CISCO Data Understanding Project

Date of Submission: 15th November

Prabhakar Narasimha Nanduri

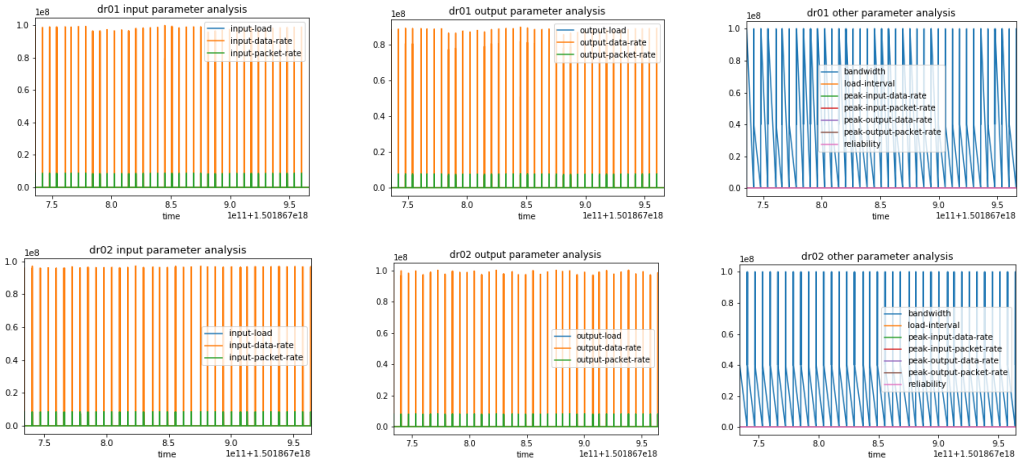
Q1) Initial Analysis

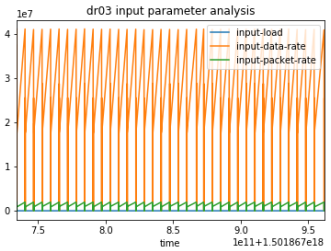
Understanding the information given:

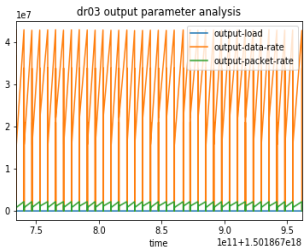
From the set of instructions and file-sets given the following aspects about the data can be deduced:-

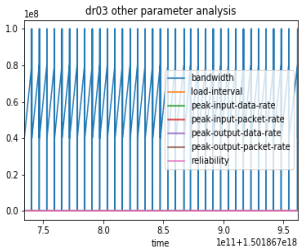
- There are 14 numeric columns of data for every producer (element of the network) of which one column is the time (timestamp) which can be taken as the reference; thus for plotting purposes can be kept as the x-axis. [The column of interface-name is being ignored since it is textual and would not help in visual analysis, network understanding on a superficial level can be obtained for it]
- It is already known that there is no 'bgp clear command' induced on the network elements dr01, dr02 and dr03. Thus for reference purposes we can use the graphs plotted for these network elements (13 parameters vs time) to understand what columns need to be considered for further analysis and what columns contain redundant data.
- Of the 13 columns there are three columns relating to the input traffic parameters sent (input-load, input-packet-rate, input-data-rate), three columns relating to the output parameters being sent(output-load, output-packet-rate, output-data-rate) and 7 other columns which are to be understood on whether the data contained in them is redundant.
- To begin the initial analysis of understanding which columns contain relevant data pertaining to the questions asked, a simple python script (attached in appendix) has been developed to plot three figures for each producer. File1 data (as a csv) format has been parsed to generate the three figures. The first figure plots the input parameters vs time, the second figure plots output parameters vs time and the third parameter plots the other parameters vs time.

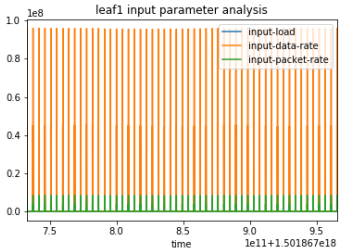
Note: For all the graphs shown below the reference (x-axis has been taken as time (in ns))

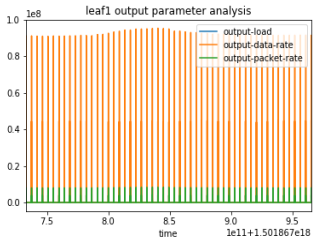


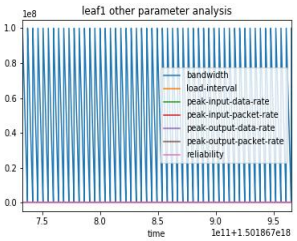

dr03 input parameter analysis
1e7
input-load
input-data-rate
input-packet-rate
time
1e11+1.501867e18

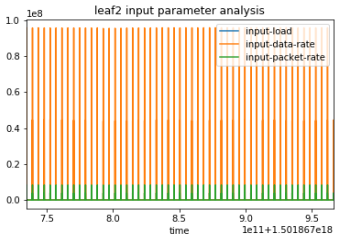

dr03 output parameter analysis
1e7
output-load
output-data-rate
output-packet-rate
time
1e11+1.501867e18

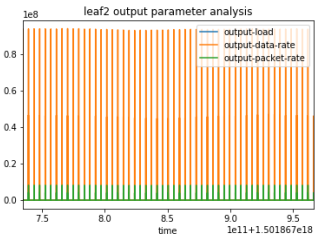

dr03 other parameter analysis
1e8
bandwidth
load-interval
peak-input-data-rate
peak-input-packet-rate
peak-output-data-rate
peak-output-packet-rate
reliability
time
1e11+1.501867e18

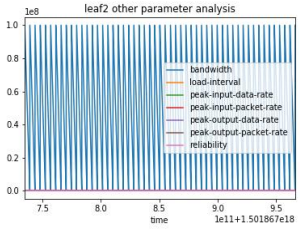

leaf1 input parameter analysis
1e8
input-load
input-data-rate
input-packet-rate
time
1e11+1.501867e18

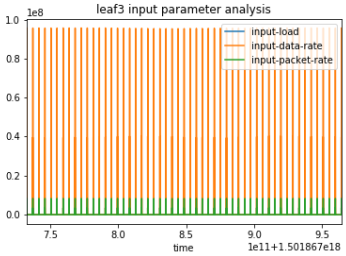

leaf1 output parameter analysis
1e8
output-load
output-data-rate
output-packet-rate
time
1e11+1.501867e18

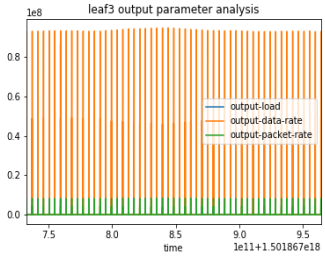

leaf1 other parameter analysis
1e8
bandwidth
load-interval
peak-input-data-rate
peak-input-packet-rate
peak-output-data-rate
peak-output-packet-rate
reliability
time
1e11+1.501867e18

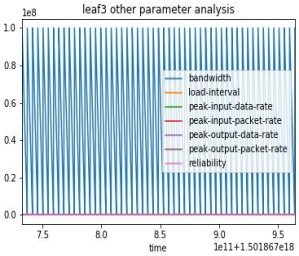

leaf2 input parameter analysis
1e8
input-load
input-data-rate
input-packet-rate
time
1e11+1.501867e18

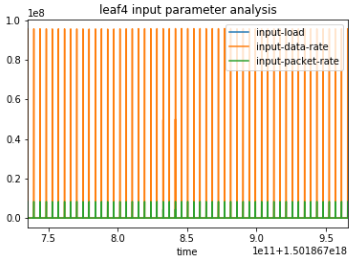

leaf2 output parameter analysis
1e8
output-load
output-data-rate
output-packet-rate
time
1e11+1.501867e18

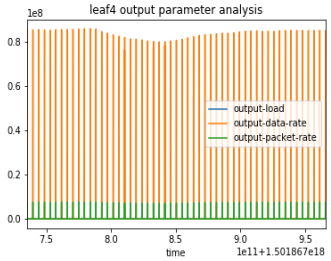

leaf2 other parameter analysis
1e8
bandwidth
load-interval
peak-input-data-rate
peak-input-packet-rate
peak-output-data-rate
peak-output-packet-rate
reliability
time
1e11+1.501867e18

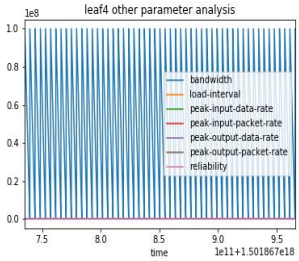

leaf3 input parameter analysis
1e8
input-load
input-data-rate
input-packet-rate
time
1e11+1.501867e18

