KB ssl[tcp/9 ports] 19.8 GB stun[udp,tcp/310 port sumo-logic[tcp/443] 7 taobao[tcp/80] 10.7 M tesla-car-app[tcp/443 traps-management-se tumblr-base[tcp/443] twitter-base[tcp/443] unknown-tcp[tcp/843] unknown-udp[udp/8 p vimeo-base[tcp/443] 2 vudu[tcp/80] 99.9 KB web-browsing[tcp/80, websocket[tcp/10843 windows-azure-base[t windows-push-notific workday-base[tcp/443 xbox-live[tcp/443] 42 yahoo-calendar[tcp/4 yahoo-mail-base[tcp/9 yahoo-web-analytics[t yelp-base[tcp/443] 12 youtube-base[tcp/443 zendesk-base[tcp/443 zoom-base[tcp/443] 4 zoom-meeting[udp/88			
✓	3] AllowLabNetwork	none	Untrusted	192.168.1.0/...	Untrusted	192.168.1.0/...	any				0

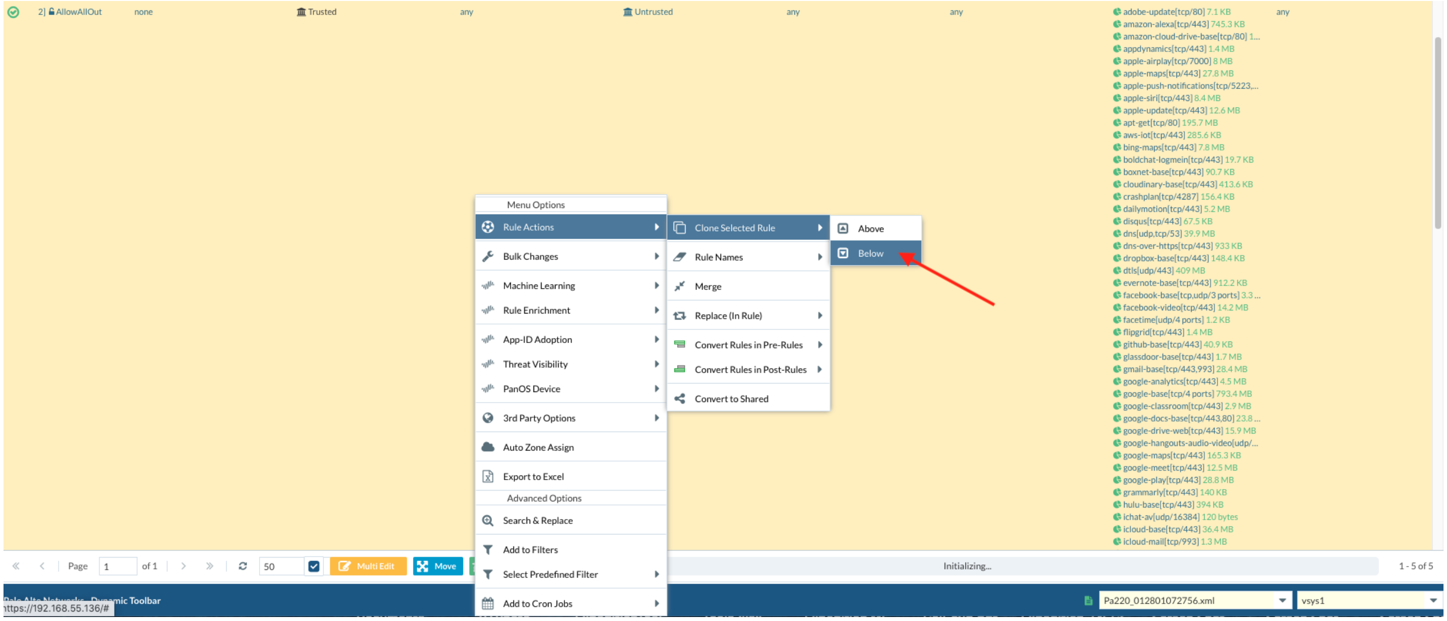
After you performed the step, You will see the original security policy got split into two rules , original rule will contain known APP-IDs in the **"APP-ID via Log"** field , Expedition cloned another rule on top of the original rule and add a prefix "Unk-" to the original rule name , the example below, the original rule name is "AllowAllOut", the "Unk-AllowAllOut" contains the **"Unknown-TCP"** and **"Unknown-UDP"** as shown in below screenshot:



	id] Name	Tag	From	Source	To	Destination	Application	App-ID via LOG	Service	Checkpoint Target	Rule Hit
✓	15422] Unk-AllowAllOut	Unknown Traffic	Trusted	any	Untrust...	any	any	unknown-tcp[tcp/843] 120.9 MB unknown-udp[udp/8 ports] 425.5 MB	any		0
✓	2] AllowAllOut	none	Trusted	any	Untrust...	any	any	adobe-update[tcp/80] 7.1 KB amazon-alexa[tcp/443] 745.3 KB amazon-cloud-drive-base[tcp/80] 138.9 KB appdynamics[tcp/443] 1.4 MB apple-airplay[tcp/7000] 6.9 MB apple-maps[tcp/443] 30.9 MB apple-push-notifications[tcp/5223,443] 11.2 MB apple-siri[tcp/443] 8.3 MB apple-update[tcp/443] 11.6 MB apt-get[tcp/80] 128.9 MB aws-iot[tcp/443] 284.3 KB bing-maps[tcp/443] 7.6 MB boxnet-base[tcp/443] 90.7 KB cloudinary-base[tcp/443] 333.2 KB crashplan[tcp/4287] 163.7 KB	any		0

Clone the existing rule

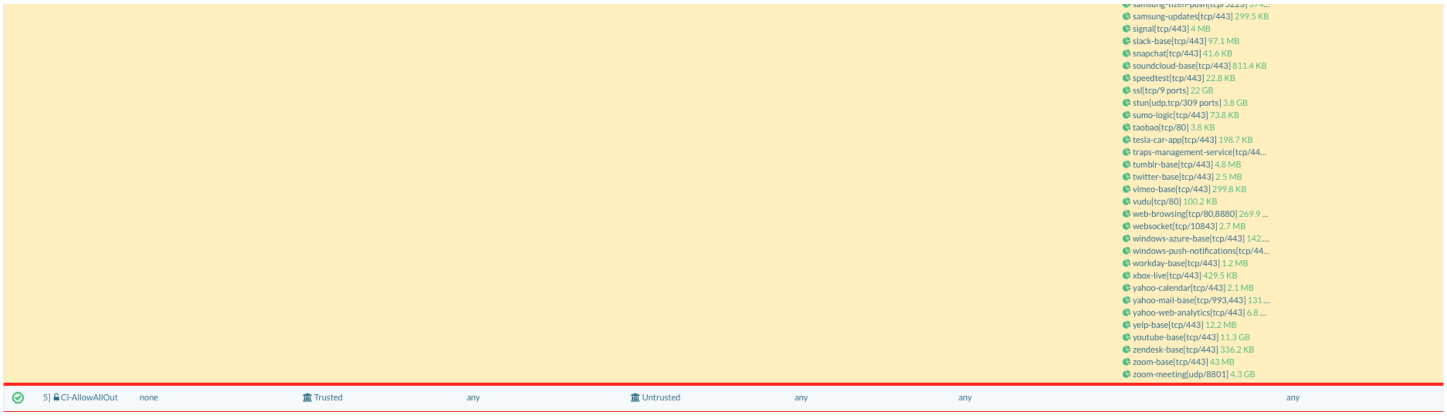
Right click on the rule that has discovered apps , select **“Rule Actions” -> “Cloned Selected Rule” -> “Below”.**



2| AllowAllOut none Trusted any Untrusted any any

adobe-update[tcp/80] 7.1 KB any
amazon-alexa[tcp/443] 745.3 KB
amazon-cloud-drive-base[tcp/80] 1...
appdynamics[tcp/443] 1.4 MB
apple-airplay[tcp/7000] 8 MB
apple-maps[tcp/443] 27.8 MB
apple-push-notifications[tcp/5223,...
apple-siri[tcp/443] 8.4 MB
apple-update[tcp/443] 12.6 MB
apt-get[tcp/80] 195.7 MB
aws-iot[tcp/443] 285.6 KB
bing-maps[tcp/443] 7.8 MB
boldchat-logmein[tcp/443] 19.7 KB
boxnet-base[tcp/443] 90.7 KB
cloudinary-base[tcp/443] 413.6 KB
crashplan[tcp/4287] 156.4 KB
dailymotion[tcp/443] 5.2 MB
disqus[tcp/443] 67.5 KB
dns[udp,tcp/53] 39.9 MB
dns-over-https[tcp/443] 933 KB
dropbox-base[tcp/443] 148.4 KB
dtls[udp/443] 409 MB
evernote-base[tcp/443] 912.2 KB
facebook-base[tcp,udp/3 ports] 3.3 ...
facebook-video[tcp/443] 14.2 MB
facetime[udp/4 ports] 1.2 KB
flipgrid[tcp/443] 1.4 MB
github-base[tcp/443] 40.9 KB
glassdoor-base[tcp/443] 1.7 MB
gmail-base[tcp/443,993] 28.4 MB
google-analytics[tcp/443] 4.5 MB
google-base[tcp/4 ports] 793.4 MB
google-classroom[tcp/443] 2.9 MB
google-docs-base[tcp/443,80] 23.8 ...
google-drive-web[tcp/443] 15.9 MB
google-hangouts-audio-video[udp/...
google-maps[tcp/443] 165.3 KB
google-meet[tcp/443] 12.5 MB
google-play[tcp/443] 28.8 MB
grammarly[tcp/443] 140 KB
hulu-base[tcp/443] 394 KB
ichat-av[udp/16384] 120 bytes
icloud-base[tcp/443] 36.4 MB
icloud-mail[tcp/993] 1.3 MB

This action will clone the existing rule and put it under the existing rule like below screenshot with prefix “CL-”, the example below, the original rule name is “AllowAllOut”, the new cloned rule will be named “CL-AllowAllOut”. This is identical to the original service-port based rule. The purpose of this step is to keep the original rule in the bottom of the APP-ID based rule.



samsung-tizen-push[tcp/5223] 374...
samsung-updates[tcp/443] 299.5 KB
signal[tcp/443] 4 MB
slack-base[tcp/443] 97.1 MB
snapchat[tcp/443] 41.6 KB
soundcloud-base[tcp/443] 811.4 KB
speedtest[tcp/443] 22.8 KB
ssl[tcp/9 ports] 22 GB
stun[udp,tcp/309 ports] 3.8 GB
sumo-logic[tcp/443] 73.8 KB
taobao[tcp/80] 3.8 KB
tesla-car-app[tcp/443] 198.7 KB
traps-management-service[tcp/44...
tumblr-base[tcp/443] 4.8 MB
twitter-base[tcp/443] 2.5 MB
vimeo-base[tcp/443] 299.8 KB
vudu[tcp/80] 100.2 KB
web-browsing[tcp/80,8880] 269.9 ...
websocket[tcp/10843] 2.7 MB
windows-azure-base[tcp/443] 142....
windows-push-notifications[tcp/44...
workday-base[tcp/443] 1.2 MB
xbox-live[tcp/443] 429.5 KB
yahoo-calendar[tcp/443] 2.1 MB
yahoo-mail-base[tcp/993,443] 131....
yahoo-web-analytics[tcp/443] 6.8 ...
yelp-base[tcp/443] 12.2 MB
youtube-base[tcp/443] 11.3 GB
zendesk-base[tcp/443] 336.2 KB
zoom-base[tcp/443] 43 MB
zoom-meeting[udp/8801] 4.3 GB

5| Cl-AllowAllOut none Trusted any Untrusted any any any

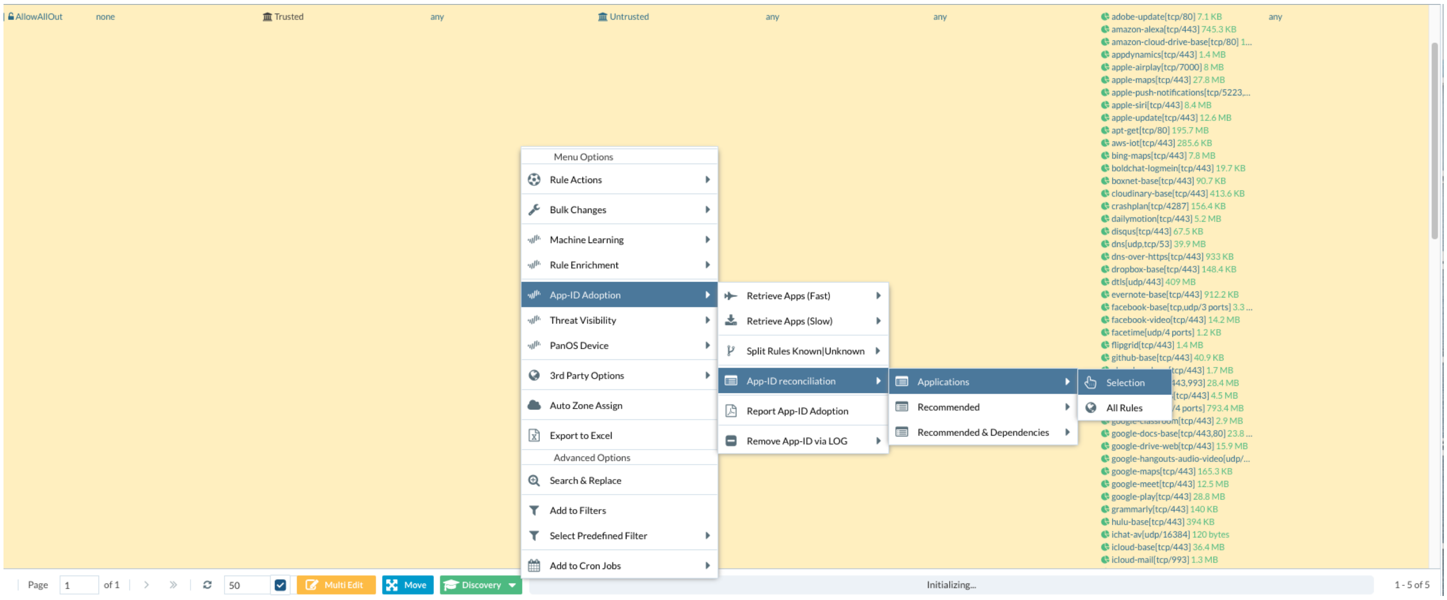
APP-ID Reconciliation

We will highlight the original security policy , right click to select **“APP-ID Adoption” -> “APP-ID Reconciliation” -> “Applications” -> “Selection”** . This step is to add the app-ID that's in the **“APP-ID via Log “** field to the **“Application”** field of the rule, so it is converting the existing service-port based rule to “APP-ID” based rule. You have three options in this process:

1. **Application** - The application field will be replaced with APP-IDs seen in the “APP-ID via Log” field.

