**PROJECT – 1 NIKHIL MOHAN (110167666)**

**Introduction:** The task is to analyze network flow traffic data and come up with various patterns, anomalies, the type and nature of data etc.

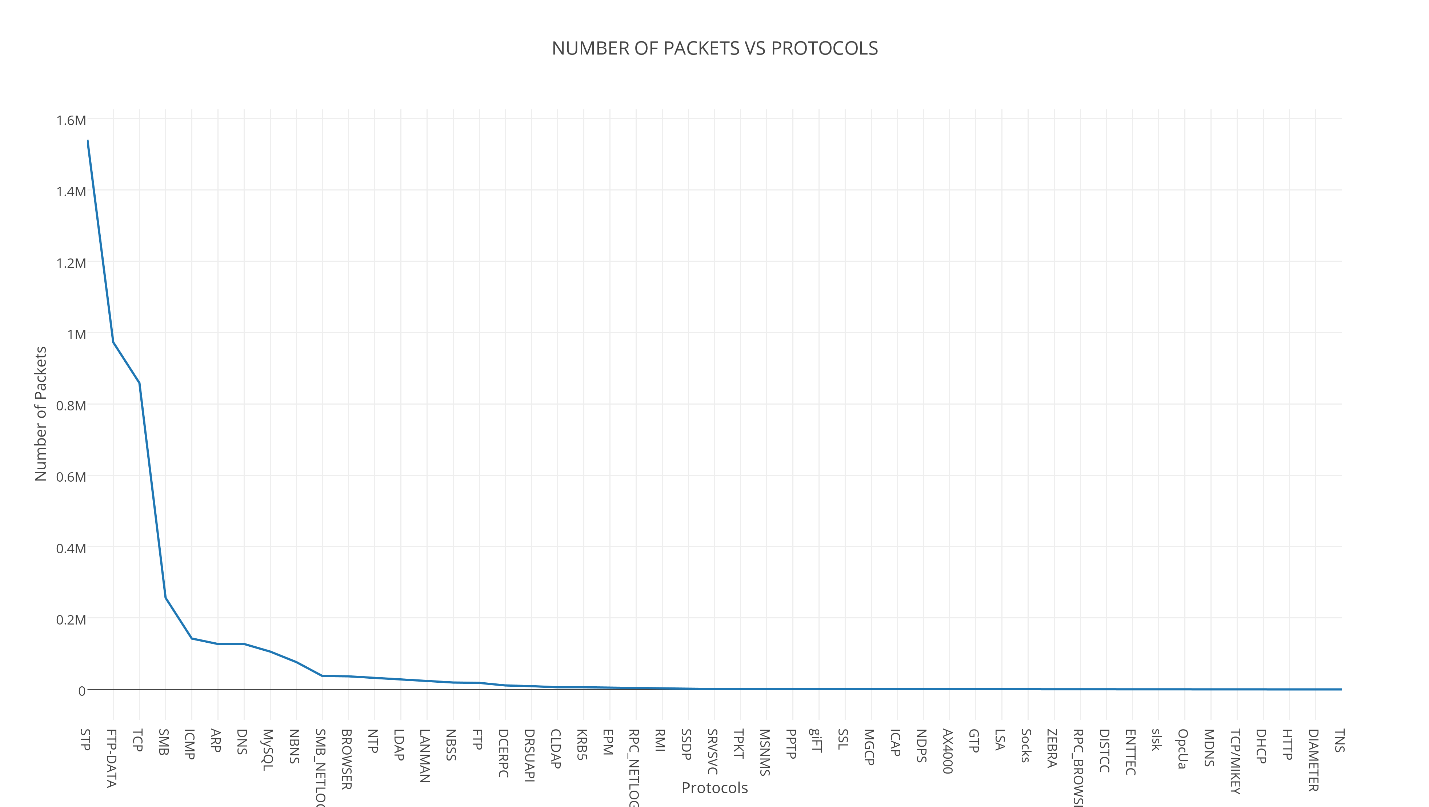
**Accomplished Tasks:**

**Fig 1 – Fig 5** gives the in-depth view of the data, the pattern and the findings in the data. This is done by loading the data into **mysql** and querying the parameters. The output is plotted using **plotly**.