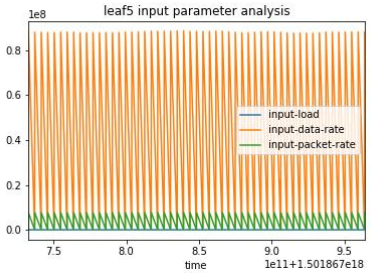

leaf3 output parameter analysis
1e8
output-load
output-data-rate
output-packet-rate
time
1e11+1.501867e18

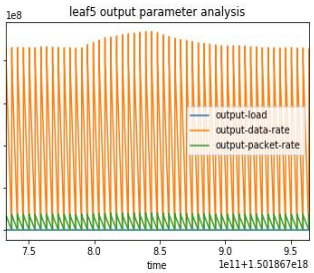

leaf3 other parameter analysis
1e8
bandwidth
load-interval
peak-input-data-rate
peak-input-packet-rate
peak-output-data-rate
peak-output-packet-rate
reliability
time
1e11+1.501867e18

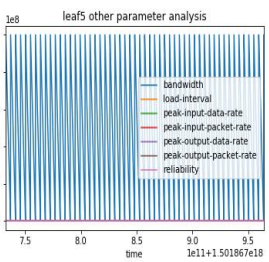

leaf4 input parameter analysis
1e8
input-load
input-data-rate
input-packet-rate
time
1e11+1.501867e18

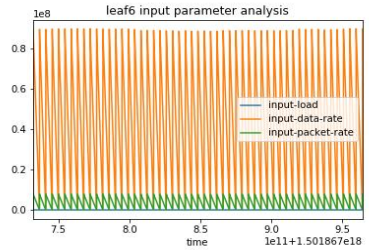

leaf4 output parameter analysis
1e8
output-load
output-data-rate
output-packet-rate
time
1e11+1.501867e18

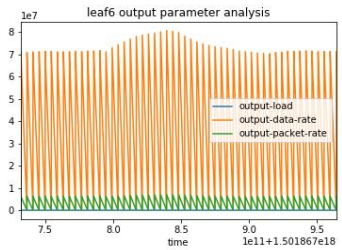

leaf4 other parameter analysis
1e8
bandwidth
load-interval
peak-input-data-rate
peak-input-packet-rate
peak-output-data-rate
peak-output-packet-rate
reliability
time
1e11+1.501867e18

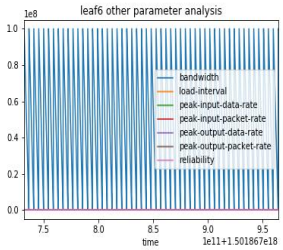

leaf5 input parameter analysis
1e8
input-load
input-data-rate
input-packet-rate
time
1e11+1.501867e18

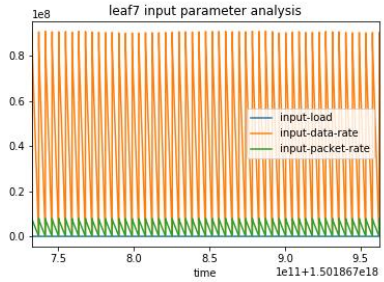

leaf5 output parameter analysis
1e8
output-load
output-data-rate
output-packet-rate
time
1e11+1.501867e18

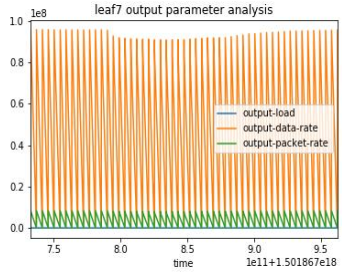

leaf5 other parameter analysis
1e8
bandwidth
load-interval
peak-input-data-rate
peak-input-packet-rate
peak-output-data-rate
peak-output-packet-rate
reliability
time
1e11+1.501867e18

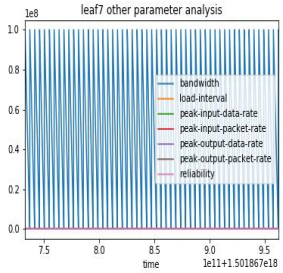

leaf6 input parameter analysis
1e8
input-load
input-data-rate
input-packet-rate
time
1e11+1.501867e18

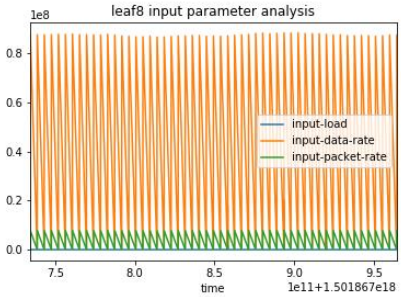

leaf6 output parameter analysis
1e7
output-load
output-data-rate
output-packet-rate
time
1e11+1.501867e18

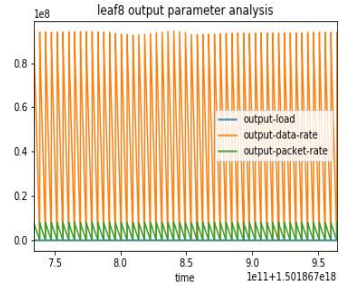

leaf6 other parameter analysis
1e8
bandwidth
load-interval
peak-input-data-rate
peak-input-packet-rate
peak-output-data-rate
peak-output-packet-rate
reliability
time
1e11+1.501867e18

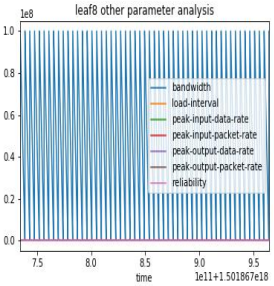

leaf7 input parameter analysis
1e8
input-load
input-data-rate
input-packet-rate
time
1e11+1.501867e18

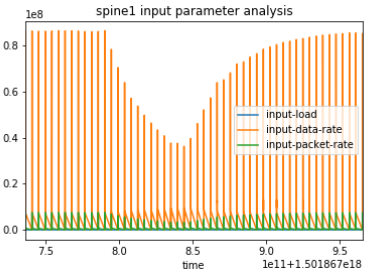

leaf7 output parameter analysis
1e8
output-load
output-data-rate
output-packet-rate
time
1e11+1.501867e18

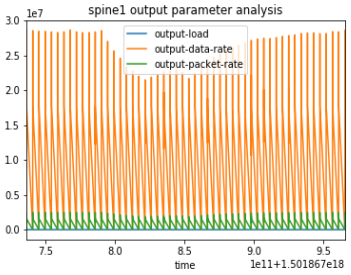

leaf7 other parameter analysis
1e8
bandwidth
load-interval
peak-input-data-rate
peak-input-packet-rate
peak-output-data-rate
peak-output-packet-rate
reliability
time
1e11+1.501867e18

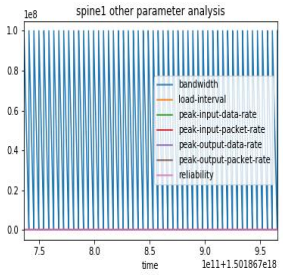

leaf8 input parameter analysis
1e8
input-load
input-data-rate
input-packet-rate
time
1e11+1.501867e18

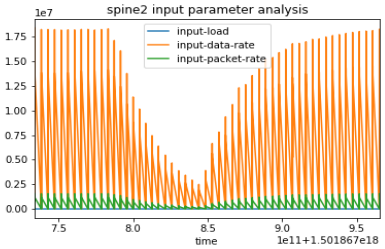

leaf8 output parameter analysis
1e8
output-load
output-data-rate
output-packet-rate
time
1e11+1.501867e18

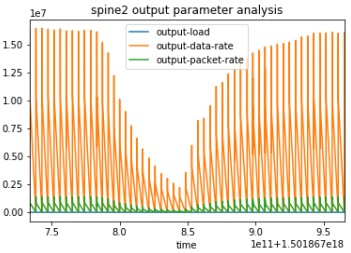

leaf8 other parameter analysis
1e8
bandwidth
load-interval
peak-input-data-rate
peak-input-packet-rate
peak-output-data-rate
peak-output-packet-rate
reliability
time
1e11+1.501867e18

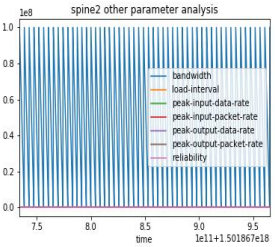

spine1 input parameter analysis
1e8
input-load
input-data-rate
input-packet-rate
time
1e11+1.501867e18

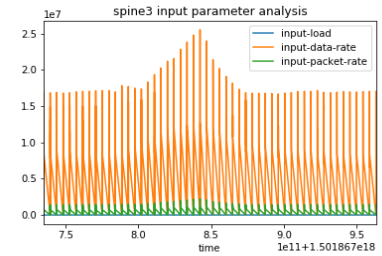

spine1 output parameter analysis
1e7
output-load
output-data-rate
output-packet-rate
time
1e11+1.501867e18