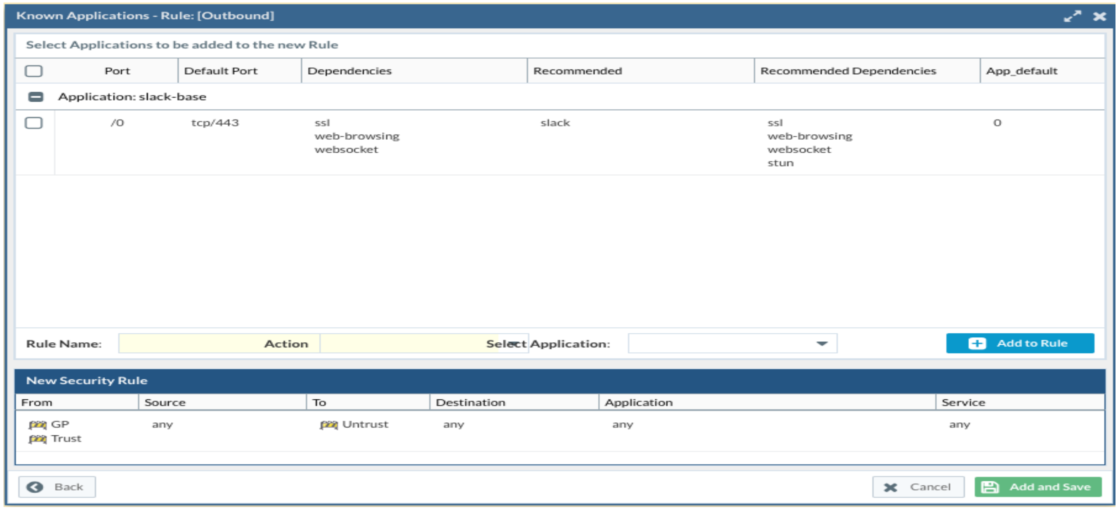
2. **Recommended Application-** The application field will be replaced with recommended APP-IDs. For instance if the APP-ID seen in the traffic log is “slack-base”, the recommended APP-ID will be **“slack”** which is a parent APP-ID that covers **“slack-base”** and other APP-IDs under the same parent.

3. **Recommended Application & Dependencies - -** The application field will be replaced with recommended APP-ID and it’s dependencies , for example : if “**gmail-chat**” is seen in the “APP-ID via log” field , and it will add it’s parent application **“gmail”** also the dependencies , **“google-base”** to the rule as well. You can find APP-ID dependencies by looking up the Application database in the PAN-OS device.



AllowAllOut	none	Trusted	any	Untrusted	any	any	adobe-update[tcp/80] 7.1 KB amazon-alexa[tcp/443] 745.3 KB amazon-cloud-drive-base[tcp/80] 1... appdynamics[tcp/443] 1.4 MB apple-airplay[tcp/7000] 8 MB apple-maps[tcp/443] 27.8 MB apple-push-notifications[tcp/5223,... apple-siri[tcp/443] 8.4 MB apple-update[tcp/443] 12.6 MB apt-get[tcp/80] 195.7 MB aws-iot[tcp/443] 285.6 KB bing-maps[tcp/443] 7.8 MB boldchat-logmein[tcp/443] 19.7 KB boxnet-base[tcp/443] 90.7 KB cloudinary-base[tcp/443] 413.6 KB crashplan[tcp/4287] 156.4 KB dailymotion[tcp/443] 5.2 MB disqus[tcp/443] 67.5 KB dns[udp,tcp/53] 39.9 MB dns-over-https[tcp/443] 933 KB dropbox-base[tcp/443] 148.4 KB dtls[udp/443] 409 MB evernote-base[tcp/443] 912.2 KB facebook-base[tcp,udp/3 ports] 3.3 ... facebook-video[tcp/443] 14.2 MB facetime[udp/4 ports] 1.2 KB flipgrid[tcp/443] 1.4 MB github-base[tcp/443] 40.9 KB [tcp/443] 1.7 MB 443,993] 28.4 MB [tcp/443] 4.5 MB /4 ports] 793.4 MB m[tcp/443] 2.9 MB google-docs-base[tcp/443,80] 23.8 ... google-drive-web[tcp/443] 15.9 MB google-hangouts-audio-video[udp/... google-maps[tcp/443] 165.3 KB google-meet[tcp/443] 12.5 MB google-play[tcp/443] 28.8 MB grammarly[tcp/443] 140 KB hulu-base[tcp/443] 394 KB ichat-av[udp/16384] 120 bytes icloud-base[tcp/443] 36.4 MB icloud-mail[tcp/993] 1.3 MB	any

Recommended Application:



Known Applications - Rule: [Outbound]

Select Applications to be added to the new Rule

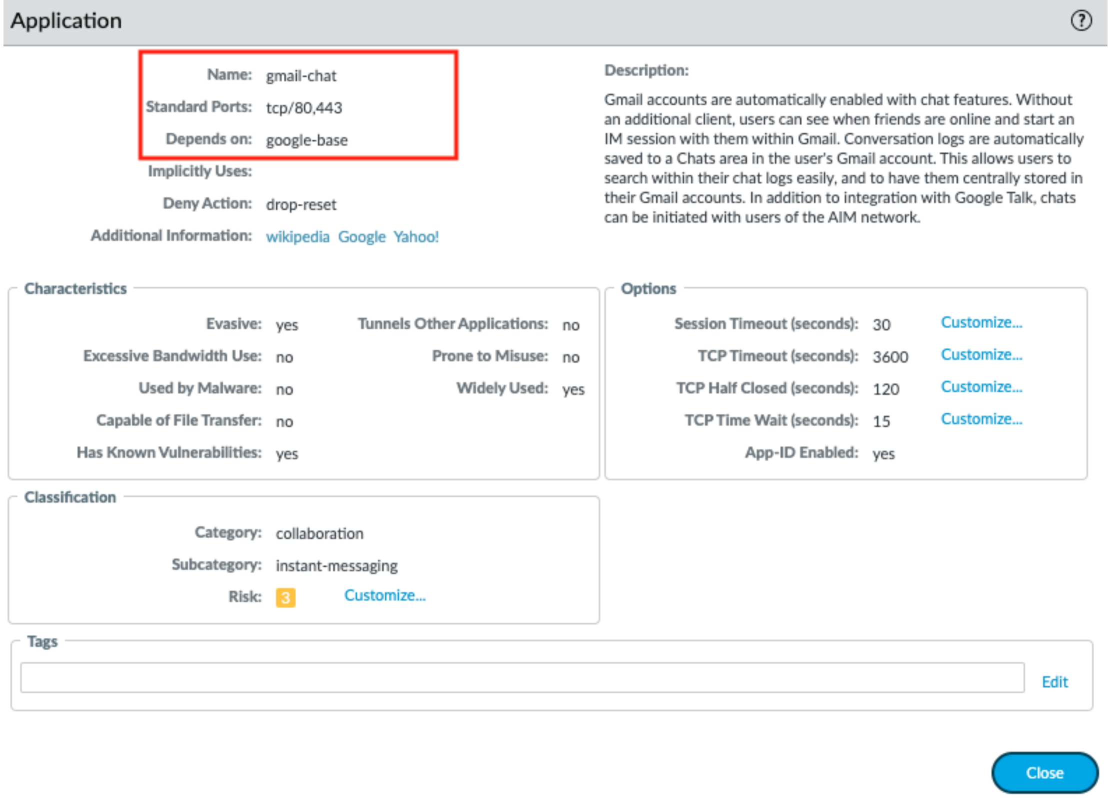
☐	Port	Default Port	Dependencies	Recommended	Recommended Dependencies	App_default
⊟	Application: slack-base					
☐	/0	tcp/443	ssl web-browsing websocket	slack	ssl web-browsing websocket stun	0

Rule Name: Action Select Application:

New Security Rule

From	Source	To	Destination	Application	Service
GP Trust	any	Untrust	any	any	any

Recommended Application and dependencies in PAN-OS application:



Application

Name: gmail-chat

Standard Ports: tcp/80,443

Depends on: google-base

Implicitly Uses:

Deny Action: drop-reset

Additional Information: wikipedia Google Yahoo!

Description:

Gmail accounts are automatically enabled with chat features. Without an additional client, users can see when friends are online and start an IM session with them within Gmail. Conversation logs are automatically saved to a Chats area in the user's Gmail account. This allows users to search within their chat logs easily, and to have them centrally stored in their Gmail accounts. In addition to integration with Google Talk, chats can be initiated with users of the AIM network.

Characteristics

Evasive: yes

Excessive Bandwidth Use: no

Used by Malware: no

Capable of File Transfer: no

Has Known Vulnerabilities: yes

Tunnels Other Applications: no

Prone to Misuse: no

Widely Used: yes

Options

Session Timeout (seconds): 30 Customize...

TCP Timeout (seconds): 3600 Customize...

TCP Half Closed (seconds): 120 Customize...

TCP Time Wait (seconds): 15 Customize...

App-ID Enabled: yes

Classification

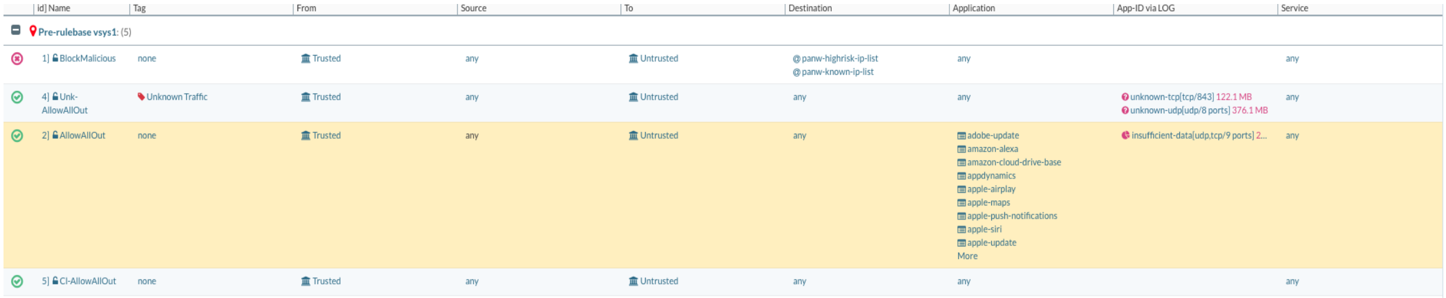
Category: collaboration

Subcategory: instant-messaging

Risk: 3 Customize...

Tags

After the step is performed , you will see the original service-port based rule become a APP-ID based rule as shown in below screenshot, and the APP-IDs shown in APP-ID via Log field will be move to “Application” field of the rule , the **“insufficient-data”** won’t be move to Application field , this is seen when Firewall does not have enough packets to determine the APP-ID.



id] Name	Tag	From	Source	To	Destination	Application	App-ID via LOG	Service
Pre-rulebase vsys1: (5)								
1] BlockMalicious	none	Trusted	any	Untrusted	@panw-highrisk-ip-list @panw-known-ip-list	any		any
4] Unk-AllowAllOut	Unknown Traffic	Trusted	any	Untrusted	any	any	unknown-tcp[tcp/843] 122.1 MB unknown-udp[udp/8 ports] 376.1 MB	any
2] AllowAllOut	none	Trusted	any	Untrusted	any	adobe-update amazon-alexa amazon-cloud-drive-base appdynamics apple-airplay apple-maps apple-push-notifications apple-siri apple-update More	insufficient-data[udp,tcp/9 ports] 2...	any
5] CI-AllowAllOut	none	Trusted	any	Untrusted	any	any		any

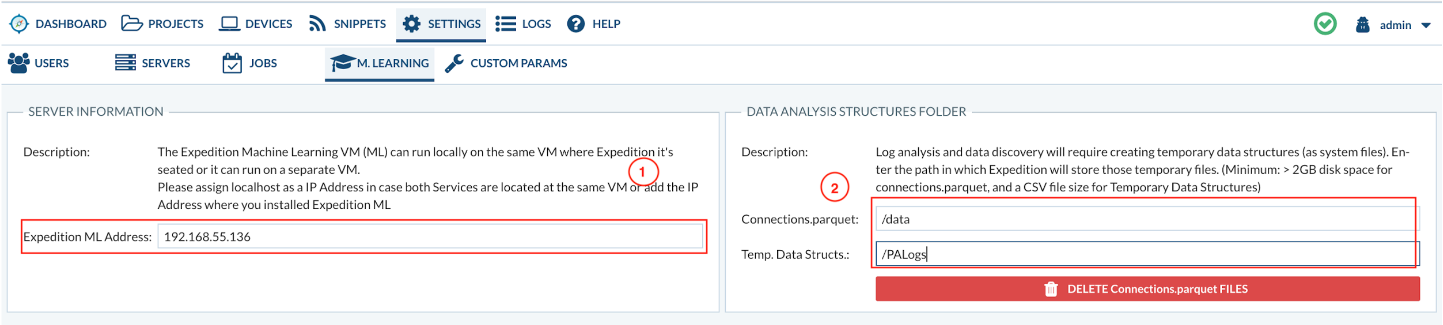
Common Steps for Machine Learning (ML) and Rule Enrichment (RE)

- ❏ Configure Machine Learning settings
- ❏ Importing traffic logs to Expedition
- ❏ Configure M.Learning on the device
- ❏ Processing the logs from the selected PAN-OS device

Machine Learning (ML) Settings

As we described in the overview, we can set up Expedition to run standalone or splitting in GUI and Analysis, that is only the case if we have a shared Expedition VM with more hardware assigned to run faster analysis.

To configure it and specify where the Parquet files will be stored after the log process we have to navigate from the Dashboard to the SETTINGS tab, then select the M.LEARNING tab.



SERVER INFORMATION

Description: The Expedition Machine Learning VM (ML) can run locally on the same VM where Expedition it's seated or it can run on a separate VM.
Please assign localhost as a IP Address in case both Services are located at the same VM or add the IP Address where you installed Expedition ML

Expedition ML Address: 192.168.55.136

DATA ANALYSIS STRUCTURES FOLDER

Description: Log analysis and data discovery will require creating temporary data structures (as system files). Enter the path in which Expedition will store those temporary files. (Minimum: > 2GB disk space for connections.parquet, and a CSV file size for Temporary Data Structures)

Connections.parquet: /data

Temp. Data Structs.: /PALogs

Step 1: Check the IP address shown in Expedition ML Address is your own IP address if you are running in standalone mode. If not, put the IP address of the other Expedition VM instance with more hardware resources than the one you are currently configuring.

Every time you enter here if the ip address is 127.0.0.1 Expedition will replace it by the one shown in the browser's URL.

Step 2: Type where you want to store the data after the process in parquet files. You could specify a different path to store your firewall traffic logs. If you don't have any folder created in Expedition let's create one, in this case we will use one named PALogs.

```
sudo mkdir /PALogs
sudo chown -R www-data.www-data /PALogs
```

with these commands we are creating and allowing "apache" to have write access to it.

Step 3: click on the Save button located at the bottom bar – right to activate the changes.

If at some point we want to start fresh and remove all the data processed as parquet files we can click on the red button "DELETE ALL DATA STRUCTURED FILES". That process cannot be undone.

Importing Logs to Expedition

There are below four ways to import traffic logs to Expedition:

1. Exporting from NGFW
2. Manually Export Logs From MONITOR
3. Expedition as Syslog server
4. Importing from Splunk

Depending on your platform and or your log storage strategy, you may want to choose between those four methods. For instance, you may have a large amount of firewalls deployed on prem and on cloud environments that you would like to analyze and that are managed by a Panorama device. In such a case, probably the best strategy would be to use Expedition as a Syslog server and request the firewalls to submit the traffic logs to Expedition via a Syslog log forwarding profile.
Maybe you already are having a Syslog log forwarding profile on your security rules to store the traffic logs in a Splunk instance. In that case, maybe the best approach would be to select the fourth option.

Below we describe the four alternatives and discuss the benefits of using each one of them.

Exporting from NGFW

There is a functionality in Palo Alto Networks Firewalls that allows you to automatically export the logs on a daily basis and import it automatically on Expedition, that functionality is called "Scheduled log export".