**Fig 6 – Fig 10** gives the overall nature of the network and explains several critical factors in the network like degree distribution, centrality and clustering of the networks. This is done by loading the network data into the **Graph Visualization tool (Gephi)** and applying various statistics to the data.

Each of the above analysis is done with respect to **AllNetworkData.csv data** and their detailed analysis is as below:

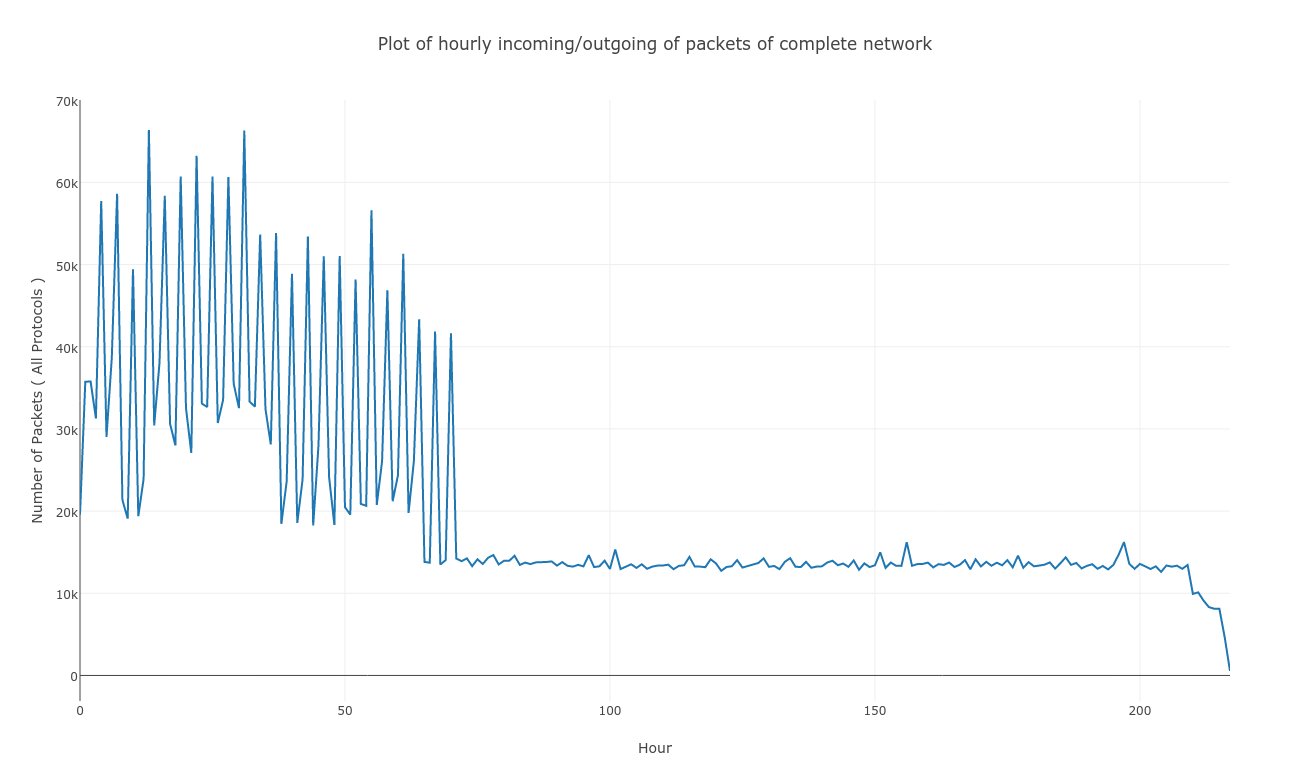
**Number of Packets vs Protocols Trend: (Fig: 1)**



**Metric: X - Axis: Protocols:** These are the unique protocols available in the given data. **Y – Axis: Number of packets:** These are the number of packets exchanged between the machines in the network.

This plot shows the usage of protocols in the network. Clearly **STP, FTP-DATA, TCP** dominate the network. These are the highly used protocols in the given data.

