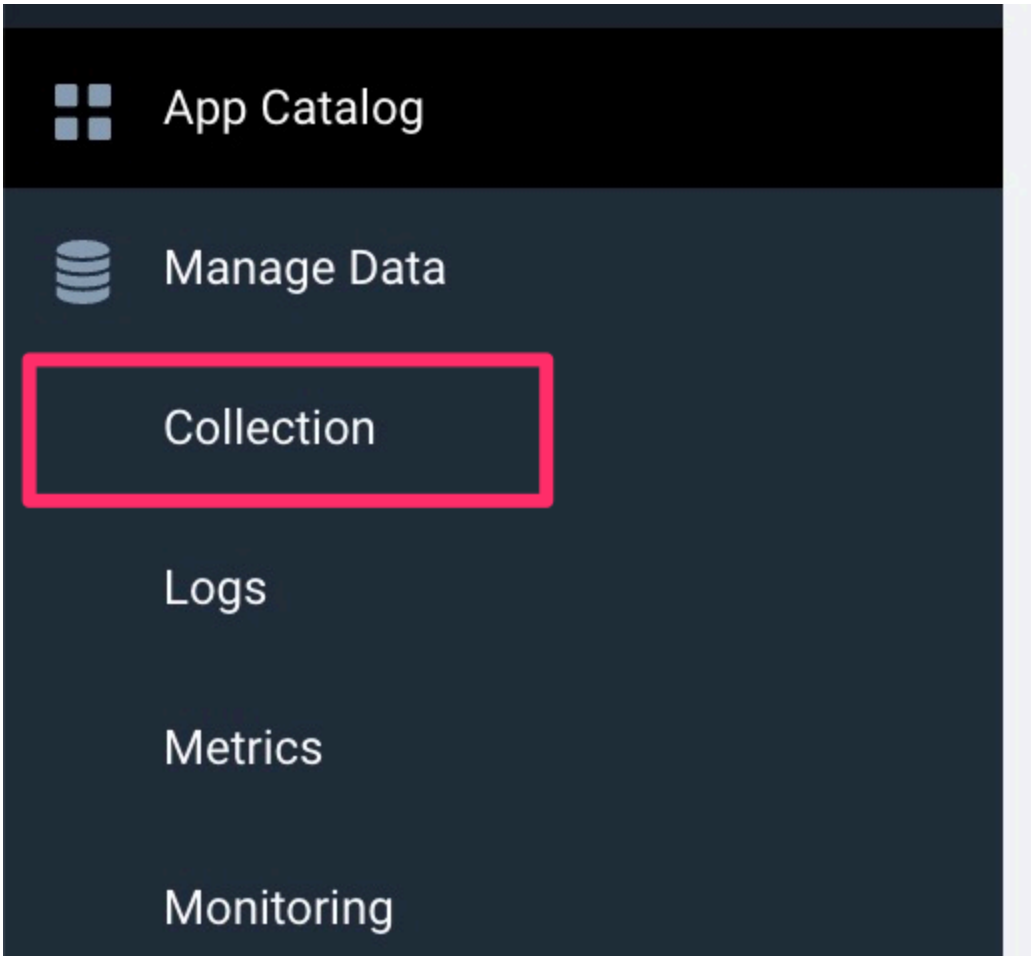
Importing only outbound VPC flow logs into Sumo Logic

dev.classmethod.jp/articles/sumo-logic_vpc-flow_processing-rules

酒井剛 June 14, 2022



When you input logs into a SIEM product, you may be concerned about the cost.
You want to keep the logs as they are as long as possible, but when you think about the log storage fee, you want to only capture the logs you really need.

Sumo Logic allows you to use Processing Rules to select logs to be ingested using regular expressions. In this article, I would like to introduce a method **to reduce Sumo Logic costs** by filtering **the ingestion of VPC flow logs , which tend to be particularly large in volume among AWS logs but are extremely effective as a target for analysis.**

What is a Processing Rule?

This function is used to control the import of messages or process messages when a specific pattern contained in a log message is matched when data is imported into Sumo Logic.

