Amazon Web Services: VPC Flow Logs

Amazon Web Services (AWS) has become an integral part of many organizations’ IT infrastructure. An Amazon Virtual Private Cloud (VPC) enables you to launch AWS resources into a virtual network that you've defined, with the benefits of using the scalable infrastructure of AWS. VPC flow log data is collected outside of the path of your network traffic, and therefore does not affect network throughput or latency. You can create or delete flow logs without any risk of impact to network performance.

VPC Flow logs contain a comprehensive record of network traffic in and out of your AWS environment. By default, the record includes values for the different components of the IP flow, including the source, destination, and protocol. They are often used for troubleshooting connectivity issues across your VPCs, intrusion detection, or anomaly detection. In the Common Information Model, VPC flow log data is typically mapped to the Network Traffic Data model.

Configuration

Guidance for onboarding data can be found in the Spunk Documentation, Getting Data In (Splunk Enterprise) or Getting Data In (Splunk Cloud). Refer to the documentation, and note the following:

- Recommended index: awsflow
- Source type: aws:cloudwatchlogs:vpcflow
- Input type: Cloudwatch logs and Kinesis. It is best to collect VPC flow logs and CloudWatch logs through Kinesis streams. However, the AWS Kinesis input has the following limitations:
  - Multiple inputs collecting data from a single stream cause duplicate events in the Splunk platform.
  - Does not support monitoring of dynamic shards repartition, which means when there is a shard split or merge, the add-on cannot automatically discover and collect data in the new shards until it is restarted. After you repartition shards, you must restart your data collection node to collect data from the partitions.
- Add-on or app: Splunk Add-on for Amazon Web Services
- Sizing estimate: The recommended maximum daily indexing volume for a typical VPC Flow log source type on a clustered indexer is 25 - 30 GB per indexer. Use this as a rough guideline to plan for the number of indexers to deploy in your clustered environment. Adding more indexers to a cluster improves indexing and search retrieval performance. Since this also incurs some additional within-cluster data replication traffic, adjust the number of indexers in your cluster based on your actual system performance. AWS limits each account to 10 requests per second, each of which returns no more than 1 MB of data. This means the data ingestion and indexing rate is no more than 10MB/s. The add-on modular input can process up to 4K events per second in a single log stream. If volume is a concern, configure the only_after parameter to limit the amount of historical data you collect. If you have high volume VPC Flow Logs, configure one or more Kinesis inputs to collect them instead of using the CloudWatch Logs input.
Validation

You can make sure that Splunk has begun ingesting the data from AWS by running Splunk searches. The Splunk Add-on for AWS also has a built-in health-overview dashboard that will provide initial troubleshooting information.

Data application

When your Splunk deployment is ingesting VPC Flow Logs, you can use the data to achieve the following:

- Managing an Amazon Web Services environment
- Detecting UDP service amplification abuse
- Monitoring AWS for suspicious traffic
- Detecting AWS network ACL activity