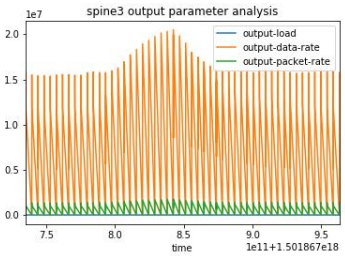

spine1 other parameter analysis
1e8
bandwidth
load-interval
peak-input-data-rate
peak-input-packet-rate
peak-output-data-rate
peak-output-packet-rate
reliability
time
1e11+1.501867e18

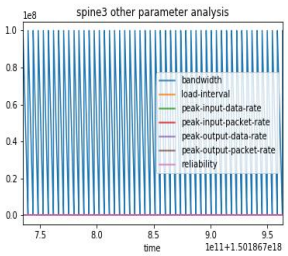

spine2 input parameter analysis
1e7
input-load
input-data-rate
input-packet-rate
time
1e11+1.501867e18

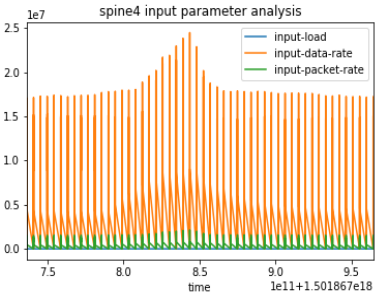

spine2 output parameter analysis
1e7
output-load
output-data-rate
output-packet-rate
time
1e11+1.501867e18

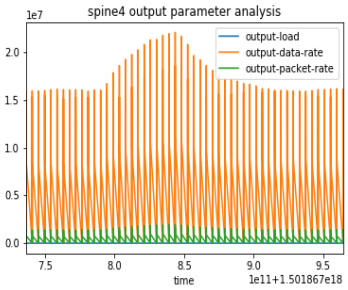

spine2 other parameter analysis
1e8
bandwidth
load-interval
peak-input-data-rate
peak-input-packet-rate
peak-output-data-rate
peak-output-packet-rate
reliability
time
1e11+1.501867e18

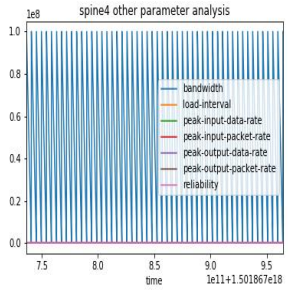












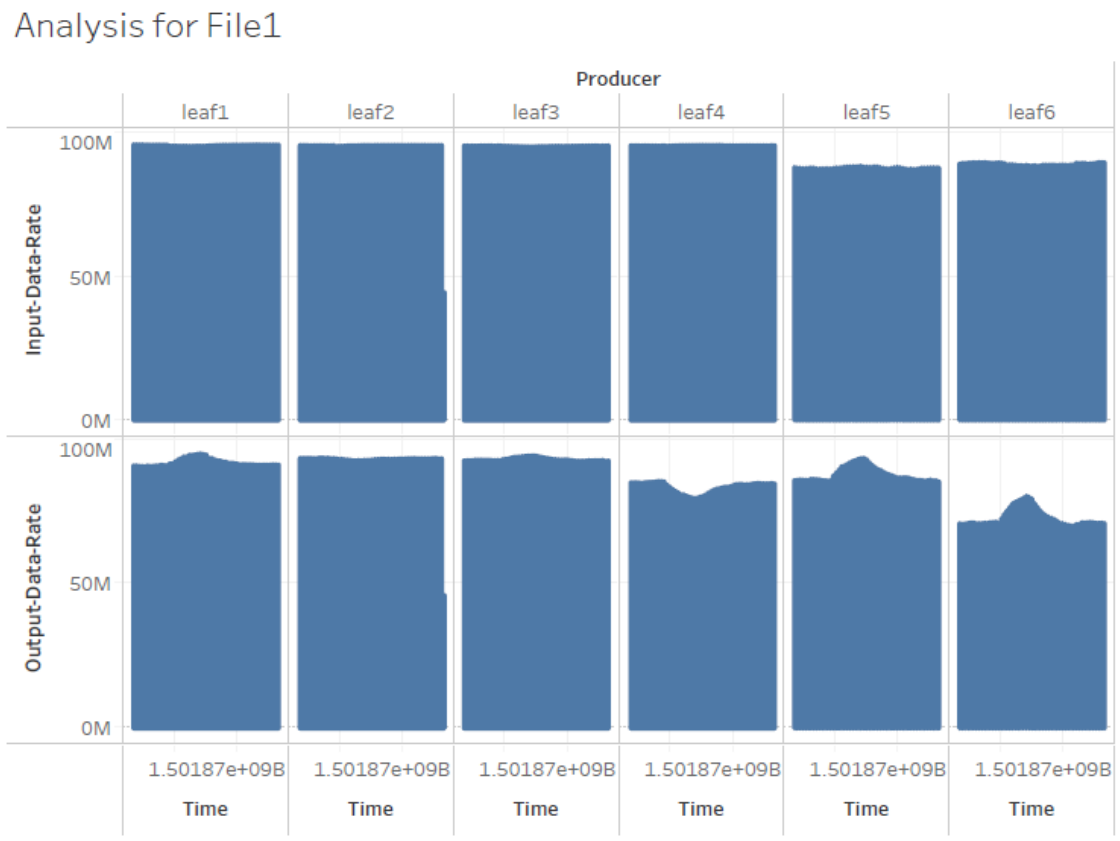
- From these plot, various insights can be drawn upon:-
    - Across the 15 producers that are showing disturbance only the input parameters and the output parameters are showing any variance and thus contain information. The other parameters (such as bandwidth, reliability, peak values) are not showing any variance even when disturbances in the input and output parameters is observed. Thus, we can conclude that for the next steps in the analysis we can consider those columns to be redundant.
    - For both the input parameters and output parameters across the 15 producers, the same trend (plot line) is being observed amongst the parameters. To reiterate the input-load plot, the input-data-rate plot and the input-packet rate plot all are following the same trend. [With the same being true for output parameters as well].
    - The input-data-rate and output-data-rate plots are deduced to be the best plots to analyze over various timestamps over the producers for every file to answer the questions being posed. [The variance of these two parameters is easily and clearly captured in the plots shown above]

With this initial analysis performed, we then plot the data (over Tableau, for ease of operation) for every producer over each file under the ‘individual file analysis’.

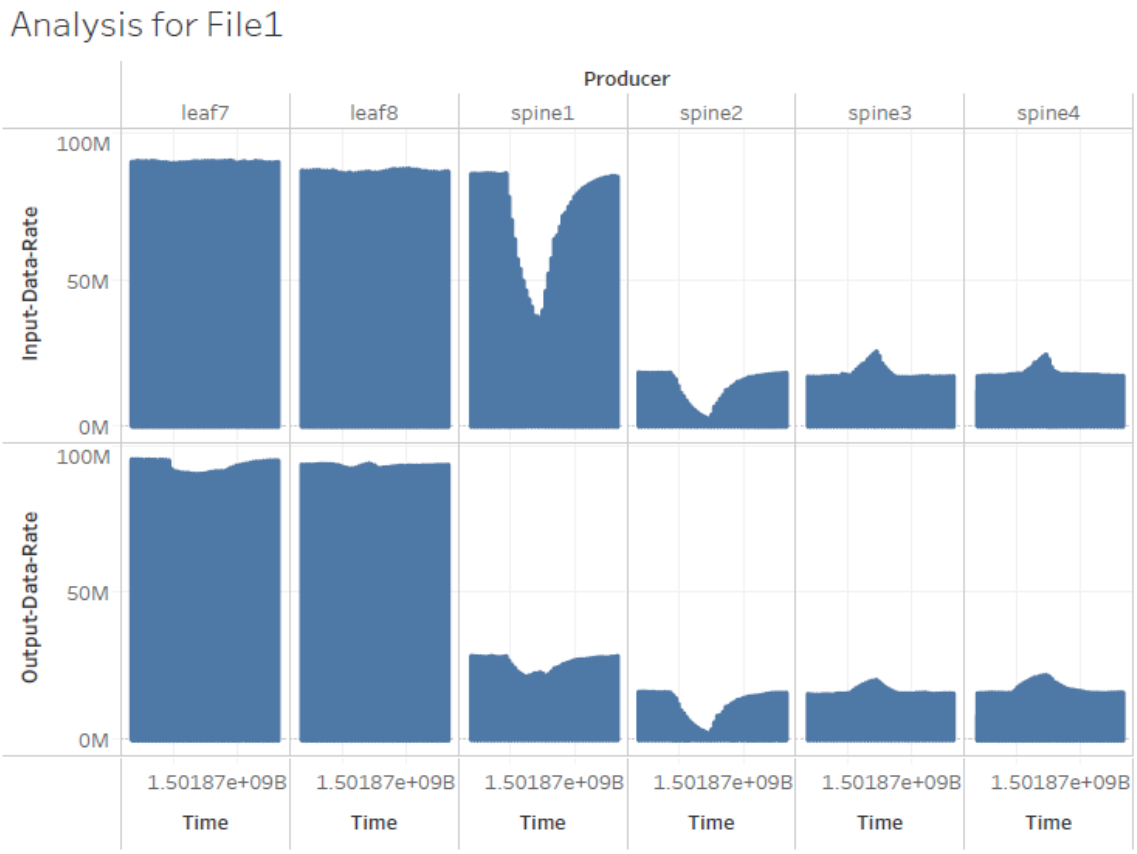
Q2) & Q3) Individual File Analysis

For all of the 7 file sets of csv data, a conjoined analysis for the input-data-rate and output-data-rate vs time for each of the 12 producers is performed and analysis about each file is detailed.

