ML Based Syslog Analytics

Business Challenge

Our client generates a large number of Syslog messages from different types of routers. Syslog is a standard for message logging. Syslogs are in semi-structured format and it was difficult to detect abnormal network behaviors. The current process involved manually going through millions of Syslog messages to identify faults in the router network. Acquiring useful insights out of it to analyze network behavior was becoming a strenuous task for the client.

Our challenge was to timely detect the abnormal network behaviors in real time thereby reducing network downtime.

Approach and Solution

- Our approach was to ingest live streaming Syslog into Azure Data Explorer(ADX) from the on-prem sources

Azure Data Explorer(ADX) is a fast, fully managed data analytics service for real-time analysis of large volumes of data streaming from applications, websites, IoT devices, and more.

- Event grouping by correlation analysis and domain inputs helped us eliminate noises in the Syslog to a great extent.
- The ingested Syslogs are parsed using regular expressions. The parsing function gets invoked as new logs are generated. The parsed Syslog is processed using Azure Data Explorer(ADX) to convert to a time series data with counts for each type of message.
- The ADX in-built anomaly detection model was used to detect anomalies in the time series data. The model uses seasonal decomposition method to achieve this. The seasonality and trends in the data was obtained by looking at past 24hours counts.
- Power BI was used for reporting dashboards for each vendor separately which shows the anomaly details for the last 24 hours.

Benefits

- Efficiently detected anomalies in Syslog in real time to take quick action
- Reduces business impact caused by the issue

Technology Used :

- Kafka, Machine learning, Timeseries, Azure ADX, Power BI