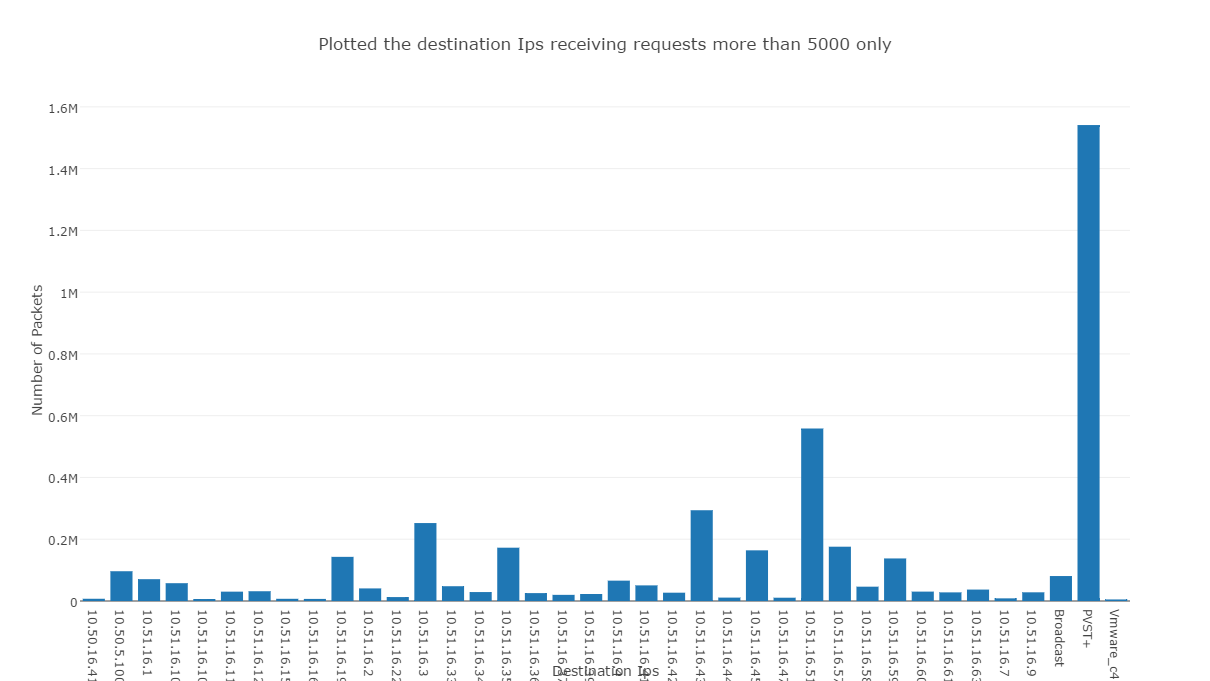
**Hourly Incoming/Outgoing packets of the complete network: (Fig: 2)**



**Metric: X - Axis: Hour:** Number of hours available in the given data. **Y – Axis: Number of packets:** These are the number of packets exchanged between the machines in the network.

This plot shows that the network was quite active for the first 50 hrs with few bursts in between and later on it has become stable with constant number of packet transactions.