Analysis for File1:-



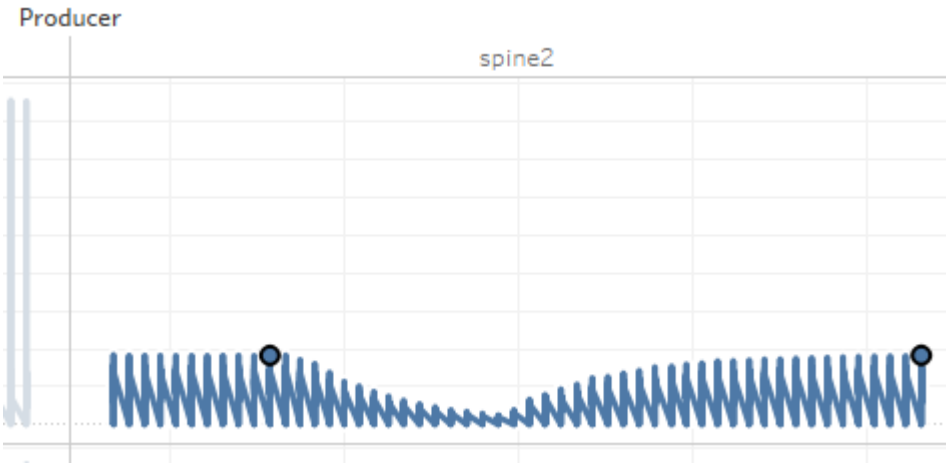
The trends of sum of Input-Data-Rate and sum of Output-Data-Rate for Time broken down by Producer. The view is filtered on Producer, which excludes dr01, dr02 and dr03.



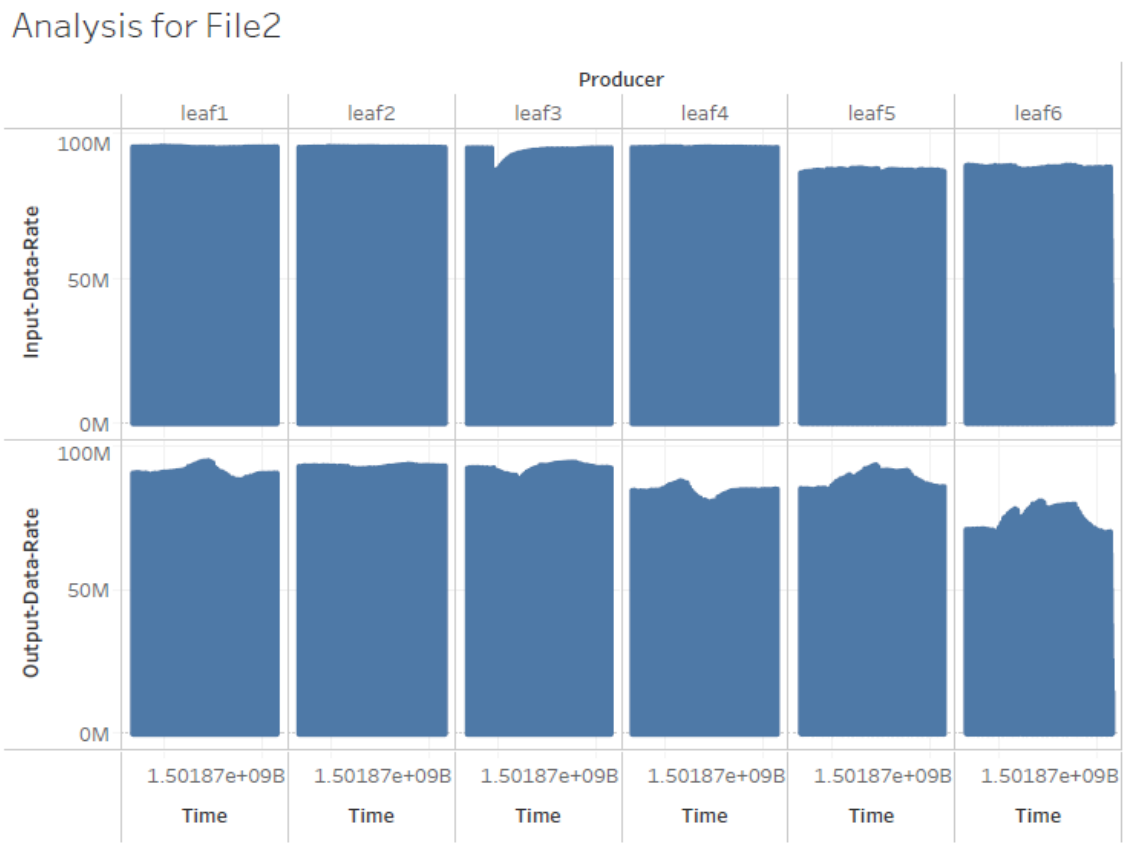
The trends of sum of Input-Data-Rate and sum of Output-Data-Rate for Time broken down by Producer. The view is filtered on Producer, which excludes dr01, dr02 and dr03.

From the plots, a clear dip in the input and output data rate of spine2 indicates that the spine has fallen. We can thus classify **this file as a case of One spine failing**. The spine failing affects the traffic at the spines 1, 3 and 4, to adjust the input and output data rates to compensate for the momentary LOS of spine2.

The time taken for the network to return to normal functioning will be the point differences (See graph below for highlighted points) between the recovery time and the crash time i.e. 1501867965347000000ns-1501867778770000000ns = 186577000192.00ns = 186.57 seconds

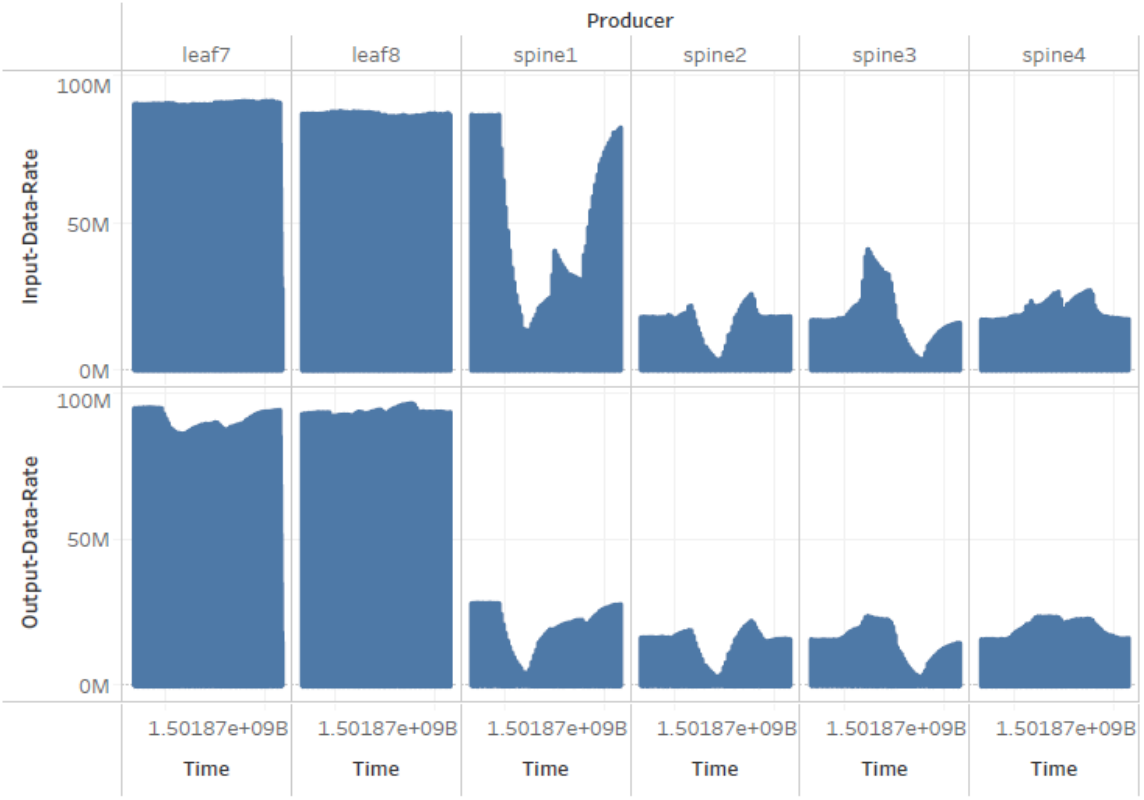


Analysis for File2:-



The trends of sum of Input-Data-Rate and sum of Output-Data-Rate for Time broken down by Producer. The view is filtered on Producer, which excludes dr01, dr02 and dr03.