Note: The PA-7000 Series and Panorama devices do not offer such log Export functionality or it is limited to the first 1.000.000 entries

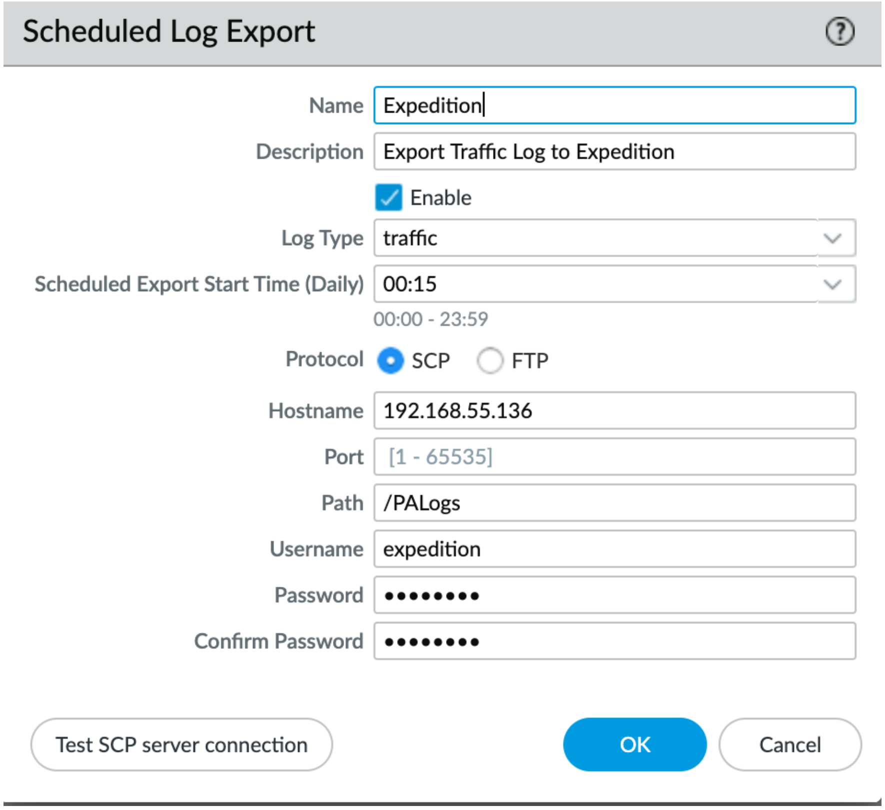
https://docs.paloaltonetworks.com/pan-os/10-0/pan-os-web-interface-help/device/device-scheduled-log-export.html

Actions

We will configure the Scheduled Log Export with the following settings.

- The log type to select is "*traffic*";
- we will set the Hostname with our *Expedition IP* or hostname (if resolved by your DNS server);
- the start time could become something like "00:15" or a time that you consider suitable for initiating a SFTP transfer that could be between few MBs up to a couple or TBs (depending on your infrastructure traffic);
- we will set the destination folder in which traffic logs will be stored in Expedition. We recommend using the folder /PALogs, which may already exist in your Expedition instance, and allows the users "*expedition*" and "*www-data*" to access it;
- we will set the username as "expedition", which it is an Expedition cli user that we have by default created in the Expedition VMs and that has privileges to store files in required file system spaces; and
- we will set up the password as "*paloalto*" or the defined password that you have for the *expedition* user in your Expedition VM.

When you create this for the first time you should run the "Test SCP Server connection" twice. The first time in order to exchange the SSH keys and the second to write a test file and validate that the PanOS Firewall can write in the folder provided under "*path*".



Scheduled Log Export

Name	Expedition
Description	Export Traffic Log to Expedition
	☑ Enable
Log Type	traffic
Scheduled Export Start Time (Daily)	00:15 00:00 - 23:59
Protocol	● SCP ○ FTP
Hostname	192.168.55.136
Port	[1 - 65535]
Path	/PALogs
Username	expedition
Password	••••••••
Confirm Password	••••••••

As this folder could need a lot of storage in the beginning it is a good practice to add a new VM DISK and mount it under /PALogs. Be sure Expedition can write if you chose to use expedition as the user to write the data. This can be done with the following command:

```
chown -R expedition.www-data /PALogs
```

Benefits

On the positive side, this action is performed only once a day and can be programmed to execute during the night or other low load period of the day.

Additionally, we can be confident that all the traffic logs will be submitted to our Expedition if there is enough space in our Expedition VM to host the file, as the transfer is performed via a SFTP connection.

You do not need to have the VM running all the time to receive the traffic log files. If you have set up the scheduled time to be at 10:00am, this gives you the chance to have the Expedition VM off during night hours and bring it up again before the SCP is going to take place.

Drawbacks

This approach requires firewall configuration changes, in order to set up a periodic log export.

Manually Export Logs From MONITOR

You can always go into any firewall from Palo Alto Networks and from the Monitor tab export the logs in CSV format and upload that CSV file to Expedition for processing, concretely, into your "*PALogs*" folder.

The format of those CSV files would be the same as executing a Scheduled Log Export.

Actions

Log into your PANOS device and, under the Monitor tab, enter the filter that you would like to use for collecting traffic log entries. For instance, specify the rule name and destinations that you would like to study.

Afterwards, click on the “Export to CSV” button, which resembles an Excel icon, to download the filtered traffic logs into your computer.
Afterwards, upload the obtained CSV into the /PALogs folder in your Expedition by using the SFTP client of your choice.

Benefits
This option offers us an opportunity to collect specific logs from the firewall, selecting a period of time, specific sources, destinations, and/or rule names that we would like to bring into Expedition and further study.

Drawbacks
However, this approach would not be suitable for a periodic task, as it requires manual connection to the Firewall, set the filter logic we want to apply, export the log in a CSV format and manually upload it into the “*/PALogs*” folder

Expedition as Syslog Server

In case the scheduled approach would not fit with your requirements, either because you are not interested in secure copying all the traffic log entries into Expedition, or because you are dealing with a platform that limits to 1 million the number of entries to be exported, then you may prefer using a *syslog message* approach.

Expedition can become a Syslog server that can receive the logs generated by platforms like the PA7000 or in any case you want. And we can submit traffic log entries to our Expedition using a *log forwarding profile*.

Actions
Become root in your Expedition instance to navigate to the following folder.

```
sudo su
cd /var/www/html/OS/rsyslog
```

In that folder you will find 3 examples for the rsyslog.conf you can use:

- rsyslog.default-tcp,
- rsyslog.default-tcpudp and
- rsyslog.default-udp

Each one listen in the specified protocol and port and allow Expedition to accept syslog from some specified networks (in the examples we define 127.0.0.1, 10.11.29.0/24, 172.16.26.0/24, *.paloaltonetworks.com) and store them in “/PALogs”.
Restart the VM to make sure that all required syslog modules are activated. Restarting the rsyslog service may not be enough as dependent modules may not have been initialized with your required settings.

Benefits
Using Expedition as a syslog server may help you determine the security rules that you are interested in studying, as we can determine the rules that are going to have the log forwarding profile applied.

Drawbacks
This option requires you to activate Expedition as a syslog server and to tune the rsyslog settings to allow connections via TCP and/or UDP from a specific set of IP addresses.
This option requires your Expedition instance to be available 24x7, as traffic logs are submitted to your Expedition in real-time. If the syslog entries are defined to be submitted using UDP protocol, there are higher chances that some traffic logs may be lost if the connection with Expedition is interrupted.

If you already have log forwarding profiles applied to your security policy rules, you may have to generate different combinations of log forwarding profiles. Rules can only have one log forwarding profile applied, even this profile can combine multiple syslog servers. Therefore, you may have to create a new log forwarding profile that contains your prior server as well as the Expedition syslog server, and apply this new log forwarding profile into the rules of your choice by replacing the prior profile applied.

Importing from Splunk

Finally, there is a chance that you have already log forwarding profiles active in your devices to submit the traffic logs to an Splunk instance. We have prepared Expedition to be able to collect traffic logs information for Splunk, generating the API queries that would be required to access those entries.
As we do not require the whole amount of information that traffic logs contain, the queries that we generate to interact with Splunk have already been constructed to request only the needed log fields and to create aggregations already in Splunk in order to reduce the amount of data that needs to be submitted from Splunk to Expedition.

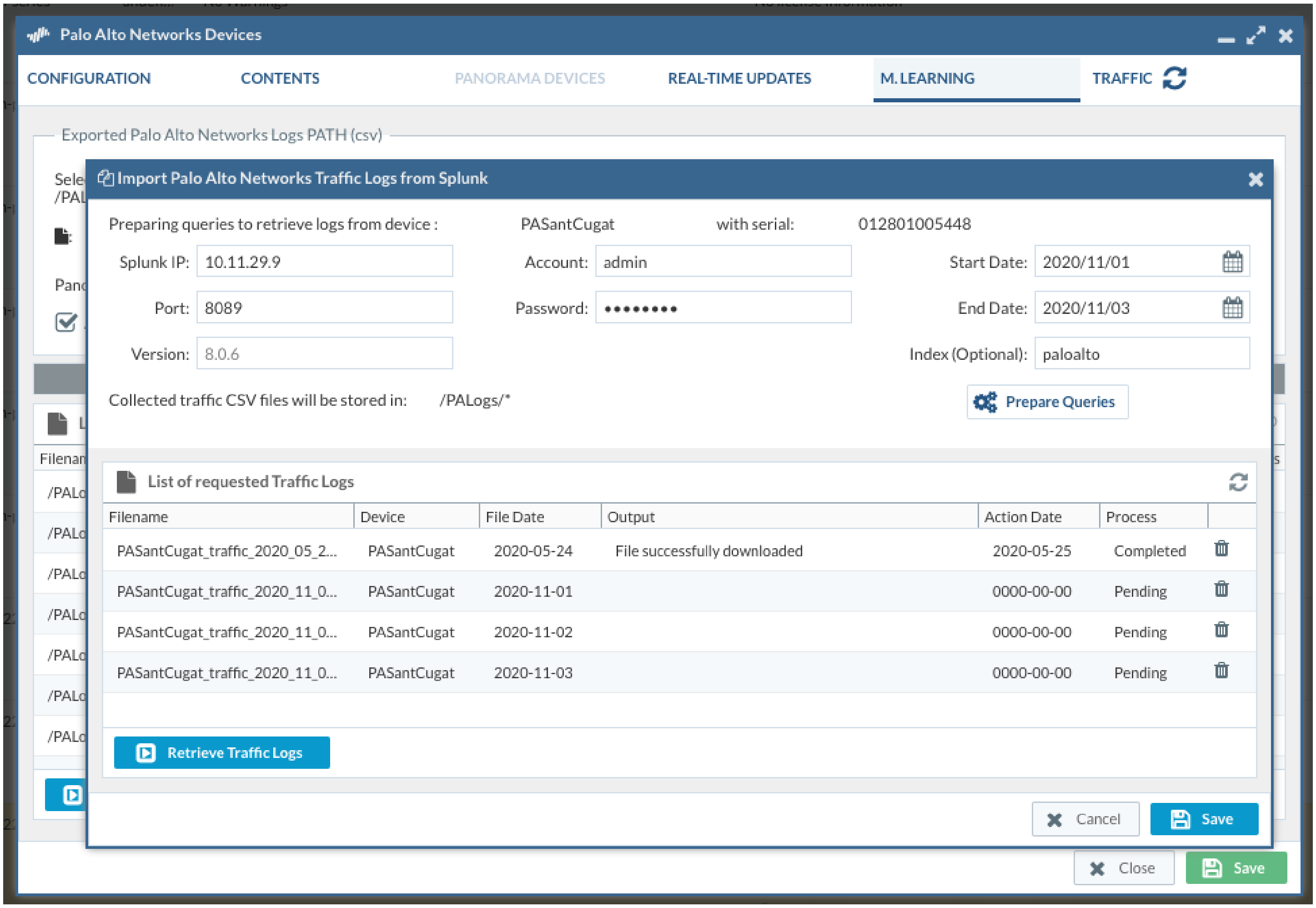
Actions
First, make sure that your Splunk instance has the **"Palo Alto Networks App for Splunk"** and **"Palo Alto Networks Add-On" in**stalled, as this will provide the required schema to query Palo Alto Networks logs. For more details on the apps, please refer to the link: https://splunk.paloaltonetworks.com/

From Expedition UI, select the device that you would like to retrieve traffic logs from. On the *M.Learning* tab, click on the "*Query Logs from Splunk®*" button to open the Splunk query window.

Fill in the form to generate the individual queries that will be sent to the Splunk instance. In the figure below, we have provided the IP address of our Splunk instance using the default's Splunk port 8089 and using the "*admin*" account that has API permits. We have defined the period of time we want to collect data from, in the example, between the first and the third of November 2020 and using the "*paloalto*" index to increase the performance on Splunk while loading the log files that refer to Palo Alto Networks devices.

To generate the queries, click on the "Prepare Queries" button. This will create one query per day and those will be presented in the bottom grid. At this point, we can fire the queries to Splunk by clicking on the "*Retrieve Traffic Logs"* button.

At this point, the queries will be submitted to splunk and Expedition will be monitoring the query status to download the results once Splunk has reported the query as completed. Depending on your Splunk settings and the number of queries submitted, some of them may be in a queue.
Once the query has been completed in Splunk, and Expedition has downloaded the results, you will be presented with the message "**File successfully downloaded"**, and you should be able to display the file in the prior *M.Learning* view.



Import Palo Alto Networks Traffic Logs from Splunk

Preparing queries to retrieve logs from device : PASantCugat with serial: 012801005448

Splunk IP: 10.11.29.9
Account: admin
Start Date: 2020/11/01
Port: 8089
Password: ••••••••
End Date: 2020/11/03
Version: 8.0.6
Index (Optional): paloalto

Collected traffic CSV files will be stored in: /PALogs/*

List of requested Traffic Logs

Filename	Device	File Date	Output	Action Date	Process
PASantCugat_traffic_2020_05_2...	PASantCugat	2020-05-24	File successfully downloaded	2020-05-25	Completed
PASantCugat_traffic_2020_11_0...	PASantCugat	2020-11-01		0000-00-00	Pending
PASantCugat_traffic_2020_11_0...	PASantCugat	2020-11-02		0000-00-00	Pending
PASantCugat_traffic_2020_11_0...	PASantCugat	2020-11-03		0000-00-00	Pending

Benefits

There is no need to interact with your Palo Alto Networks devices, as those are already submitting the traffic logs to Splunk. instead, the interactions are performed via API with your Splunk instance.

If you have indexes created, requests can make use of them to speed up the process.

We can specify the periods of time that we want to study, and only retrieve the traffic logs that were generated during those dates.

Drawbacks

It requires a user account with API querying permission in order to interact with Splunk.

Currently, this approach offers manual query generation and execution and it needs to be executed for every firewall that we want to collect traffic logs from.

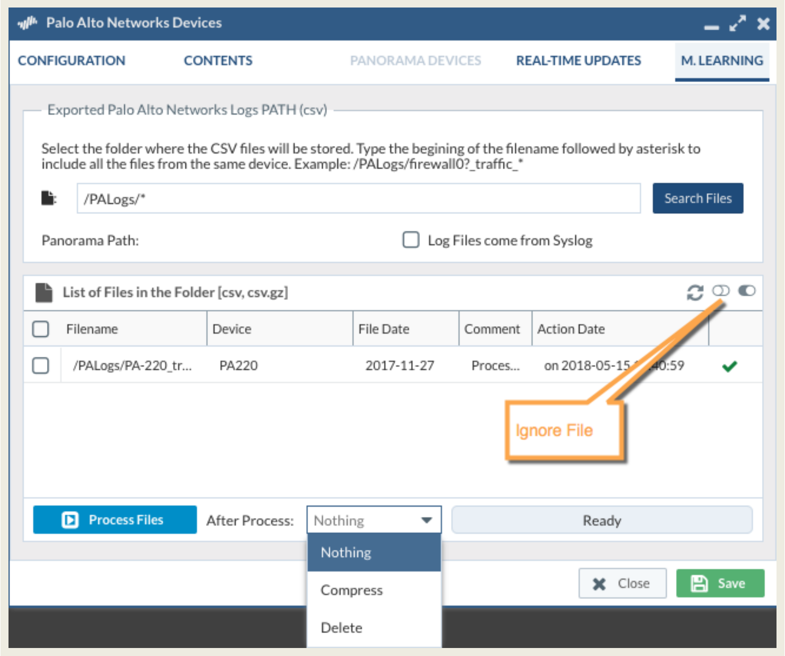
In the future, we will provide opportunities to retrieve Device Groups traffic logs directly from the Panorama point of view.