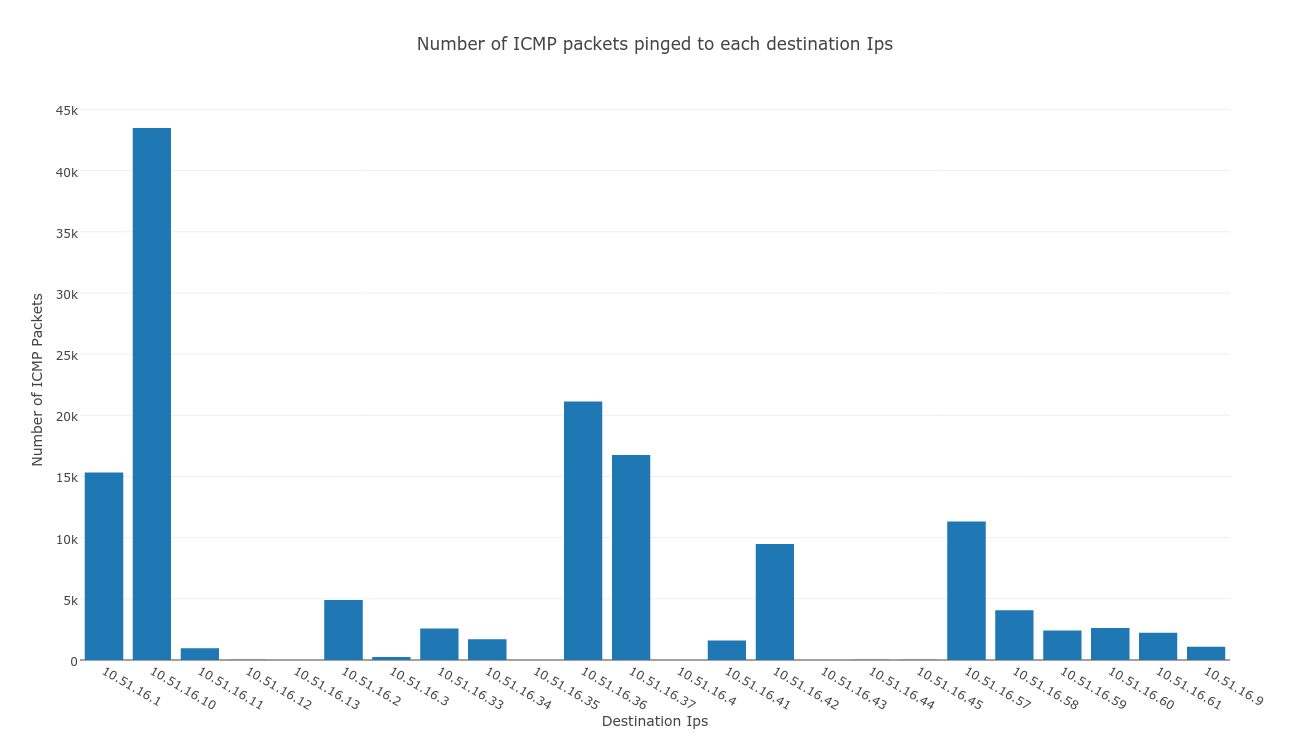
**Destination Ips receiving more than 5000 requests: (Fig: 3)**



**Metric: X - Axis: Destination Ips:** These are the destination Ips in the given data. **Y – Axis: Number of Packets:** These are the number of packets exchanged between the machines in the network.

The above plot shows destination Ips getting more than 5000 requests over the complete time given in the data. As shown, **PVST+** gets abnormally high incoming requests followed by 10.51.16.51.

**Number of ICMP packets reaching destination Ips: (Fig: 4)**



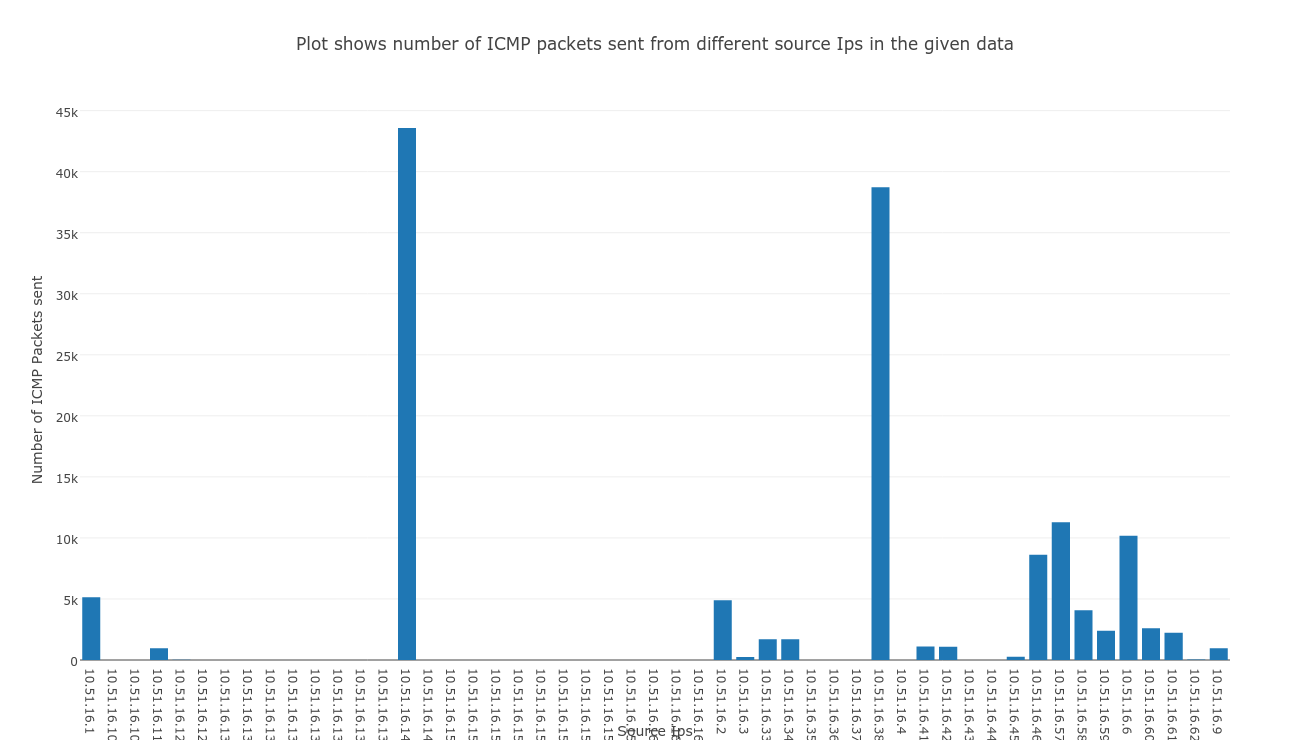
**Metric: X - Axis: Destination Ips:** These are the destination Ips receiving packets through ICMP protocol in the given data. **Y – Axis: Number of ICMP Packets:** These are the number of packets exchanged between the machines in the network.