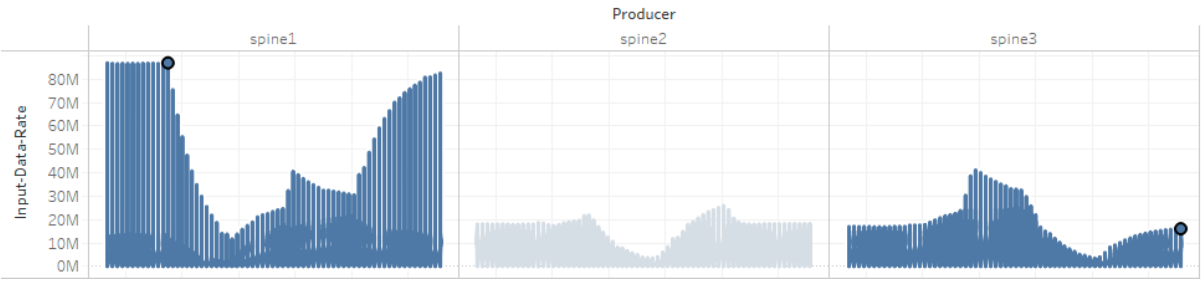
Analysis for File2



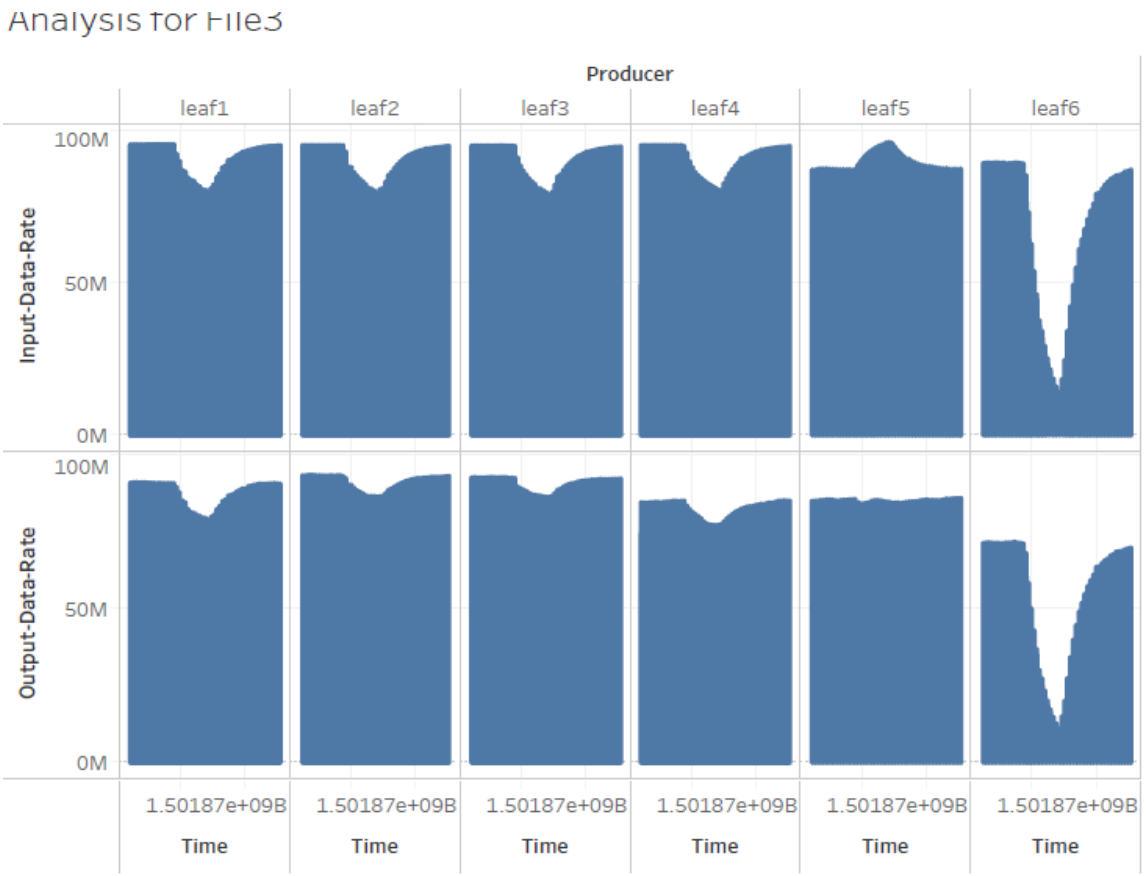
The trends of sum of Input-Data-Rate and sum of Output-Data-Rate for Time broken down by Producer. The view is filtered on Producer, which excludes dr01, dr02 and dr03.

From the plots, a clear dip in the input and output data rate of spine2 followed by spine3and then followed by spine 4 in immediate intervals of time indicate that three spine have fallen in sequence. We can thus classify **this file as a case of three spines failing in sequence**. The spines failing heavily increase the traffic at the spine 4(along with minor disturbances at the leaf level), to adjust the input and output data rates to compensate for the momentary LOS.

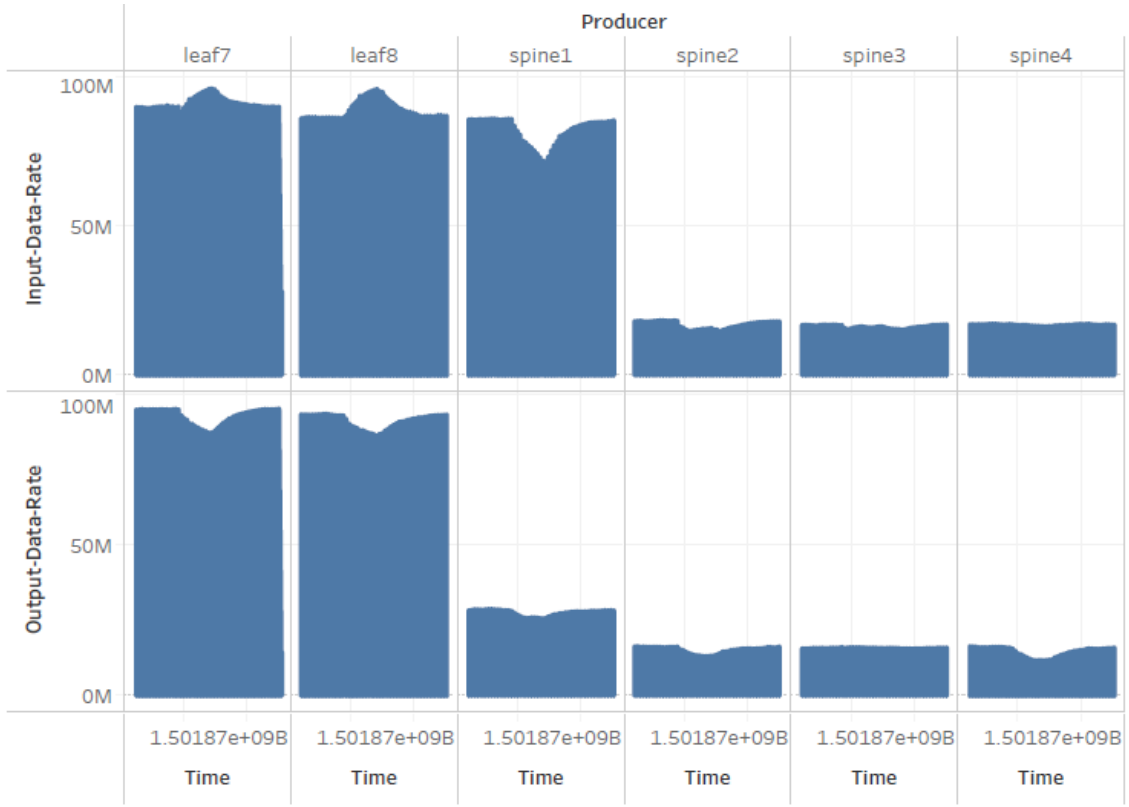
The time taken for the network to return to normal functioning will be the point differences (See graph below for highlighted points) between the recovery time and the crash time i.e. 1501868387642000000.00ns - 1501868628517000000.00ns = 240875000064ns = 240.87 seconds



Analysis for File3:-



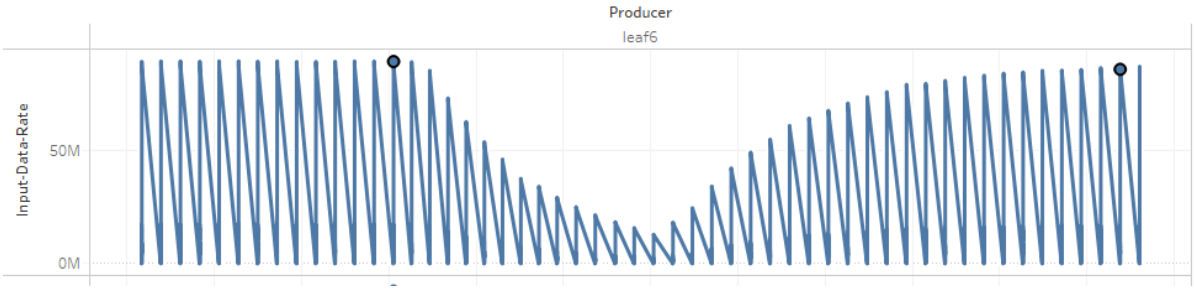
The trends of sum of Input-Data-Rate and sum of Output-Data-Rate for Time broken down by Producer. The view is filtered on Producer, which excludes dr01, dr02 and dr03.