Configure M. LEARNING on the device

Once the device has been added in the previous step, we have to edit it and go to the M.Learning Tab.

From here we will assign the PATH where our logs have been imported in Expedition, or via SCP or SYSLOG, then click on Search Files to allow Expedition to search on it and show us all the related logs to the device we are configuring.

Expedition will filter out and show you only the files from the configured folder (and recursively from there) only if the serial number configured on the device (or the HA Serial) matches the serial seen in the csv files. So, if you configure the device and you set the serial "123" but the CSV content talks about the serial number "567" ; those logs won't be shown under the view.



Exported Palo Alto Networks Logs PATH (csv)

Select the folder where the CSV files will be stored. Type the begining of the filename followed by asterisk to include all the files from the same device. Example: /PALogs/firewall0?_traffic_*

/PALogs/*

Panorama Path:

□ Log Files come from Syslog

List of Files in the Folder [csv, csv.gz]

□	Filename	Device	File Date	Comment	Action Date	
□	/PALogs/PA-220_tr...	PA220	2017-11-27	Proces...	on 2018-05-15 [illegible]40:59	✓

After Process: Nothing

By default, the first time Expedition sees a new file is flagged as a NEW file. When clicking on Process files all the NEW files will be converted to Parquet.

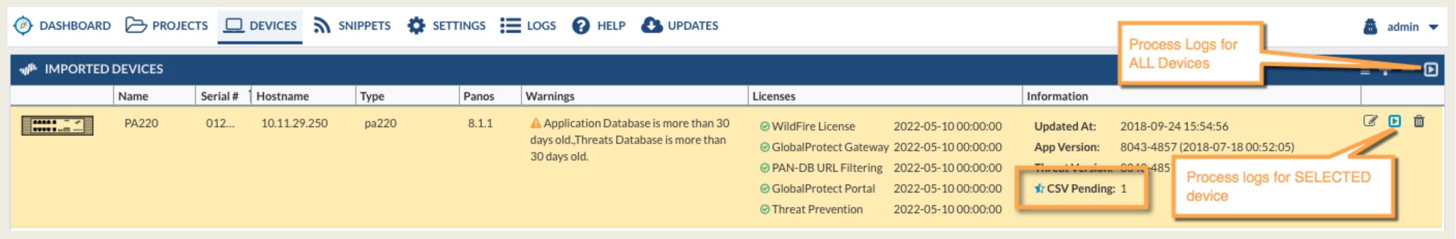
We can select an action for "After Process" like we want to delete the files after transformation to parquet to reduce the space we need to store the raw data.

If we want to avoid processing some file we have to just select it and click the icon (Ignore File), that will tell Expedition to skip that file at the Process time.

In case you are using SYSLOG to receive logs in Expedition check the "Log Files come from Syslog" this will prevent you to transform the logs from Today since it's a file is still opened and used by the syslog server before the log rotation that will occur when we change the day at 00:00, so this check will hide the log from Today until Tomorrow.

Processing Logs From Selected Devices

We can enter on each firewall where we are importing more files for the process and click on “Process Logs” or we can do it from the Devices view by selecting the Device and clicking on the blue Icon.



Name	Serial #	Hostname	Type	Panos	Warnings	Licenses	Information
PA220	012...	10.11.29.250	pa220	8.1.1	Application Database is more than 30 days old.,Threats Database is more than 30 days old.	WildFire License 2022-05-10 00:00:00 GlobalProtect Gateway 2022-05-10 00:00:00 PAN-DB URL Filtering 2022-05-10 00:00:00 GlobalProtect Portal 2022-05-10 00:00:00 Threat Prevention 2022-05-10 00:00:00	**Updated At:** 2018-09-24 15:54:56 **App Version:** 8043-4857 (2018-07-18 00:52:05) **Threat Version:** 8043-485 **CSV Pending:** 1

Process Logs for ALL Devices

Process logs for SELECTED device

Debugging Log Processing

While Expedition is processing the logs, we can view the progress and potential issues from the CLI by doing this:

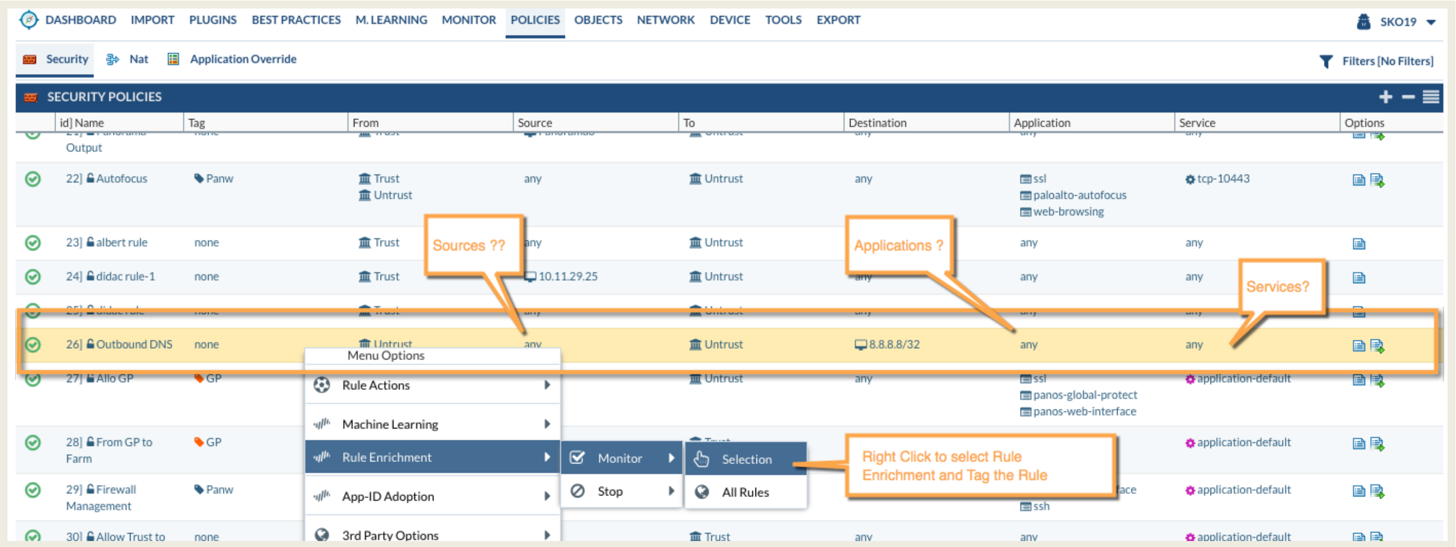
```
tail -f /tmp/error_logCoCo
```

Rule Enrichment(RE)

As we described before RE it's useful to reduce the surface attack in your current security policies, today with RE we can remove all the “any” from our Rules. One of the most popular use cases is to help to enrich a policy based on services with App-ID for instance.

- ❏ Enable Rule Enrichment
- ❏ Rule Enrichment Discovery
- ❏ Import Discovered Rules back to Project

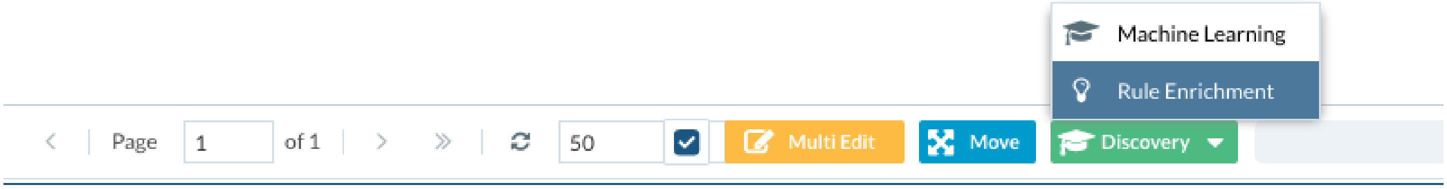
Enable Rule Enrichment



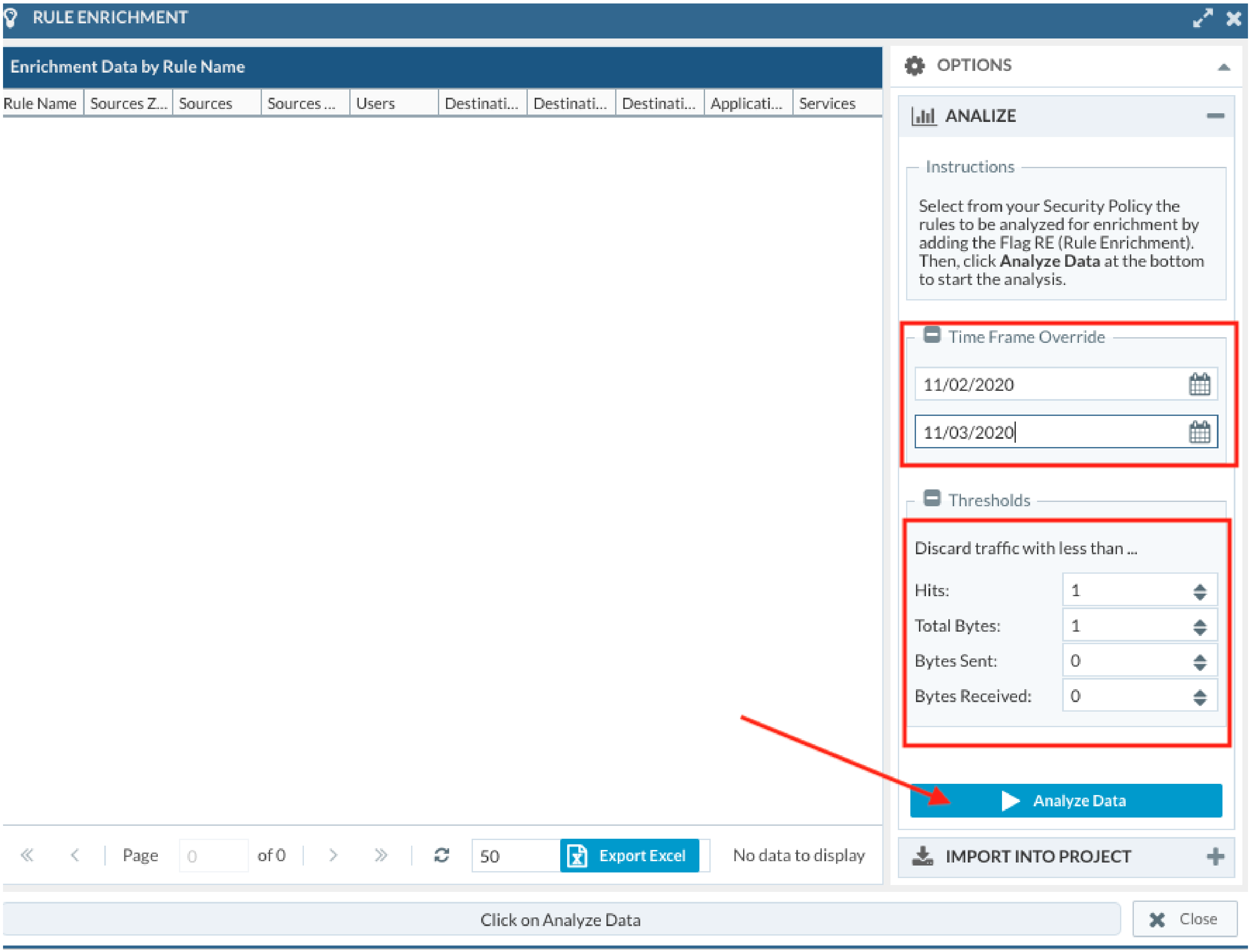
id] Name	Tag	From	Source	To	Destination	Application	Service	Options
21] Panorama Output	none	Trust	Panorama	Untrust	any	any	any	
22] Autofocus	Panw	Trust Untrust	any	Untrust	any	ssl paloalto-autofocus web-browsing	tcp-10443	
23] albert rule	none	Trust	any	Untrust	any	any	any	
24] didac rule-1	none	Trust	10.11.29.25	Untrust	any	any	any	
25] didac rule	none	Trust	any	Untrust	any	any	any	
26] Outbound DNS	none	Untrust	any	Untrust	8.8.8.8/32	any	any	
27] Allo GP	GP			Untrust	any	ssl panos-global-protect panos-web-interface	application-default	
28] From GP to Farm	GP			Trust			application-default	
29] Firewall Management	Panw					ace ssh	application-default	
30] Allow Trust to	none			Trust	any	any	application-default	

In this example we have Rule Enrichment Discovery

After tag the Rule click on **DISCOVERY** green bottom located in the bottom bar (center)



A new window will pop up. Click the TAB called **RULE ENRICHMENT**. Then you have the option to override the Time Range defined previously in the LOG CONNECTOR, if you define like the last 30 days in the connector you can select a specific time frame like example and that will take preference from the connector configuration. Also you can specify the thresholds to discard traffic when it's less than specific bytes or hits. Then click on "**Analyze Data**" and wait



Enrichment Data by Rule Name

Rule Name	Sources Z...	Sources	Sources ...	Users	Destinati...	Destinati...	Destinati...	Applicati...	Services

No data to display

Instructions

Select from your Security Policy the rules to be analyzed for enrichment by adding the Flag RE (Rule Enrichment). Then, click **Analyze Data** at the bottom to start the analysis.

Time Frame Override

11/02/2020

11/03/2020

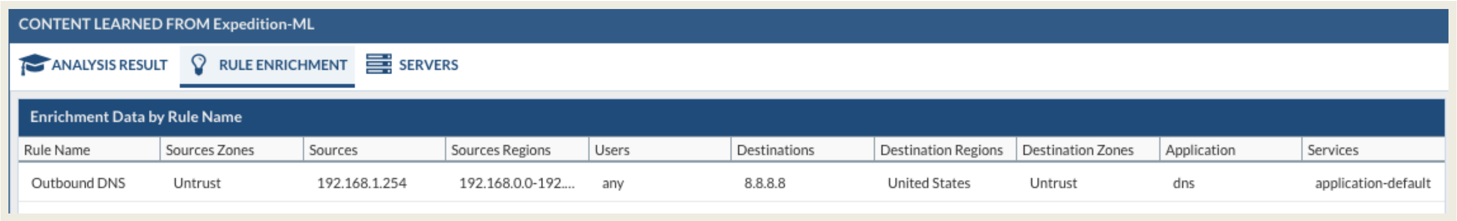
Thresholds

Discard traffic with less than ...