Regarding **import control , there are two types of filters:**

- Exclude messages that match : Works as a blacklist and does not include messages that match the specified pattern (includes all messages that do not match)
- Include messages that match : This acts as a whitelist and includes only messages that match the specified pattern (all messages that do not match will not be included).

Regarding message processing , there are two types of functions called actions:

- Hash messages that match : This acts as a hashing mechanism for secret information, converting message parts that match a specified pattern into a random hash value, and then capturing the message.
- Mask messages that match : This masks confidential information and converts the parts of the message that match the specified pattern into customizable characters before capturing the message.

Understand the format of the logs you want to ingest

To utilize Processing Rules, you need to define the message pattern that will be the condition for specific processing. To do this, you need to understand the target log format. In this case, we are using VPC flow logs, so you can check the format of the output log in the official AWS documentation.

When configuring VPC flow logs, you can output logs in the default format or specify fields to output in a custom format. This time, we will assume that logs are output in the default format.
According to the documentation,

> The following table lists all available fields for a flow log record. The Version column shows the VPC Flow Logs version in which the field was introduced. The default format includes all Version 2 fields, in the same order as the table.

So, if you check the table in the document, you will see that the log format is as follows:

```
version account-id interface-id srcaddr dstaddr srcport dstport protocol packets
bytes start end action log-status
```

Now that we understand the VPC log format, let's think about the conditional pattern to identify the outbound communication logs, which is what we want to do this time.

About the communication direction of VPC flow logs

From the message content of the VPC flow log, it is necessary to determine whether the log message is outbound or inbound communication (or internal communication, i.e. communication between AWS). In the log output settings above, it is possible to select a custom format and determine this from the value of "flow-direction" in version 5 , but this time we will only consider the fields in version 2 of the default format.
Therefore , we will focus on "srcaddr" and "dstaddr" .

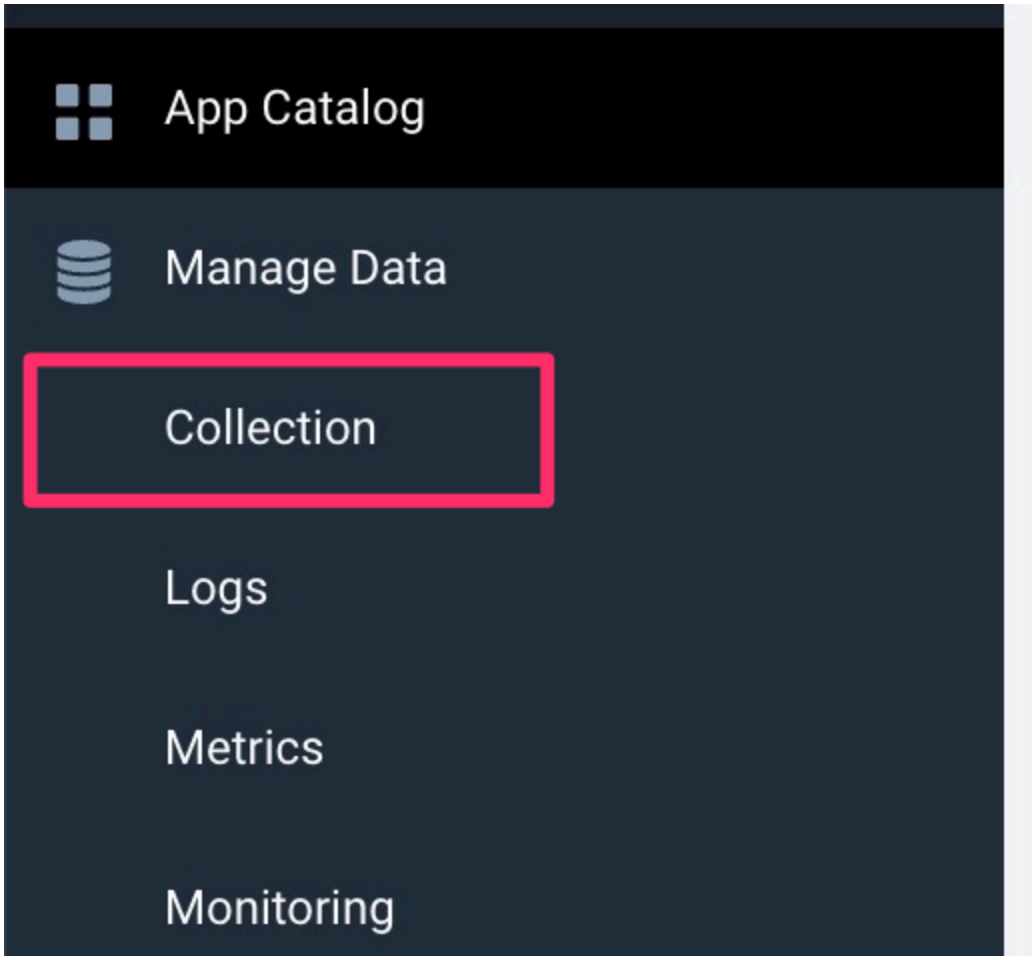
In these two cases, if **"srcaddr" is a global IP** and **"dstaddr" is a private IP** , the message is an **inbound**
communication. Conversely, if **"srcaddr" is a private IP** and **"dstaddr"** is a global IP, the message is an **outbound**
communication. Also, if **both "srcaddr" and "dstaddr" are private IPs** , it is possible to determine that the communication is **internal communication**
. Therefore, if **"dstaddr" is a private IP , it can be determined that the communication is inbound or internal** . If you create a rule that rejects the import of messages where "dstaddr" is a private IP and imports all other messages, it may be possible to import **only outbound communication .**