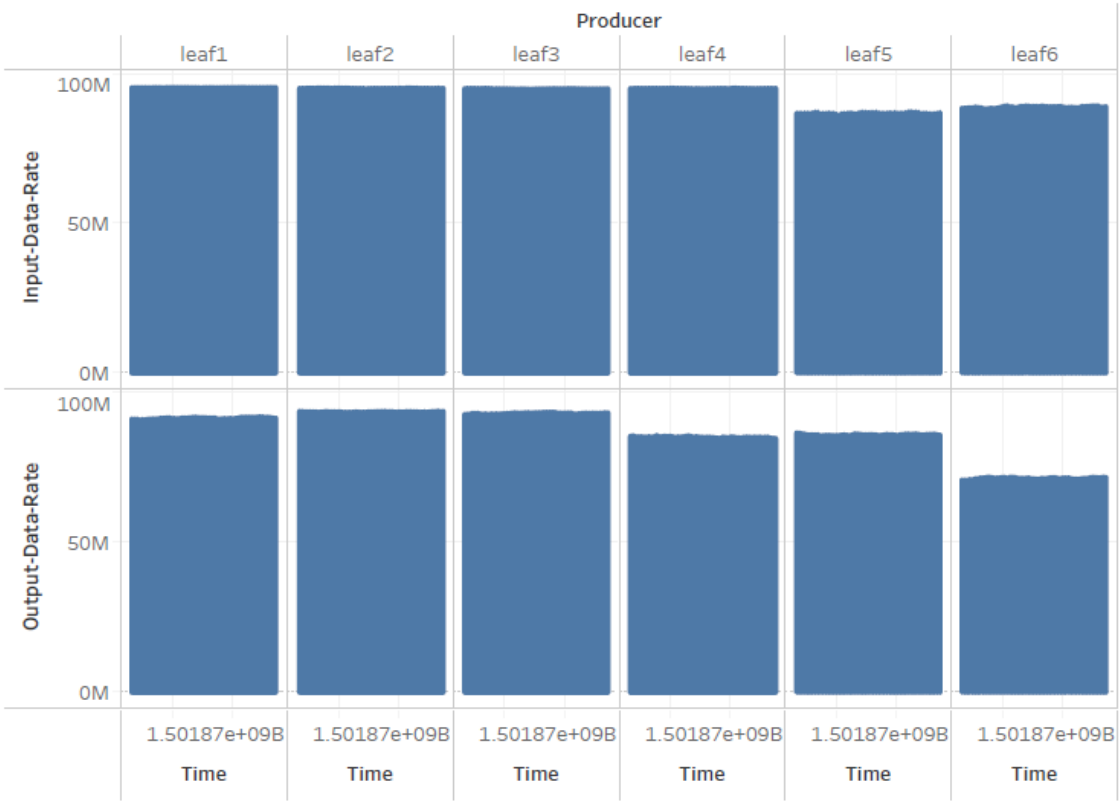
The trends of sum of Input-Data-Rate and sum of Output-Data-Rate for Time broken down by Producer. The view is filtered on Producer, which excludes dr01, dr02 and dr03.

From the plots, a clear dip and raise (recovery) in the input and output data rate of Leaf6 indicates that this leaf has crashed. We can thus classify **this file as a case of One leaf failing**. This leaf increases the input rates of the neighbouring leafs (5, 7 and 8) and correspondingly the data traffic at other leafs (1, 2, 3, 4) and the spines decreases.

The time taken for the network to return to normal functioning will be the point differences (See graph below for highlighted points) between the recovery time and the crash time i.e. 1501867481091000000ns- 1501867647649000000ns = 166557999872ns = 166.56 seconds



Analysis for File4:-

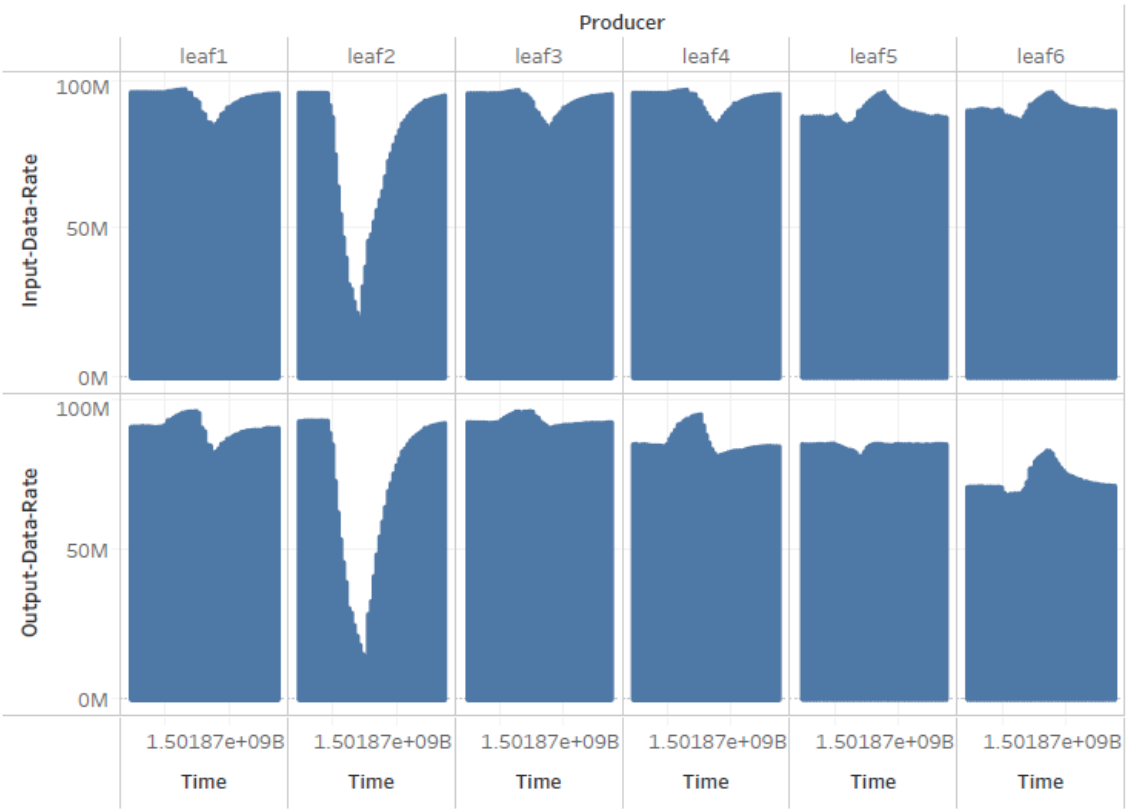


The trends of sum of Input-Data-Rate and sum of Output-Data-Rate for Time broken down by Producer. The view is filtered on Producer, which excludes dr01, dr02 and dr03.