ICMP protocol in particular was considered as it is a ping protocol and a highly used protocol to cause DOS attack. According to the analysis, 10.51.16.10 receives abnormally high incoming requests for the given total time in the data.

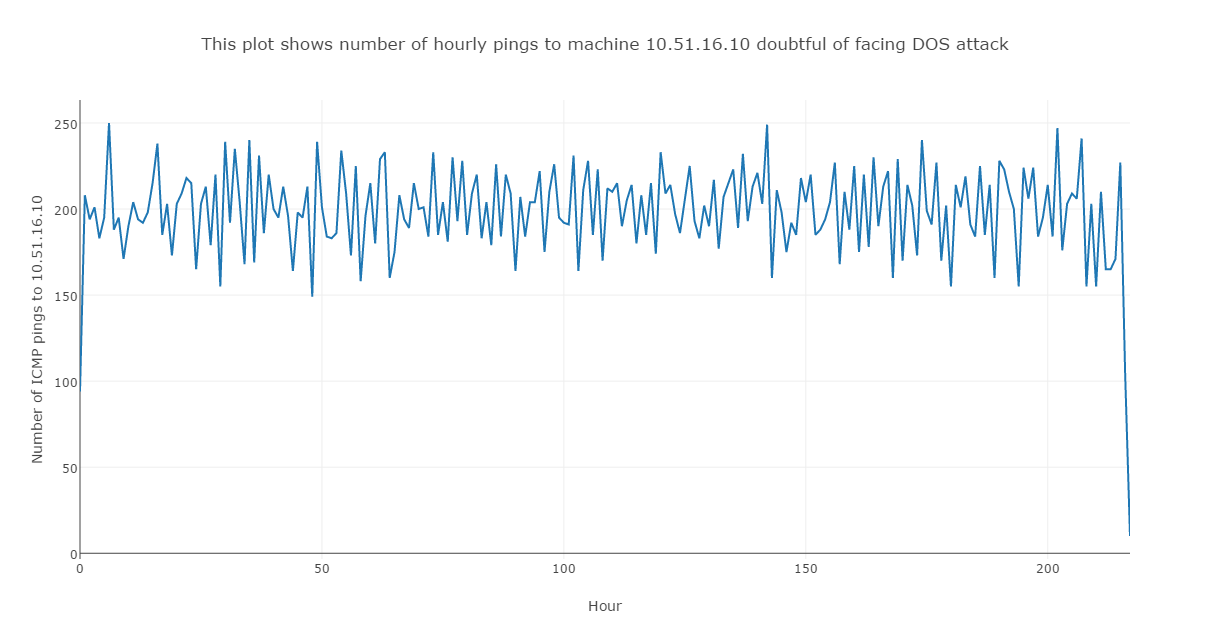
**ICMP pings from sources (Fig: 5)**

**Metric: X - Axis: Source Ips:** These are the source Ips in the given data related to ICMP protocol. **Y – Axis: Number of ICMP packets sent:** These are the number of ICMP packets exchanged between the machines in the network.