The previous applies here.

Give it a try

Now, let's configure log import on the Sumo Logic side.

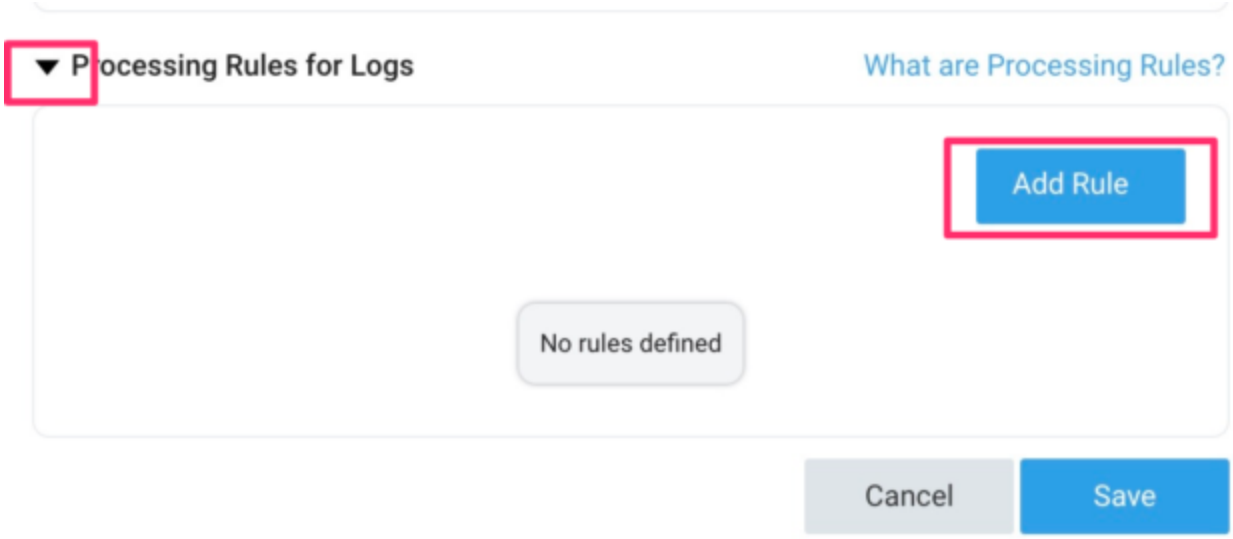
Log in to Sumo Logic and go to Manage Data > Collection in the left pane.