Hits:	1
Total Bytes:	1
Bytes Sent:	0
Bytes Received:	0

Click on Analyze Data

Once the analysis has finished we will get something like this, this information is based on the logs previously processed and stored as parquet and filtered by the log connector and the name of the rule tagged as RE (you can tag as much rules as you want and perform the analysis at the same time for all the rules, each one will create its own results)



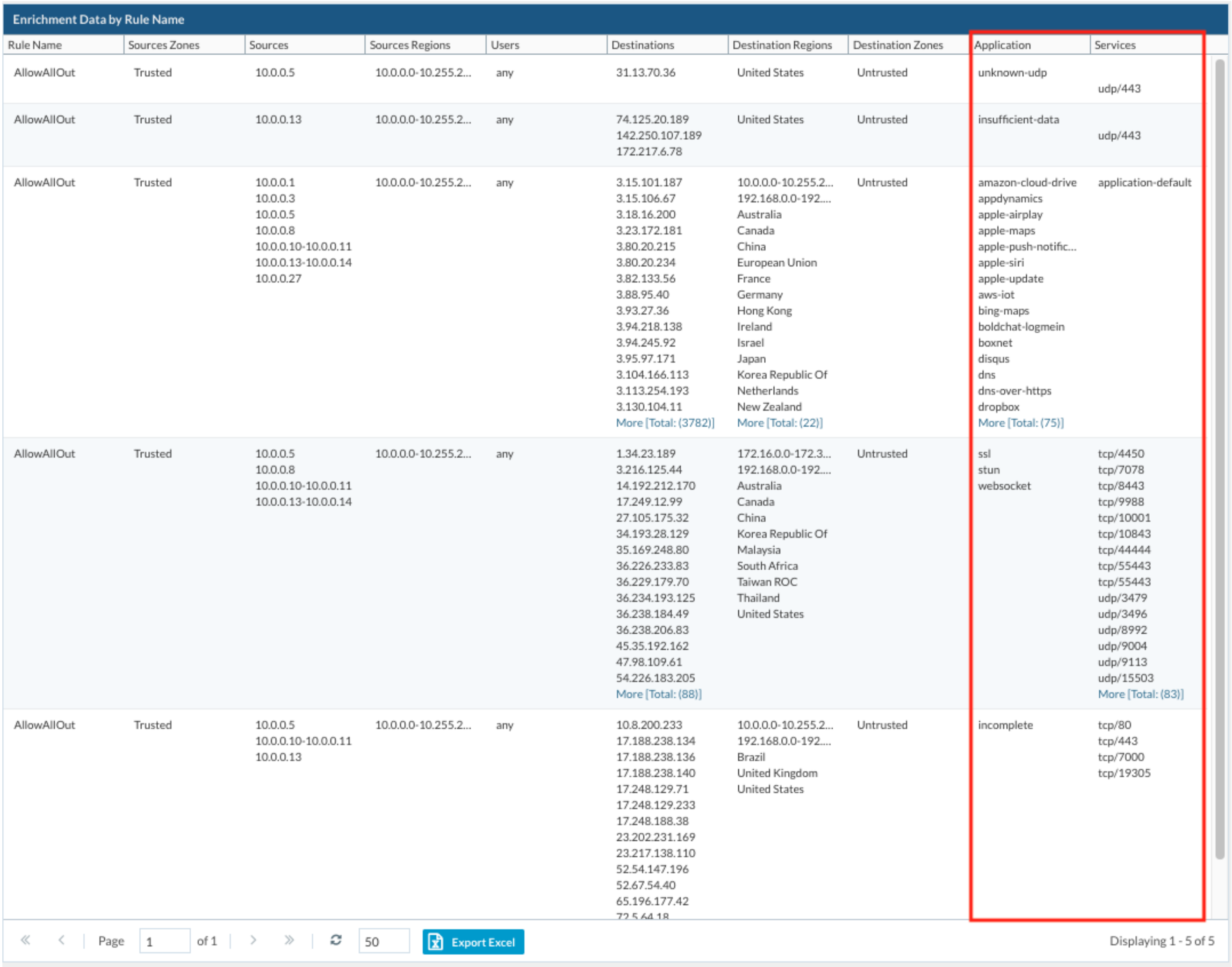
Enrichment Data by Rule Name

Rule Name	Sources Zones	Sources	Sources Regions	Users	Destinations	Destination Regions	Destination Zones	Application	Services
Outbound DNS	Untrust	192.168.1.254	192.168.0.0-192....	any	8.8.8.8	United States	Untrust	dns	application-default

From the output we can read there is only one Source IP address, I can read the applications (dns) and Expedition has calculated if the port seen using the applications is the same as defined as application-default to tell us the service can be considered “application-default” if cannot be identified you will see the port discovered (in this example should be udp/43)

The major difference between RE and ML is RE produced less rules than ML, it will only produce 4 -6 rules based on Application and service ports , the rule will be group to below 6 categories based on below:

1. application is Unknown-udp
2. application is Unknown-tcp
3. application is insufficient data
4. application with application-default port
5. application with custom service port.
6. application is incomplete



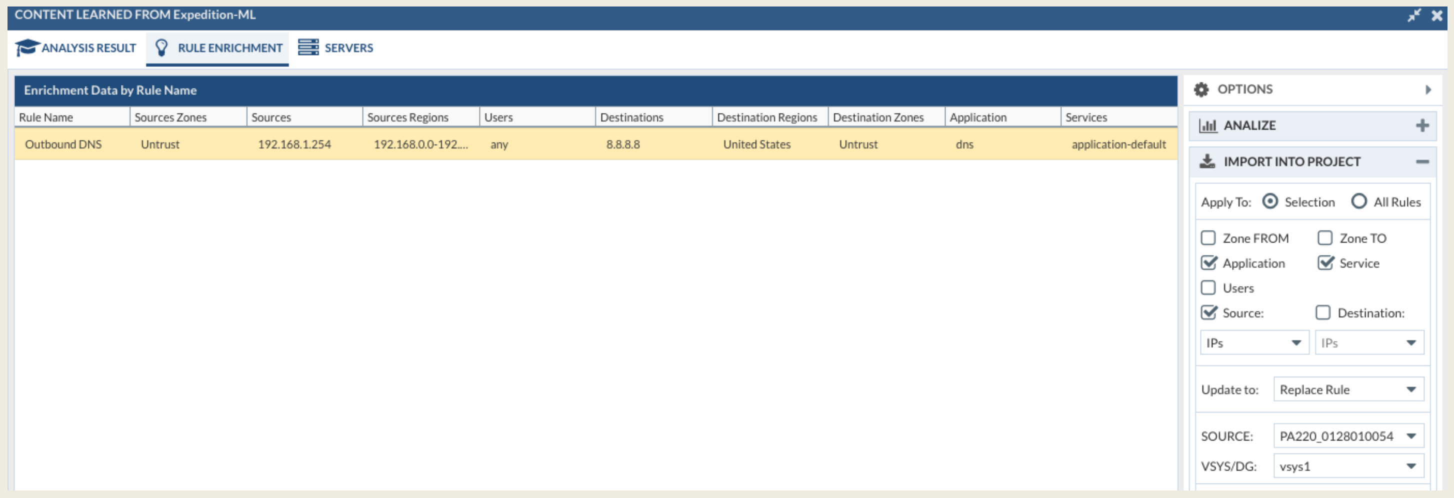
Enrichment Data by Rule Name

Rule Name	Sources Zones	Sources	Sources Regions	Users	Destinations	Destination Regions	Destination Zones	Application	Services
AllowAllOut	Trusted	10.0.0.5	10.0.0.0-10.255.2...	any	31.13.70.36	United States	Untrusted	unknown-udp	udp/443
AllowAllOut	Trusted	10.0.0.13	10.0.0.0-10.255.2...	any	74.125.20.189 142.250.107.189 172.217.6.78	United States	Untrusted	insufficient-data	udp/443
AllowAllOut	Trusted	10.0.0.1 10.0.0.3 10.0.0.5 10.0.0.8 10.0.0.10-10.0.0.11 10.0.0.13-10.0.0.14 10.0.0.27	10.0.0.0-10.255.2...	any	3.15.101.187 3.15.106.67 3.18.16.200 3.23.172.181 3.80.20.215 3.80.20.234 3.82.133.56 3.88.95.40 3.93.27.36 3.94.218.138 3.94.245.92 3.95.97.171 3.104.166.113 3.113.254.193 3.130.104.11 More [Total: (3782)]	10.0.0.0-10.255.2... 192.168.0.0-192.... Australia Canada China European Union France Germany Hong Kong Ireland Israel Japan Korea Republic Of Netherlands New Zealand More [Total: (22)]	Untrusted	amazon-cloud-drive appdynamics apple-airplay apple-maps apple-push-notific... apple-siri apple-update aws-iot bing-maps boldchat-logmein boxnet disqus dns dns-over-https dropbox More [Total: (75)]	application-default
AllowAllOut	Trusted	10.0.0.5 10.0.0.8 10.0.0.10-10.0.0.11 10.0.0.13-10.0.0.14	10.0.0.0-10.255.2...	any	1.34.23.189 3.216.125.44 14.192.212.170 17.249.12.99 27.105.175.32 34.193.28.129 35.169.248.80 36.226.233.83 36.229.179.70 36.234.193.125 36.238.184.49 36.238.206.83 45.35.192.162 47.98.109.61 54.226.183.205 More [Total: (88)]	172.16.0.0-172.3... 192.168.0.0-192.... Australia Canada China Korea Republic Of Malaysia South Africa Taiwan ROC Thailand United States	Untrusted	ssl stun websocket	tcp/4450 tcp/7078 tcp/8443 tcp/9988 tcp/10001 tcp/10843 tcp/44444 tcp/55443 tcp/55443 udp/3479 udp/3496 udp/8992 udp/9004 udp/9113 udp/15503 More [Total: (83)]
AllowAllOut	Trusted	10.0.0.5 10.0.0.10-10.0.0.11 10.0.0.13	10.0.0.0-10.255.2...	any	10.8.200.233 17.188.238.134 17.188.238.136 17.188.238.140 17.248.129.71 17.248.129.233 17.248.188.38 23.202.231.169 23.217.138.110 52.54.147.196 52.67.54.40 65.196.177.42 72.5.64.18	10.0.0.0-10.255.2... 192.168.0.0-192.... Brazil United Kingdom United States	Untrusted	incomplete	tcp/80 tcp/443 tcp/7000 tcp/19305

Import Discovered Rules back to Project

To import the information to our Rule we have to click on the option **“IMPORT INTO PROJECT”** under the **“Analyze Data”** button.

I will select Applications, Service and Source, for the Source I will check IPs that means bring the IP address, I can choose the regions found instead.



CONTENT LEARNED FROM Expedition-ML

ANALYSIS RESULT | RULE ENRICHMENT | SERVERS

Enrichment Data by Rule Name

Rule Name	Sources Zones	Sources	Sources Regions	Users	Destinations	Destination Regions	Destination Zones	Application	Services
Outbound DNS	Untrust	192.168.1.254	192.168.0.0-192....	any	8.8.8.8	United States	Untrust	dns	application-default

OPTIONS

ANALIZE

IMPORT INTO PROJECT

Apply To: ● Selection ○ All Rules

☐ Zone FROM ☐ Zone TO
☑ Application ☑ Service
☐ Users
☑ Source: ☐ Destination:
IPs IPs

Update to: Replace Rule

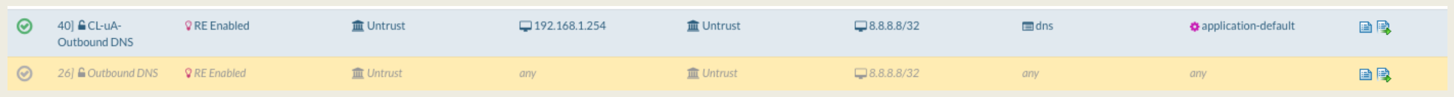
SOURCE: PA220_0128010054

VSYS/DG: vsys1

We can override the current (original) rule with the new data to improve that rule or we can let Expedition clone the existing rule and import the new data on the new cloned rule, that cloned rule will be stored on top of the original one. This will allow you to run more analysis in the future to ensure you don’t miss anything because the period of time analyzed was not enough to capture all the data.

Click on “**Import”.**

How we left “Replace Rule” instead of cloning, Expedition will automatically Clone the Original Rule, bring the new data on top and disable the current Rule for you to validate that everything looks good after the enrichment, you can remove the disabled rule at the end of the process.



40] CL-uA-Outbound DNS	RE Enabled	Untrust	192.168.1.254	Untrust	8.8.8.8/32	dns	application-default
26] Outbound DNS	RE Enabled	Untrust	any	Untrust	8.8.8.8/32	any	any

When we use Rule Enrichment Expedition will group all the data by Rule Name and it will create a rule with all the traffic seen with Users, another one with the traffic seen without users, another one with the traffic where the applications where found through their default-port and another one with the traffic where the applications were found on a port different the default port that means with Rule Enrichment we will get a maximum of 4 rules by each rule we are analyzing.

Debug the Rule Enrichment

```
tail -f /tmp/error_SecRulesEnrich
```

on the red button “DELETE ALL DATA STRUCTURED FILES”. That process cannot be undone.

MACHINE LEARNING (ML)

Machine Learning is a more sophisticated process than RE. ML was designed to help companies to know what is going on in the network and translate that to security policies based from the beginning on App-ID and User-ID. The most common case we cover is a “Greenfield” where you have one rule covering all the traffic and Expedition suggests all the new rules that cover the traffic seen.

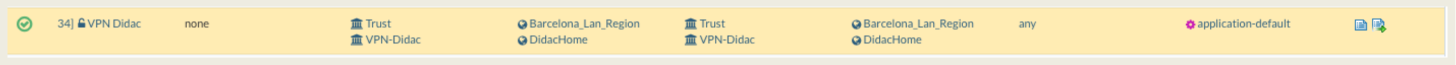
In the ML process Expedition identifies the Servers in the networks based on some analysis since we support environments with asymmetric traffic this analysis is critical at the time to suggest new rules. Here are the specific steps for ML:

- ❏ Enable Machine Learning
- ❏ Machine Learning Discovery
- ❏ Import Suggested Rules back to Project

Enable Machine Learning

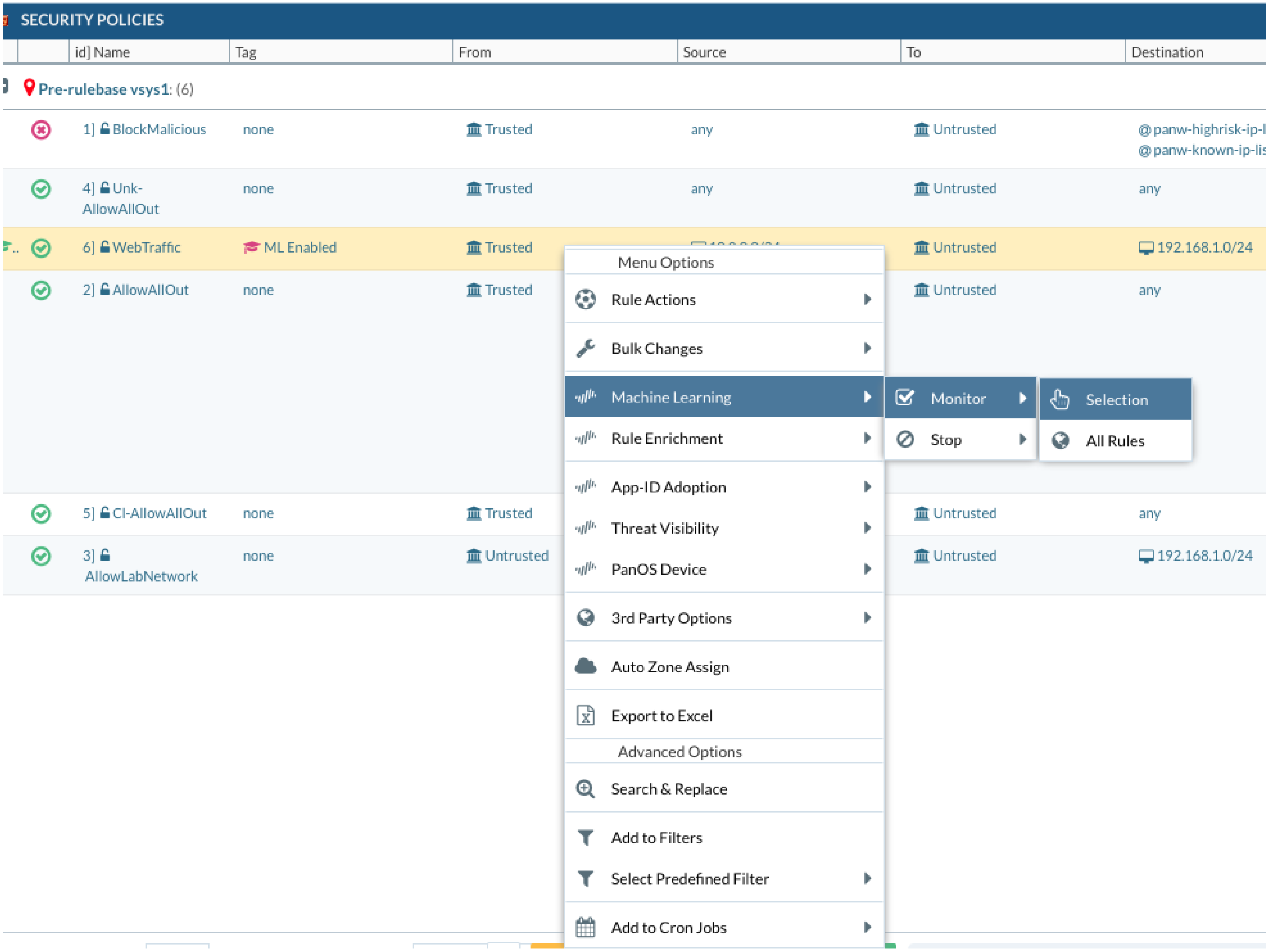
Example like below , we want to replace below one rule that is too wide open since it allows everything between two networks.