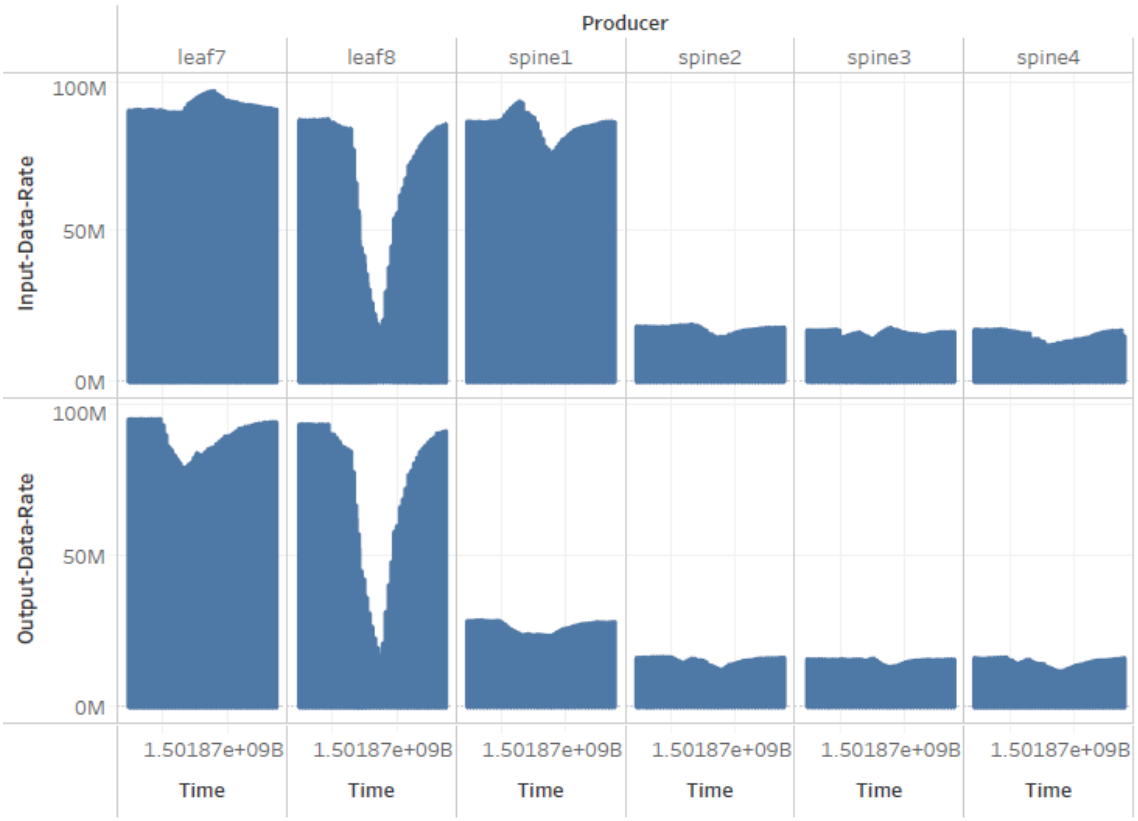
The trends of sum of Input-Data-Rate and sum of Output-Data-Rate for Time broken down by Producer. The view is filtered on Producer, which excludes dr01, dr02 and dr03.

Observing the plots for file4, we see that there are no major dips in the input and output data parameters. The levels of traffic are also steady across the time intervals measured and thus this can be concluded as a case of **NORMAL OPERATION**. This file from now on will be used as a reference for normal operation understanding. Certain remarks about the spine operation can be made, that spine1 is the primary data input and output channel (for between the leafs) and the other three spines are up for redundancy.

Analysis for File5:-



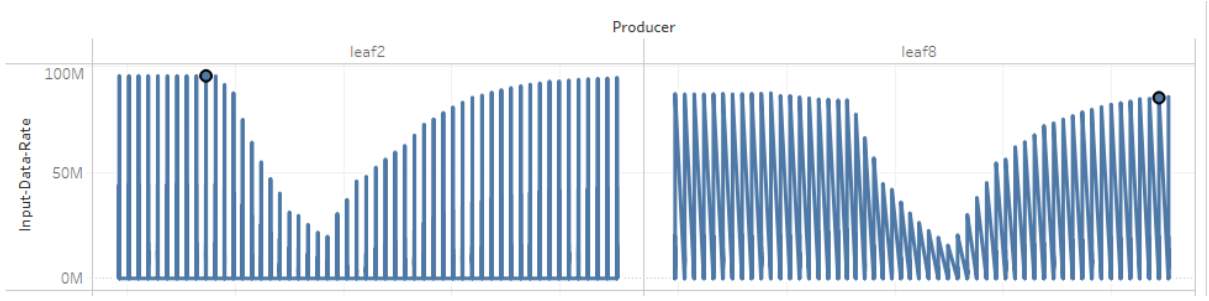
The trends of sum of Input-Data-Rate and sum of Output-Data-Rate for Time broken down by Producer. The view is filtered on Producer, which excludes dr01, dr02 and dr03.



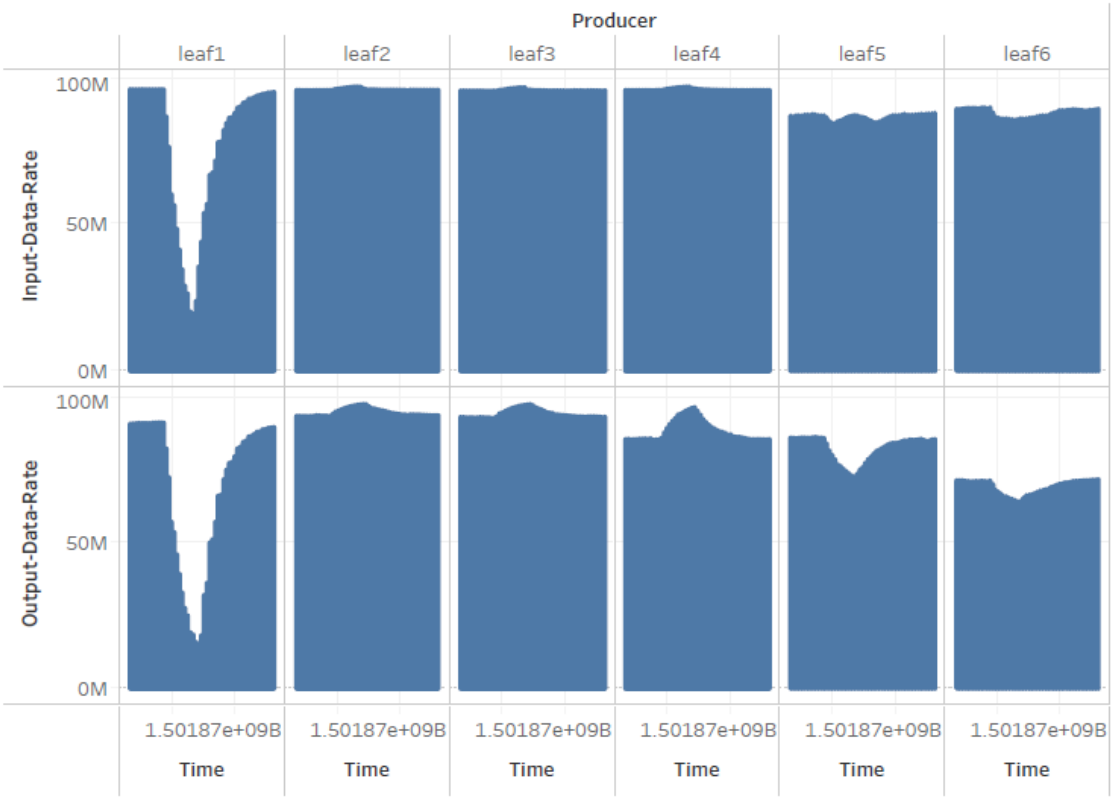
The trends of sum of Input-Data-Rate and sum of Output-Data-Rate for Time broken down by Producer. The view is filtered on Producer, which excludes dr01, dr02 and dr03.

From the plots, a clear dip in the input and output data rate of leaf2 followed by leaf8 in immediate intervals of time indicate that the two leafs have fallen in sequence. We can thus classify **this file as a case of two leafs failing in sequence**. To offset the failure of leaf2 and leaf8 the neighbouring leafs (leafs 1, 3 and 4 for leaf2 and leafs 5, 6 and 7 for leaf8) increase their output-data-rates by a considerable margin. There is a slight adjustment of input/output data rates at spine1 to tackle for the failure and recovery of the aforementioned two leafs.

The time taken for the network to return to normal functioning will be the point differences (See graph below for highlighted points) between the recovery time and the crash time i.e. 1501868086393000000ns - 1501868272007000000ns = 185614000128ns = 185.61 seconds



Analysis for File6:-

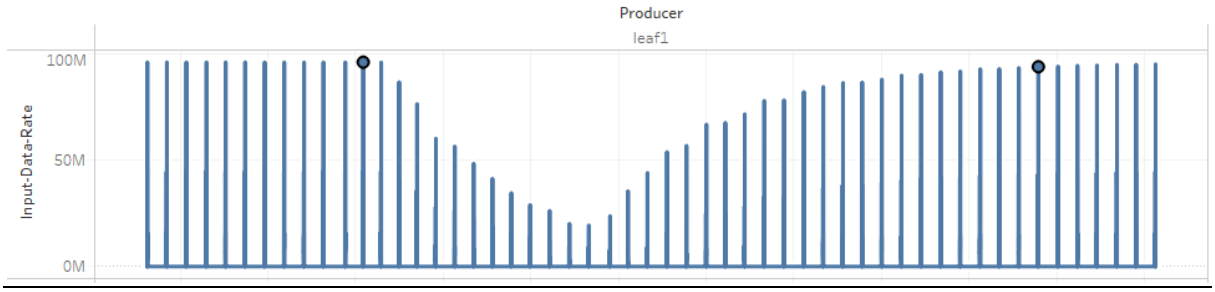


The trends of sum of Input-Data-Rate and sum of Output-Data-Rate for Time broken down by Producer. The view is filtered on Producer, which excludes dr01, dr02 and dr03.