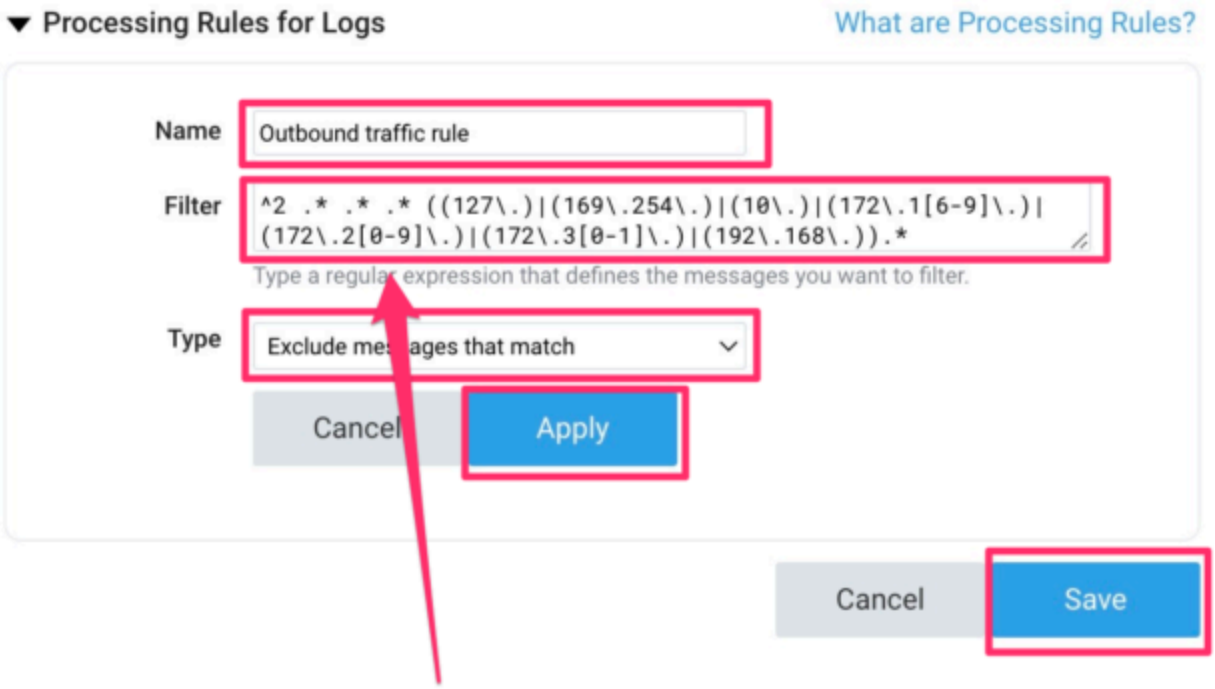
Click Edit for the VPC Flow Log Source of the Hosted Collector.

*If you have not created a Hosted Collector, please create one. If you have not yet configured data import settings (created a source) for VPC flow logs, please create a source. Our blog also explains how to import AWS logs stored in an S3 bucket. Please refer to [this page if you want to create a new one.]

Next, scroll down to the bottom of the screen and you will see a collapsed section called Processing Rules for Logs. Open it and select Add Rule.



Enter an arbitrary rule name and the condition pattern you thought up earlier in Filter. The regular expressions that can be used in the condition pattern are RE2 syntax. Check [here for the syntax that can be used.]




```
^2 .* .* .* ((127\.)|(169\.254\.)|(10\.)|(172\.1[6-9]\.)|(172\.2[0-9]\.)|(172\.3[0-1]\.)|(192\.168\.)).*
```

Each field is separated by a space, and "dstaddr" is the fifth field. We want to reject messages where the fifth field is a private IP address. Therefore, select **"Exclude messages that match" for Type, click Apply, and then click Save.**

Let's check it out

Now that we've configured the system to capture only outbound communication, let's verify that the logs are being captured as expected.
Try pinging an EC2 instance in the VPC from the PC console. Since this is inbound communication from outside, we assume that the VPC flow logs for this communication will not be captured by Sumo Logic. *Although it shows a timeout, this is not a problem as it is recorded as DENY in the VPC flow logs.

```
$ ping <VPC内のインスタンスのPublic IP>
PING <VPC内のインスタンスのPublic IP> (<VPC内のインスタンスのPublic IP>): 56 data bytes
Request timeout for icmp_seq 0
Request timeout for icmp_seq 1
Request timeout for icmp_seq 2
^C
```

Next, connect to an EC2 instance in the VPC and ping the private IP address from this instance. Since this is internal communication to another VPC within AWS, the VPC flow logs for this communication will not be ingested into Sumo Logic.