We want to know what is exactly happening and create the rules based on what we have learned.



✓	34] VPN Didac	none	Trust VPN-Didac	Barcelona_Lan_Region DidacHome	Trust VPN-Didac	Barcelona_Lan_Region DidacHome	any	application-default	

First step is to enable the “Machine Learning feature “ by highlighting the rule and right-clicking to select “**Machine Learning” -> “Monitor” -> “Selection”** .



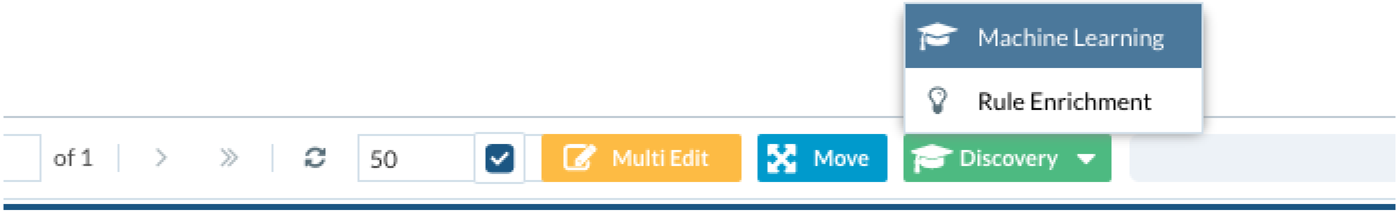
SECURITY POLICIES

Pre-rulebase vsys1: (6)

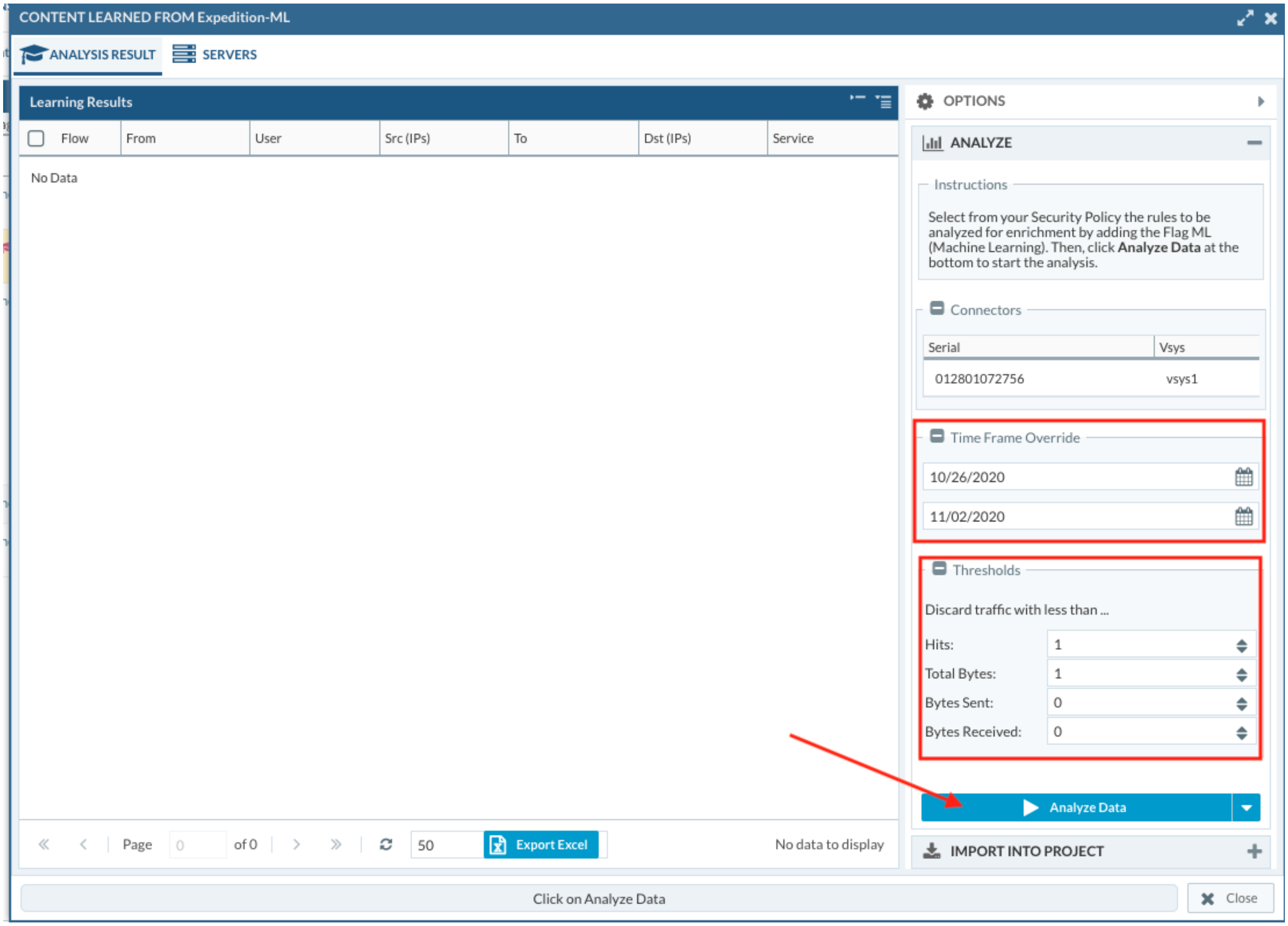
id] Name	Tag	From	Source	To	Destination
1] BlockMalicious	none	Trusted	any	Untrusted	@panw-highrisk-ip-l @panw-known-ip-lis
4] Unk-AllowAllOut	none	Trusted	any	Untrusted	any
6] WebTraffic	ML Enabled	Trusted		Untrusted	192.168.1.0/24
2] AllowAllOut	none	Trusted		Untrusted	any
5] CI-AllowAllOut	none	Trusted		Untrusted	any
3] AllowLabNetwork	none	Untrusted		Untrusted	192.168.1.0/24

Machine Learning Discovery

After tag the Rule click on **DISCOVERY** green bottom located in the bottom bar (center) , select **“Machine Learning”**



A new window will show up and we can override the time to analyze overriding the one configured in the LOG CONNECTOR (remember to check you have a LOG CONNECTOR configured and Active). Also you can select to discard traffic if it's less than bytes or less than certain hit counts you specified in the threshold.



☐ Flow	From	User	Src (IPs)	To	Dst (IPs)	Service
No Data						

No data to display

Instructions

Select from your Security Policy the rules to be analyzed for enrichment by adding the Flag ML (Machine Learning). Then, click **Analyze Data** at the bottom to start the analysis.

Connectors

Serial	Vsys
012801072756	vsys1

Time Frame Override

10/26/2020

11/02/2020

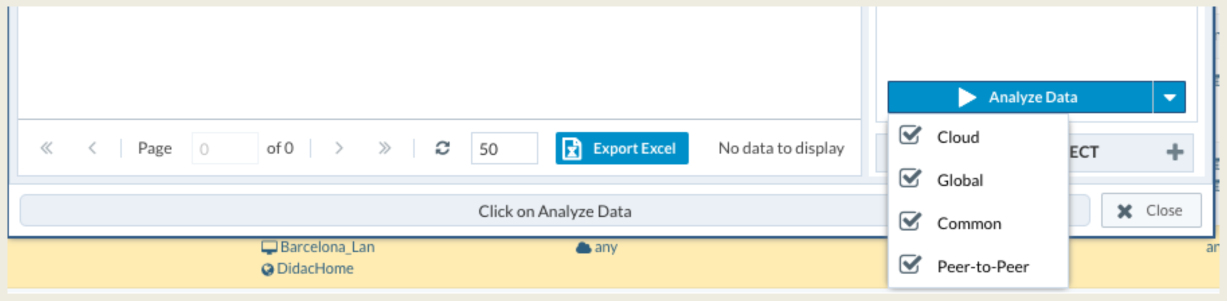
Thresholds

Discard traffic with less than ...

Hits:	1
Total Bytes:	1
Bytes Sent:	0
Bytes Received:	0

Click on Analyze Data

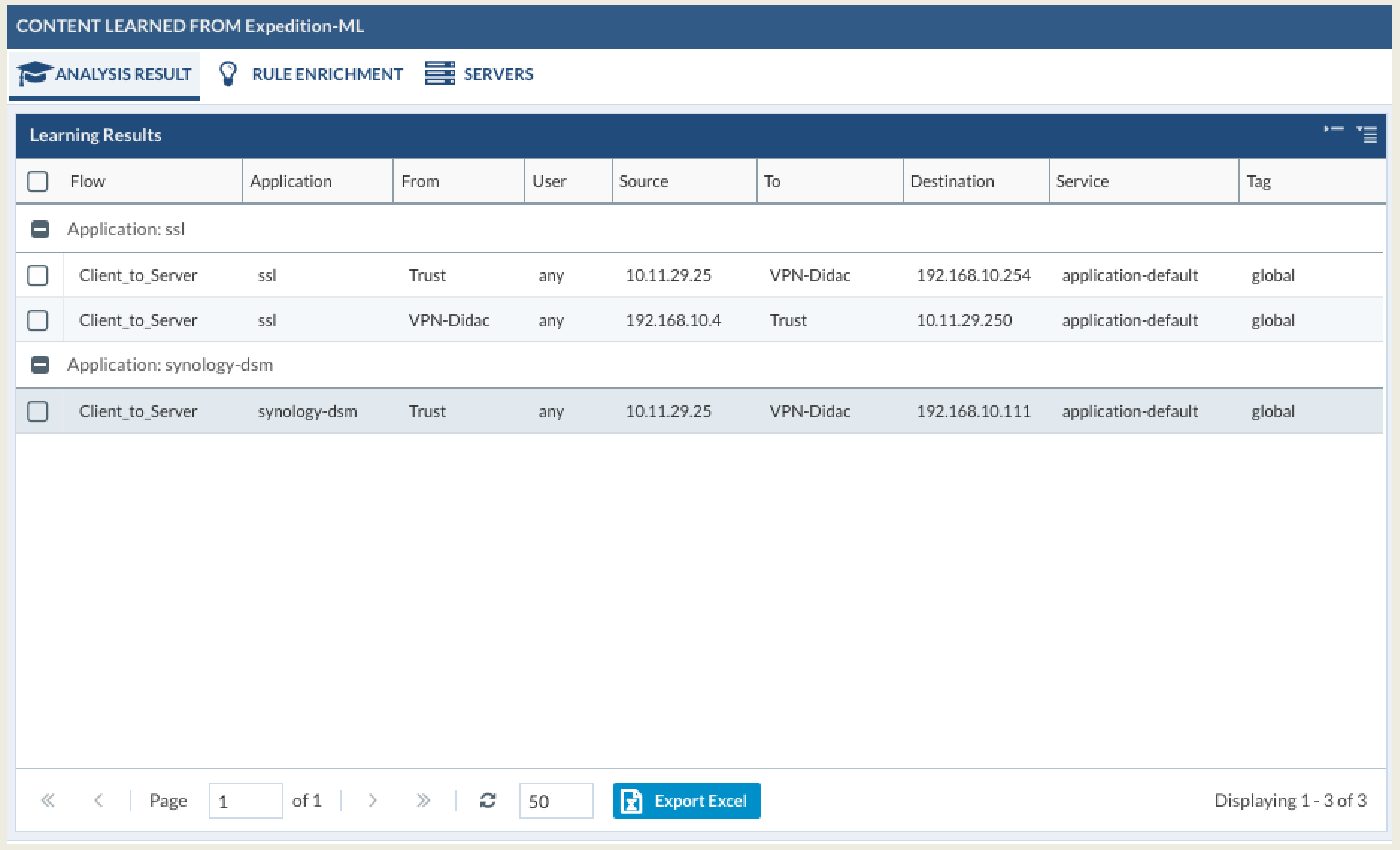
The **"Analyze Data"** button can be extended by clicking on the arrow.



- **Cloud:** Means Expedition will consider some applications as Cloud, when found in the traffic the destination ip addresses will be considered as "any" since they can dynamically change and doesn't make sense to keep the ones the network resolved in the moment we captured the traffic.
- **Common:** Are considered common applications that are present in all the networks and generate a huge volume of logs. Ex: ping, dns, ldap. In some cases you won't want to waste resources to analyze logs related to those applications to speed up the analysis for other applications. In case Expedition finds traffic regarding those applications will be considered as source "any '' and destination "any ''. Ex: ping Rule suggested ANY – ANY – ping – ALLOW
- **Peer-to-Peer:** All the applications classified as peer to peer by Palo Alto Networks.
- **Global:** All the other applications. Expedition will analyze sources, destinations, users, service ports to suggest Rules based on how they are consumed.

Once you select the Application categories, you can then click on **"Analyze Data"**

Once the analysis has finished:



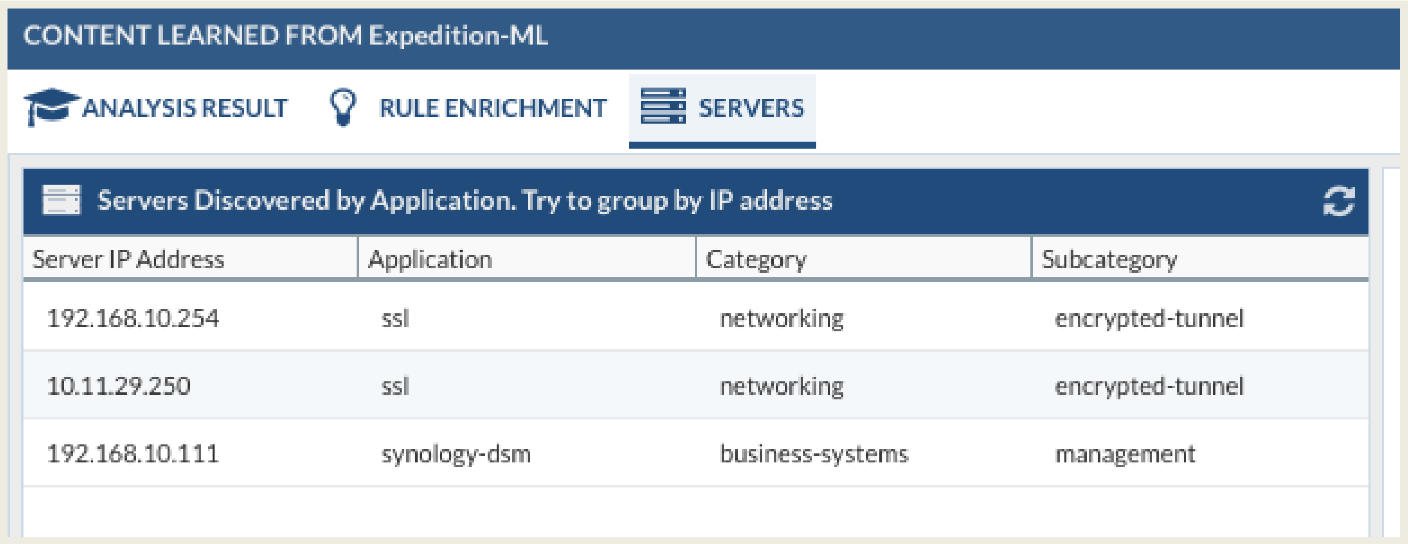
	Flow	Application	From	User	Source	To	Destination	Service	Tag
⊟	Application: ssl								
□	Client_to_Server	ssl	Trust	any	10.11.29.25	VPN-Didac	192.168.10.254	application-default	global
□	Client_to_Server	ssl	VPN-Didac	any	192.168.10.4	Trust	10.11.29.250	application-default	global
⊟	Application: synology-dsm								
□	Client_to_Server	synology-dsm	Trust	any	10.11.29.25	VPN-Didac	192.168.10.111	application-default	global

This is the result of the analysis for a single rule. This is all the Rules suggested based on how the applications have been consumed. In this case there are no users. The ML result will generate Rules based on Application , so for example if you have multiple sources and destinations with **"SSL"** as application , it will generate multiple security rules based on different sources and destinations, in contrast to RE feature, RE feature will group all different sources and destinations to one rule. The Flow has been calculated after figuring out who the servers are on the networks. Tag tells you the container for the APP. In this case all were **Global.**

Discovered Servers

This is the list of the servers we found from the logs after analyzing who is who in the network. This is important to Expedition to understand the flow of the communications.