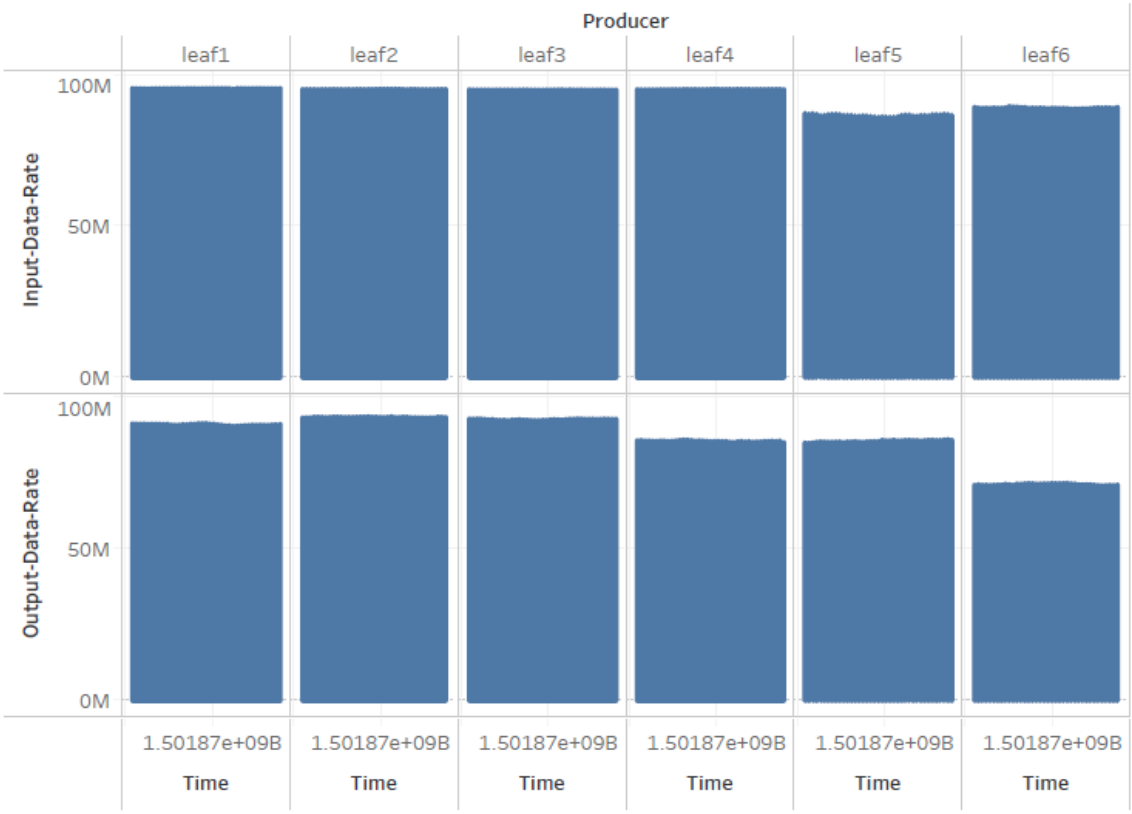
The trends of sum of Input-Data-Rate and sum of Output-Data-Rate for Time broken down by Producer. The view is filtered on Producer, which excludes dr01, dr02 and dr03.

From the plots, a clear dip and raise (recovery) in the input and output data rate of Leaf1 indicates that this leaf has crashed. We can thus classify **this file as a case of One leaf failing**. This leaf increases the input rates of the neighbouring leafs (3 and 4) and correspondingly the data traffic at other leafs (5, 6, 7, 8) decreases. The increase of data rate at spine1 seems to be a minor disturbance and cannot be accounted for in the analysis of the leaf1 crash.

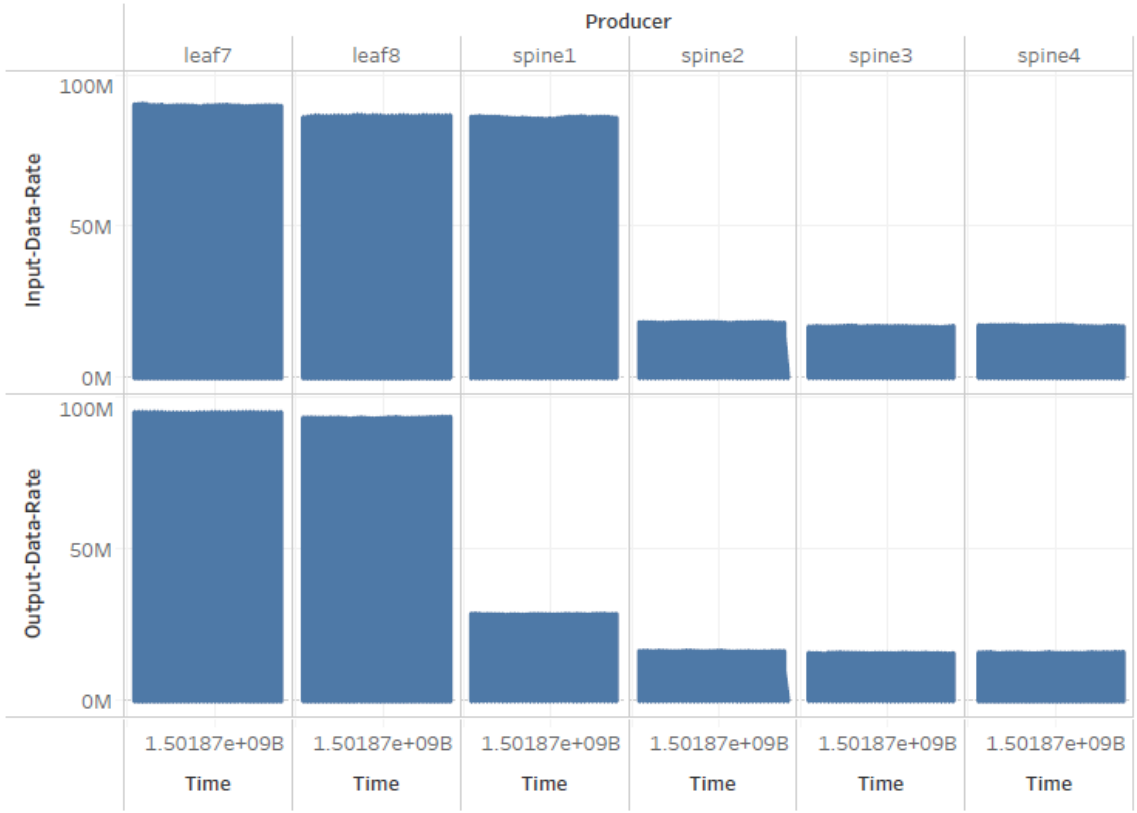
The time taken for the network to return to normal functioning will be the point differences (See graph below for highlighted points) between the recovery time and the crash time i.e. 1501867181790000000ns - 1501867335834000000ns = 154043999744ns = 154.04 seconds



Analysis for File7:-



The trends of sum of Input-Data-Rate and sum of Output-Data-Rate for Time broken down by Producer. The view is filtered on Producer, which excludes dr01, dr02 and dr03.



The trends of sum of Input-Data-Rate and sum of Output-Data-Rate for Time broken down by Producer. The view is filtered on Producer, which excludes dr01, dr02 and dr03.

Observing the plots for file7, we see that there are no major dips in the input and output data parameters. The levels of traffic are also steady across the time intervals measured and thus this can be concluded as a case of **NORMAL OPERATION**.

This concludes the analysis for each file-sets drawing conclusions and answering all of the questions asked for.

Appendix:

Python code for the initial file1 data visualizations:-

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
## Importing all the relevant libraries
raw_data_file1 = pd.read_csv('File1.csv')
##reading the  csv file using pandas
data_frame=pd.DataFrame(raw_data_file1)
data_frame=data_frame.drop(data_frame.columns[[0,1]],axis=1)
##initializing a dataframe and dropping irrelevant columns
no_of_elements=data_frame.Producer.unique()
for i in no_of_elements:
   data_frame_temp=data_frame[data_frame.Producer==i]
   data_frame_temp.plot(data_frame_temp.time,y=['input-load','input-data-rate','input-
packet-rate'])
   plt.title(i+' input parameter analysis')
   plt.savefig('parameter_visualizations_python\\input_parameter_analysis %s.png' %i)
   data_frame_temp.plot(data_frame_temp.time,y=['output-load','output-data-
rate','output-packet-rate'])
   plt.title(i+' output parameter analysis')
   plt.savefig('parameter_visualizations_python\\output_parameter_analysis %s.png' %i)
   data_frame_temp.plot(data_frame_temp.time,y=['bandwidth','load-interval','peak-input-
data-rate','peak-input-packet-rate','peak-output-data-rate','peak-output-packet-rate','reliability'])
   plt.title(i+' other parameter analysis')
   plt.savefig('parameter_visualizations_python\\other_parameter_analysis %s.png' %i)

##The crux of the code-segment: For each producer plot three graphs, 1.Taking all input
parameters 2.Taking all output parameters 3.Taking other parameters w.r.t time as the x-axis. After
plotting save the graphs in the mentioned path.
```

THANK YOU