This plot shows the possible number of sources of ICMP protocol to analze the anomaly in the number of requests sent through ICMP protocol. The analysis shows that **10.51.16.14 and 10.51.16.36** are the highest contributors.

****

**ICMP pings to a particular destination: (Fig: 6)**



**Metric: X - Axis: Hour:** Number of hours available in the given data. **Y – Axis: Number of ICMP packets to 10.51.16.10:** These are the number of ICMP packets exchanged between various machines and 10.51.16.10.

The above plot is the number of hourly ICMP packets sent to the Ip: **10.51.16.10** that received the abnormally high incoming ICMP packets (shown in Fig: 5). The number of ICMP packets is consistently high through the time frame to this particular Ip. And the info shows **“destination unreachable host unreachable”** for almost evert packet that sent **(43625 packets out of 43271)**. This could be a potential DOS attack as there are previous cases pointing to the same judgement which is called **“ICMP destination unreachable host unreachable attack”**.

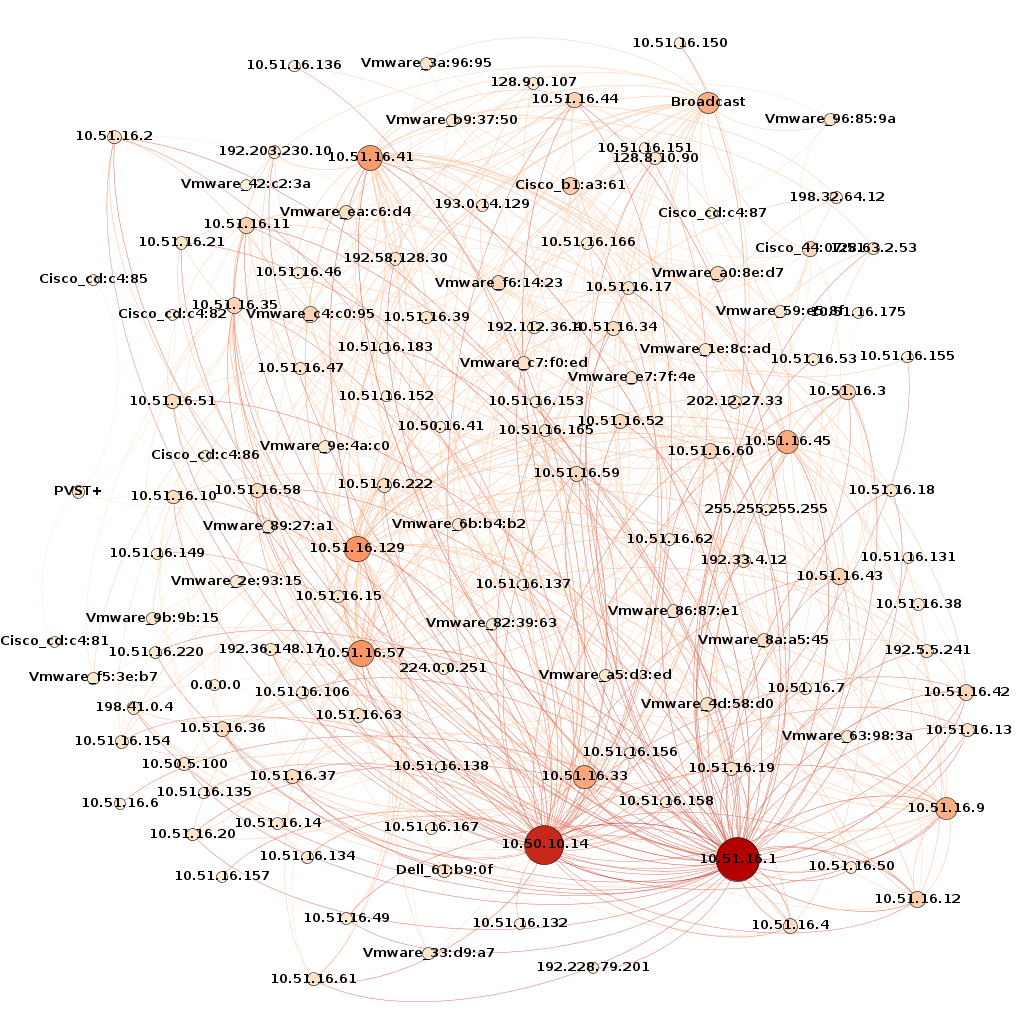
**Graph Distance Report of network and Other Observations:**