This list of servers can be exported for an offline review by clicking on the Export Excel.

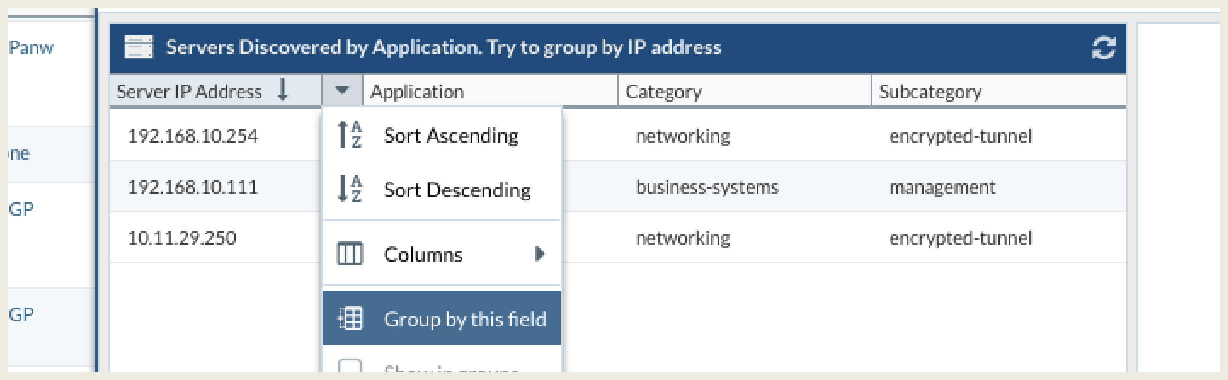


Servers Discovered by Application. Try to group by IP address

Server IP Address	Application	Category	Subcategory
192.168.10.254	ssl	networking	encrypted-tunnel
10.11.29.250	ssl	networking	encrypted-tunnel
192.168.10.111	synology-dsm	business-systems	management

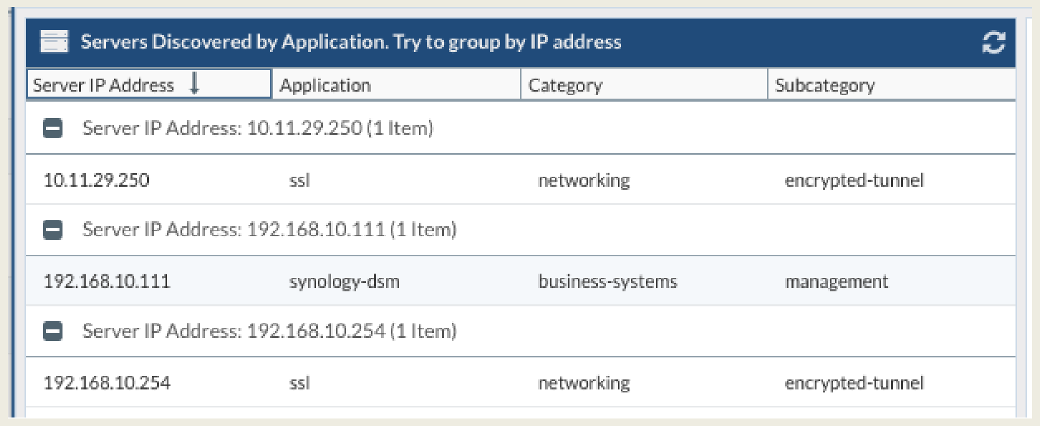
We can group the discovered Servers by Applications and see how many Applications each Server has.

- Point your mouse over the column "Server IP Address" and click on the arrow and click on the "Group by this field" In this example nothing will change but in real life environments will give you the applications served by each server we used.



Servers Discovered by Application. Try to group by IP address

Server IP Address	Application	Category	Subcategory
192.168.10.254		networking	encrypted-tunnel
192.168.10.111		business-systems	management
10.11.29.250		networking	encrypted-tunnel

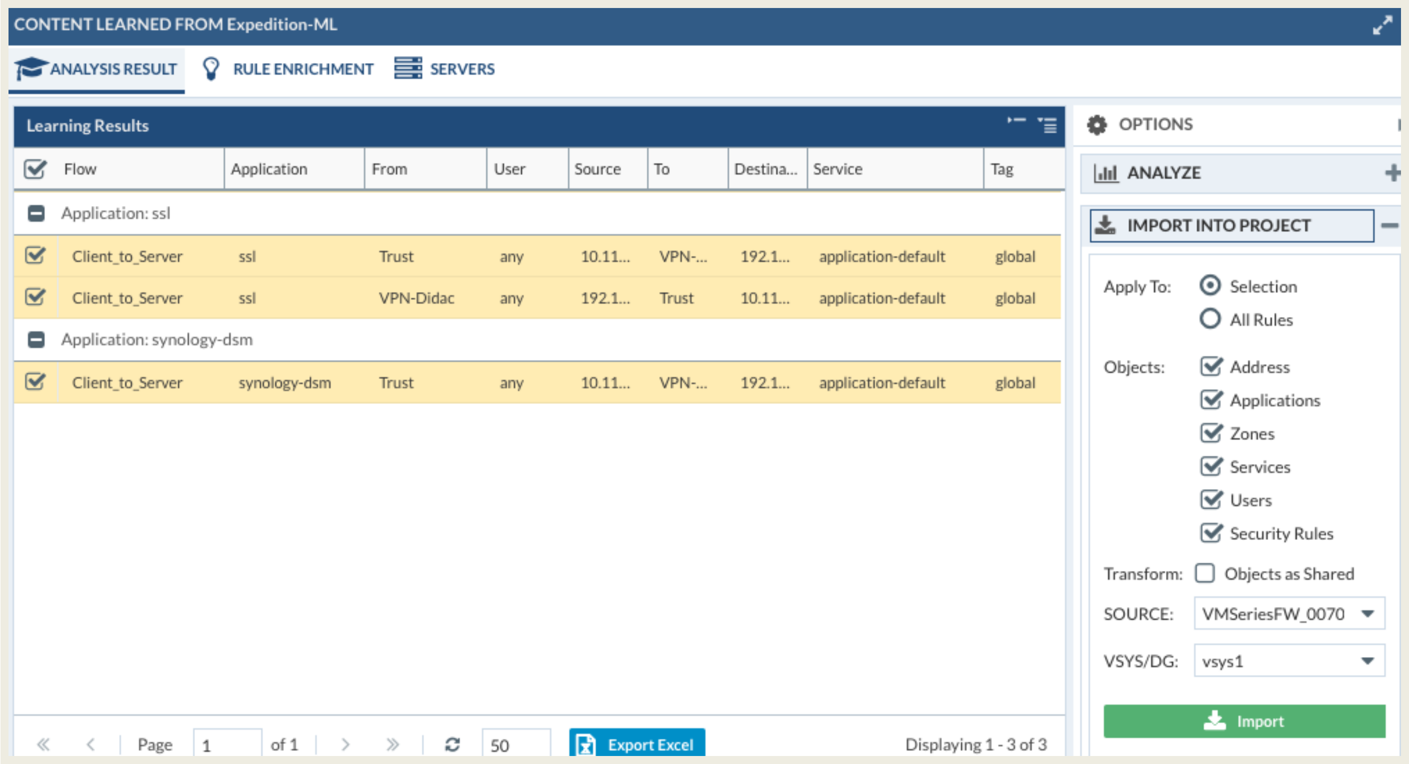


Servers Discovered by Application. Try to group by IP address

Server IP Address	Application	Category	Subcategory
Server IP Address: 10.11.29.250 (1 Item)			
10.11.29.250	ssl	networking	encrypted-tunnel
Server IP Address: 192.168.10.111 (1 Item)			
192.168.10.111	synology-dsm	business-systems	management
Server IP Address: 192.168.10.254 (1 Item)			
192.168.10.254	ssl	networking	encrypted-tunnel

Import suggested rules back to Project

Review the new suggested rules and select the ones you want to import to your security policy



☑ Flow	Application	From	User	Source	To	Destina...	Service	Tag
Application: ssl								
☑ Client_to_Server	ssl	Trust	any	10.11...	VPN-...	192.1...	application-default	global
☑ Client_to_Server	ssl	VPN-Didac	any	192.1...	Trust	10.11...	application-default	global
Application: synology-dsm								
☑ Client_to_Server	synology-dsm	Trust	any	10.11...	VPN-...	192.1...	application-default	global

Page 1 of 1 — 50 — Displaying 1 - 3 of 3

Apply To: ● Selection ○ All Rules

Objects: ☑ Address ☑ Applications ☑ Zones ☑ Services ☑ Users ☑ Security Rules

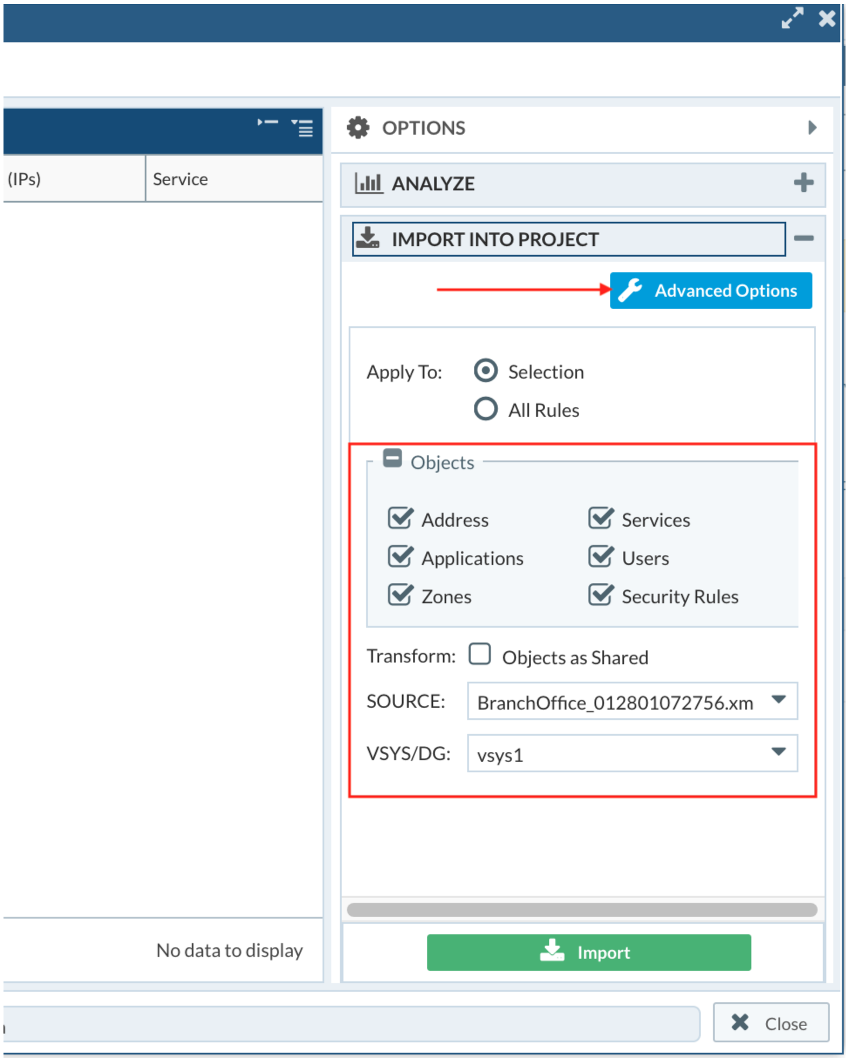
Transform: □ Objects as Shared

SOURCE: VMSeriesFW_0070

VSYS/DG: vsys1

In our example we will select them all and we will import everything, you can be more selective and convert all the new objects as part of your shared objects instead of creating them under the vsys / device-group selected at the time to import. Remember you can always export as Excel and review them offline.

In some cases , you would like to implement the security policy based on source or destination subnet instead of single IP, you can click on **"Advanced Options"**



Apply To: ● Selection ○ All Rules

Objects

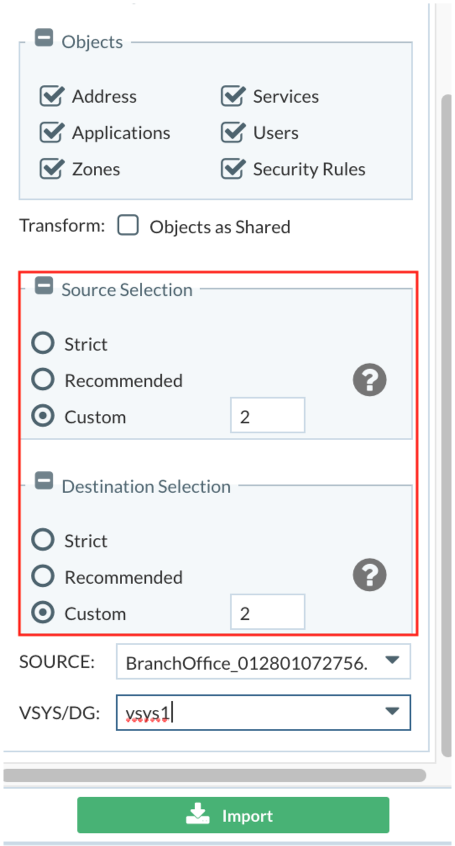
☑ Address ☑ Services
☑ Applications ☑ Users
☑ Zones ☑ Security Rules

Transform: □ Objects as Shared

SOURCE: BranchOffice_012801072756.xm

VSYS/DG: vsys1

It will then switch to the advanced option window like below, you could specify you want to import source or destination as /24 subnet when you see more than "x" IPs in the same /24 subnet. In the example, we use "2" , if the suggested rules contains more than 2 different IPs (10.0.0.1 and 10.0.0.2) in the same subnet , the import will change the source to "10.0.0.0/24" subnet instead of import them as 2 single IPs.