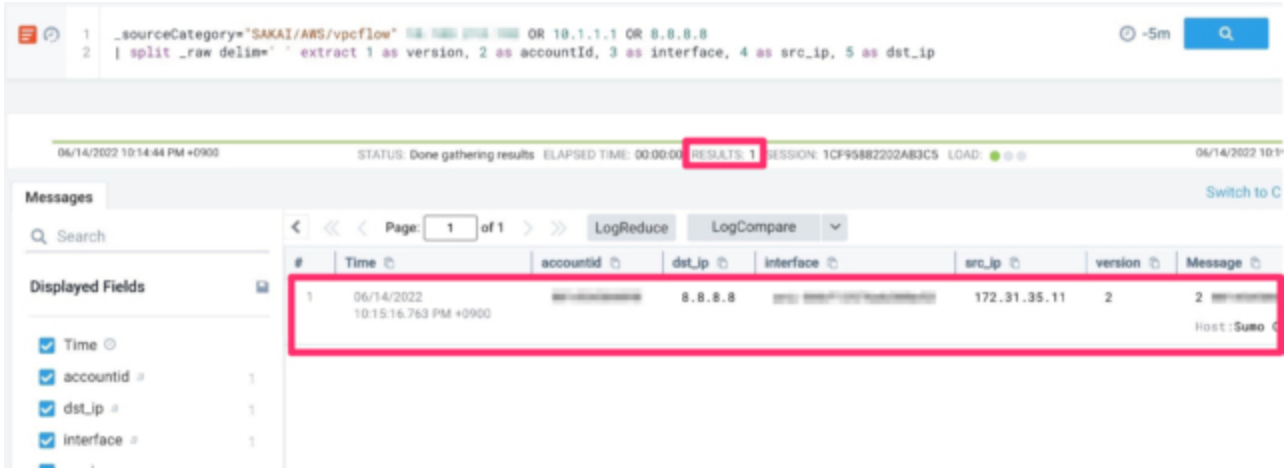
```
$ ping 10.1.1.1
PING 10.1.1.1 (10.1.1.1) 56(84) bytes of data.
^C
--- 10.1.1.1 ping statistics ---
17 packets transmitted, 0 received, 100% packet loss, time 16372ms
```

Finally, try pinging a public IP address from an EC2 instance in the VPC. Since this is outbound communication, only the VPC flow logs for this communication will be imported into Sumo Logic.

```
$ ping 8.8.8.8
PING 8.8.8.8 (8.8.8.8) 56(84) bytes of data.
64 bytes from 8.8.8.8: icmp_seq=1 ttl=105 time=2.68 ms
64 bytes from 8.8.8.8: icmp_seq=2 ttl=105 time=2.64 ms
^C
--- 8.8.8.8 ping statistics ---
2 packets transmitted, 2 received, 0% packet loss, time 1001ms
rtt min/avg/max/mdev = 2.641/2.661/2.681/0.020 ms
```

Let's search the logs in the Sumo Logic console. By searching for the IP addresses that were pinged, we were able to confirm that only outbound communications were being

captured, as expected.



summary

This time, we introduced how to use Sumo Logic's Processing Rules to import only logs that meet specific conditions. Importing logs into Sumo Logic enables continuous log visualization and enhanced security through flexible search drill-down, but costs can be a concern. Understanding the
logs necessary to increase security insights and then filtering out unnecessary logs to reduce costs are key challenges in SIEM operation.

We hope you will use this feature to help you further your optimization efforts.