Objects

☑ Address ☑ Services

☑ Applications ☑ Users

☑ Zones ☑ Security Rules

Transform: □ Objects as Shared

Source Selection

○ Strict

○ Recommended

● Custom 2

Destination Selection

○ Strict

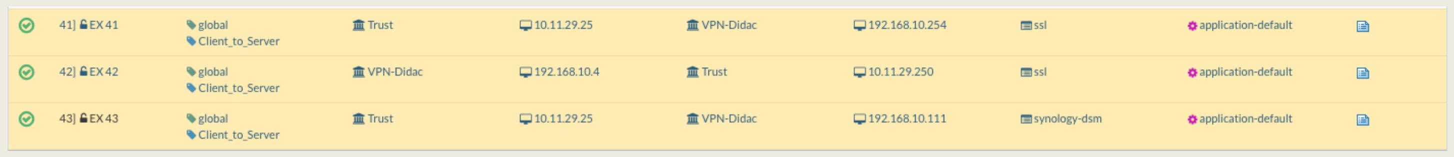
○ Recommended

● Custom 2

SOURCE: BranchOffice_012801072756.

VSYS/DG: vsys1

After clicking on **“Import”** , the rules imported will be named as **“EX-XX”** and will be placed at the very end of your current policy. We can move them after that where we prefer, usually before the rule that was under investigation



✓	41] EX 41	global Client_to_Server	Trust	10.11.29.25	VPN-Didac	192.168.10.254	ssl	application-default	
✓	42] EX 42	global Client_to_Server	VPN-Didac	192.168.10.4	Trust	10.11.29.250	ssl	application-default	
✓	43] EX 43	global Client_to_Server	Trust	10.11.29.25	VPN-Didac	192.168.10.111	synology-dsm	application-default	

Debug the Machine Learning

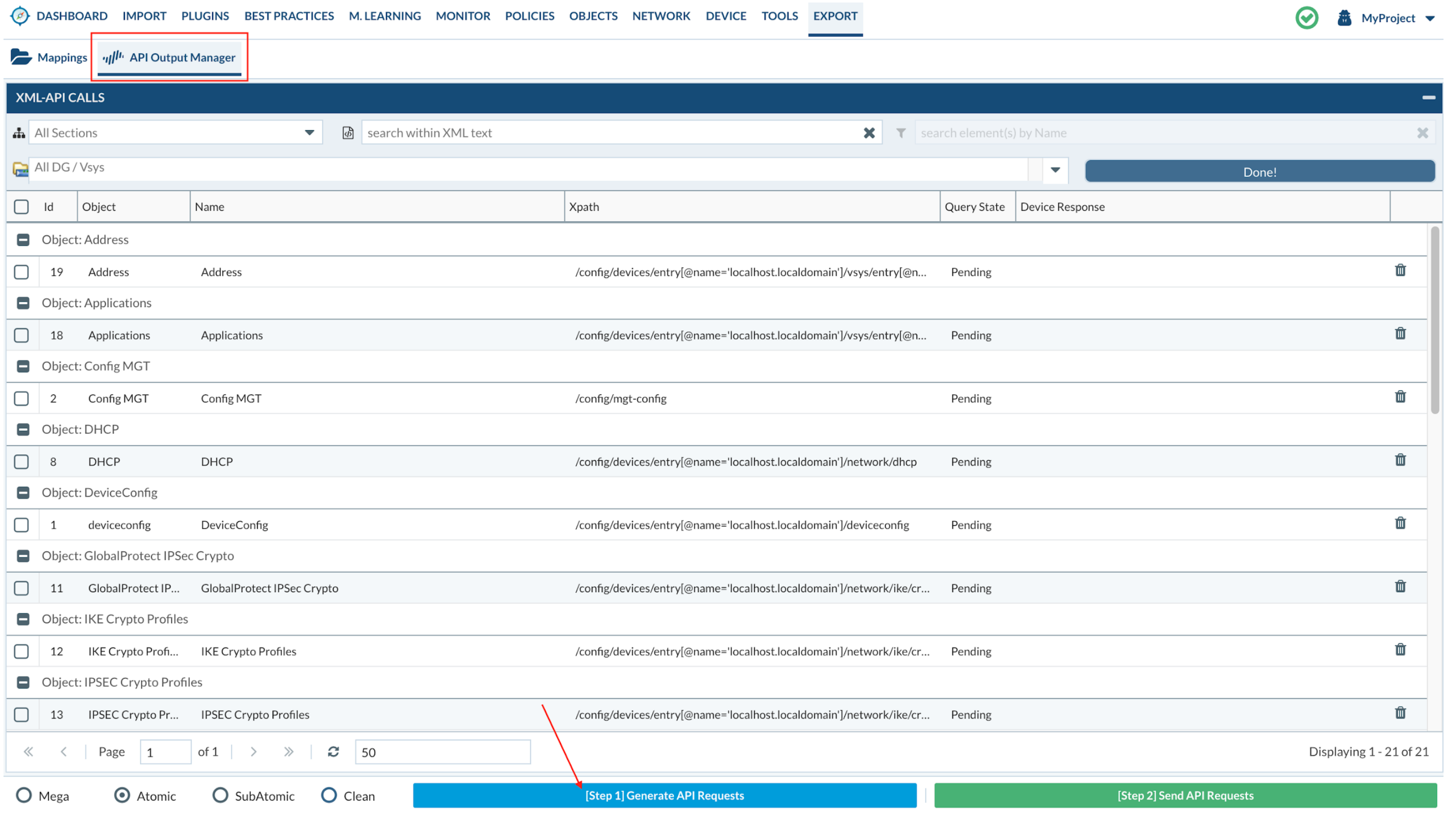
```
tail -f /tmp/error_SecRulesLearn
```

Push Modified Security Policies back to PAN-OS Devices

Generate API Requests

After you reviewed the modified security rules and ready to push the changes back to the PAN-OS device, you will go to **“Export”** ->**”API Output Manager”**, click on the **“[Step1] Generate API Requests”** to generate all the API calls. By default the mode will be **“Atomic”** **,** here are the details of each mode:

- **Mega**: generates one single API call containing the complete XML configuration.
- **Atomic:** generate one API call for each section in the configuration, such as address objects section, service object section, etc. In case of multiple DGs, we may obtain one API call for each section in the DG.
- **Subatomic**: generate an API call for every element in a configuration. For instance, one API call for each address object in a configuration. This may be required when only certain objects should be submitted to the device.
- **Clear**: generate API calls to delete the content of the desired section



☐	Id	Object	Name	Xpath	Query State	Device Response
Object: Address						
☐	19	Address	Address	/config/devices/entry[@name='localhost.localdomain']/vsys/entry[@n...	Pending	
Object: Applications						
☐	18	Applications	Applications	/config/devices/entry[@name='localhost.localdomain']/vsys/entry[@n...	Pending	
Object: Config MGT						
☐	2	Config MGT	Config MGT	/config/mgt-config	Pending	
Object: DHCP						
☐	8	DHCP	DHCP	/config/devices/entry[@name='localhost.localdomain']/network/dhcp	Pending	
Object: DeviceConfig						
☐	1	deviceconfig	DeviceConfig	/config/devices/entry[@name='localhost.localdomain']/deviceconfig	Pending	
Object: GlobalProtect IPSec Crypto						
☐	11	GlobalProtect IP...	GlobalProtect IPSec Crypto	/config/devices/entry[@name='localhost.localdomain']/network/ike/cr...	Pending	
Object: IKE Crypto Profiles						
☐	12	IKE Crypto Profi...	IKE Crypto Profiles	/config/devices/entry[@name='localhost.localdomain']/network/ike/cr...	Pending	
Object: IPSEC Crypto Profiles						
☐	13	IPSEC Crypto Pr...	IPSEC Crypto Profiles	/config/devices/entry[@name='localhost.localdomain']/network/ike/cr...	Pending	

Page 1 of 1 | 50 | Displaying 1 - 21 of 21

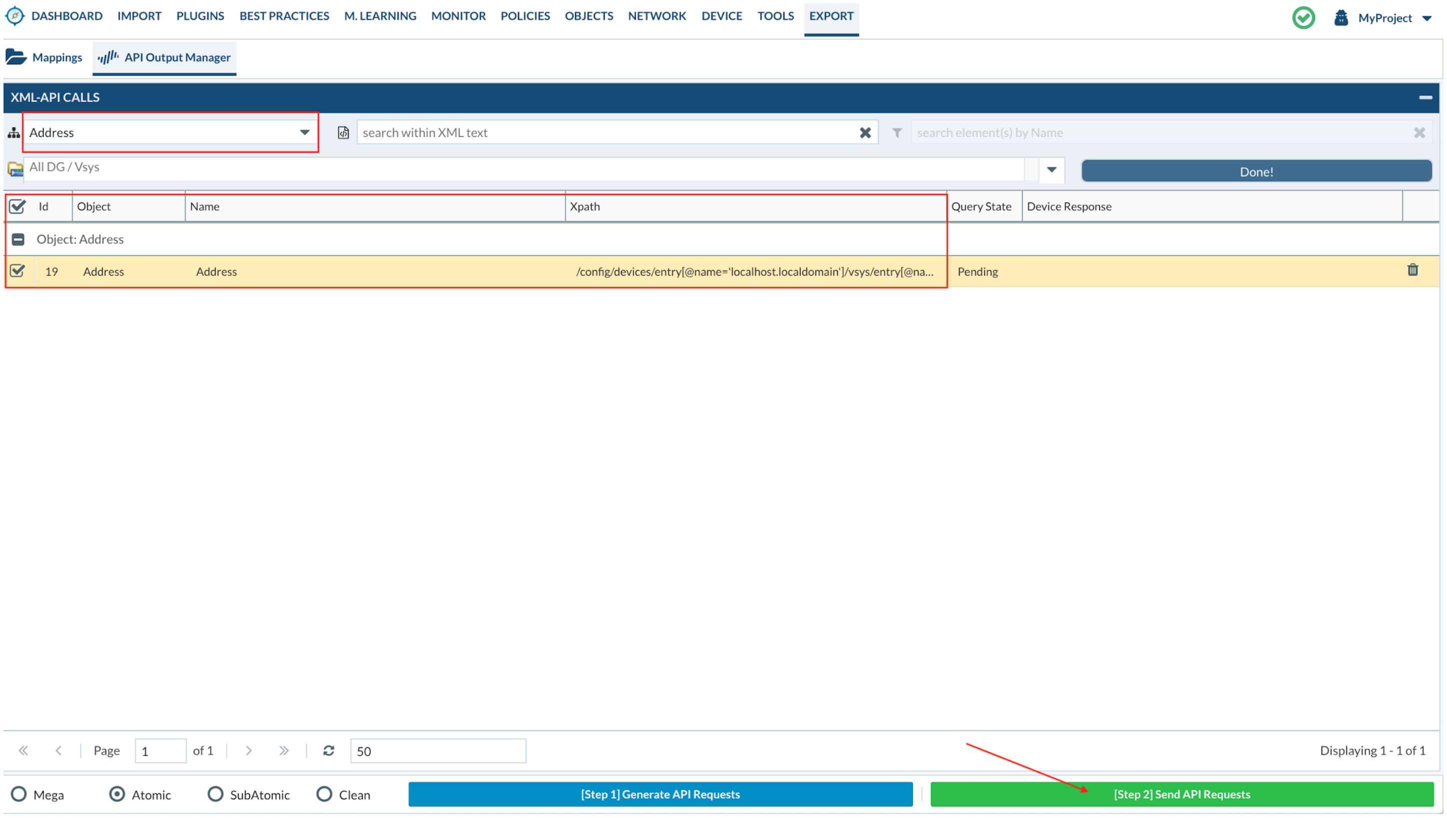
○ Mega ● Atomic ○ SubAtomic ○ Clean

Send API Requests

You can then select the API calls you would like to send to the PAN-OS device by selecting the checkbox in front of the API calls , the order of the API calls are listed in the **"Id"** column , you will start with the API calls with lower Id, for example, if the changes in Expedition include below new objects and policies, you will push those API calls one by one per below order:

Tag
Address objects
Address-Group objects
Service objects
Service-Group objects
Security Rules

You can filter the section of the configuration by selecting the section on the filter dropdown menu. check on the API calls you would like to send to the PAN-OS device, click on **"[Step 2] Send API Requests"**



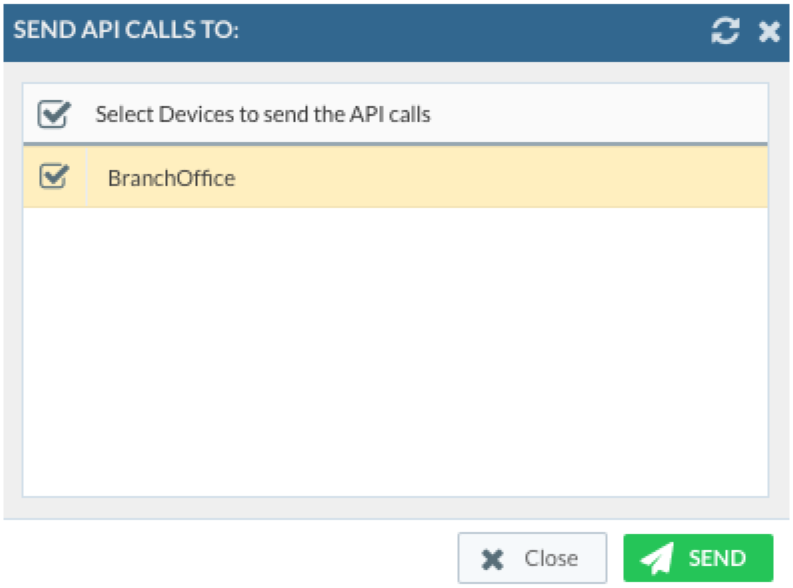
Filter: Address
All DG / Vsys

☑	Id	Object	Name	Xpath	Query State	Device Response
	Object: Address					
☑	19	Address	Address	/config/devices/entry[@name='localhost.localdomain']/vsys/entry[@na...	Pending	

Page 1 of 1 | 50 | Displaying 1 - 1 of 1

○ Mega ● Atomic ○ SubAtomic ○ Clean

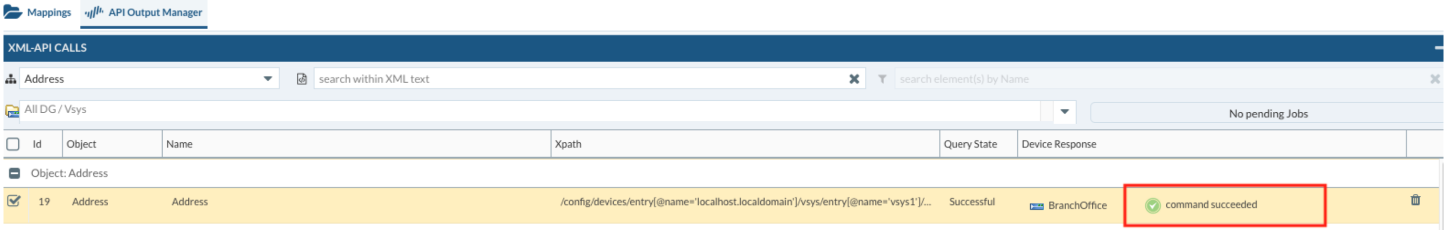
Select the Device you want to send the API call to , then click on"**SEND**"


SEND API CALLS TO:

☑ Select Devices to send the API calls

☑ BranchOffice

After the command , the Screen will refresh with the status of the API calls, in the case of successful API call, th status will show “**command succeeded** “ like below screenshot:



Address

All DG / Vsys

No pending Jobs

☐	Id	Object	Name	Xpath	Query State	Device Response		
⊟	Object: Address							
☑	19	Address	Address	/config/devices/entry[@name='localhost.localdomain']/vsys/entry[@name='vsys1']/...	Successful	BranchOffice	✓ command succeeded	

After you send all the required API calls, you can then verify changes in the PAN-OS device and **Commit** the changes.

Revision History

Date	Revision	Comment
Nov,5, 2020	A	First release of the documentation
Apr, 11, 2022	B	Adding Splunk app info