**Diameter:** 5

**Average Path length:** 2.239785197291618

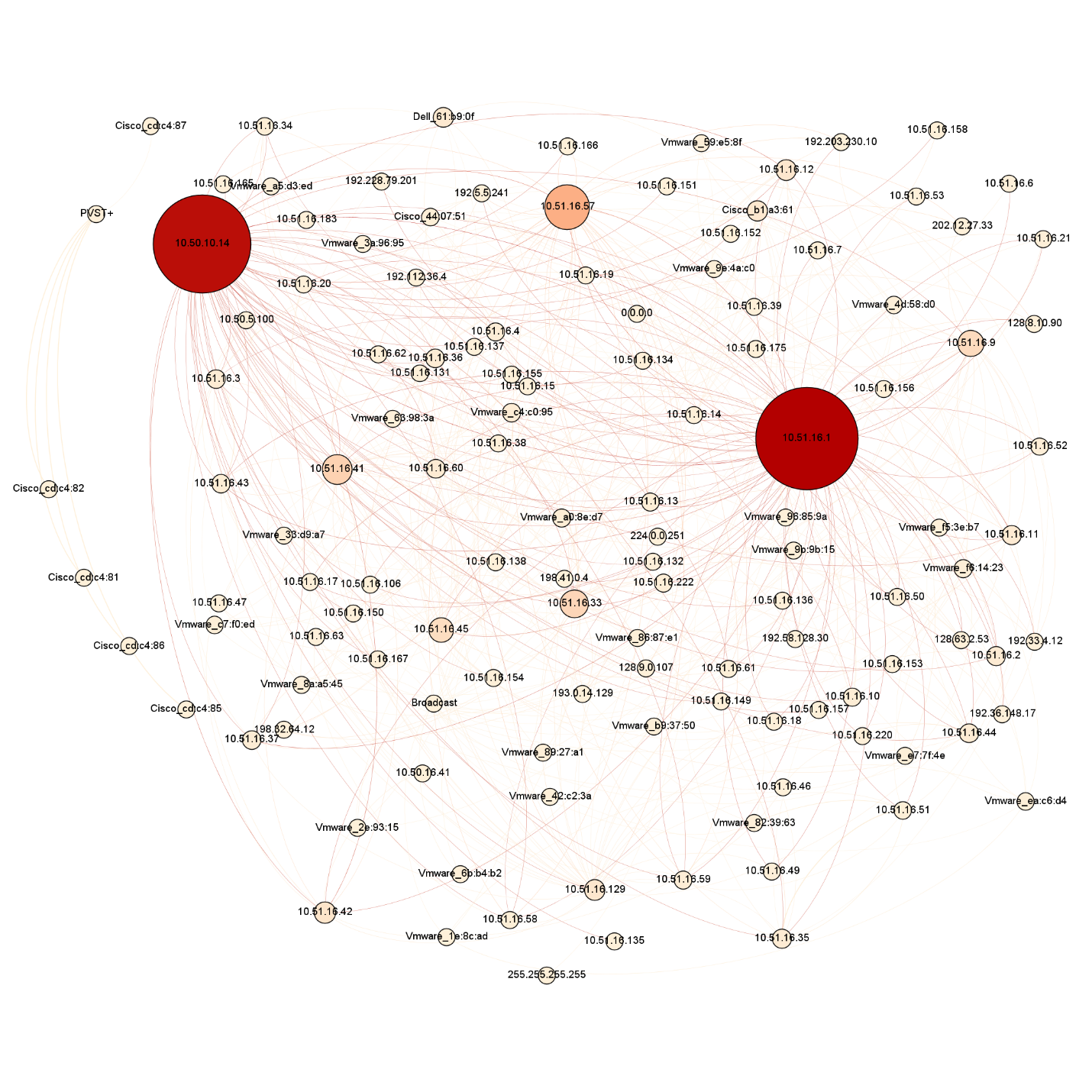
**Number of shortest paths:** 4283

**Network Based on Degree: (Fig 7)**



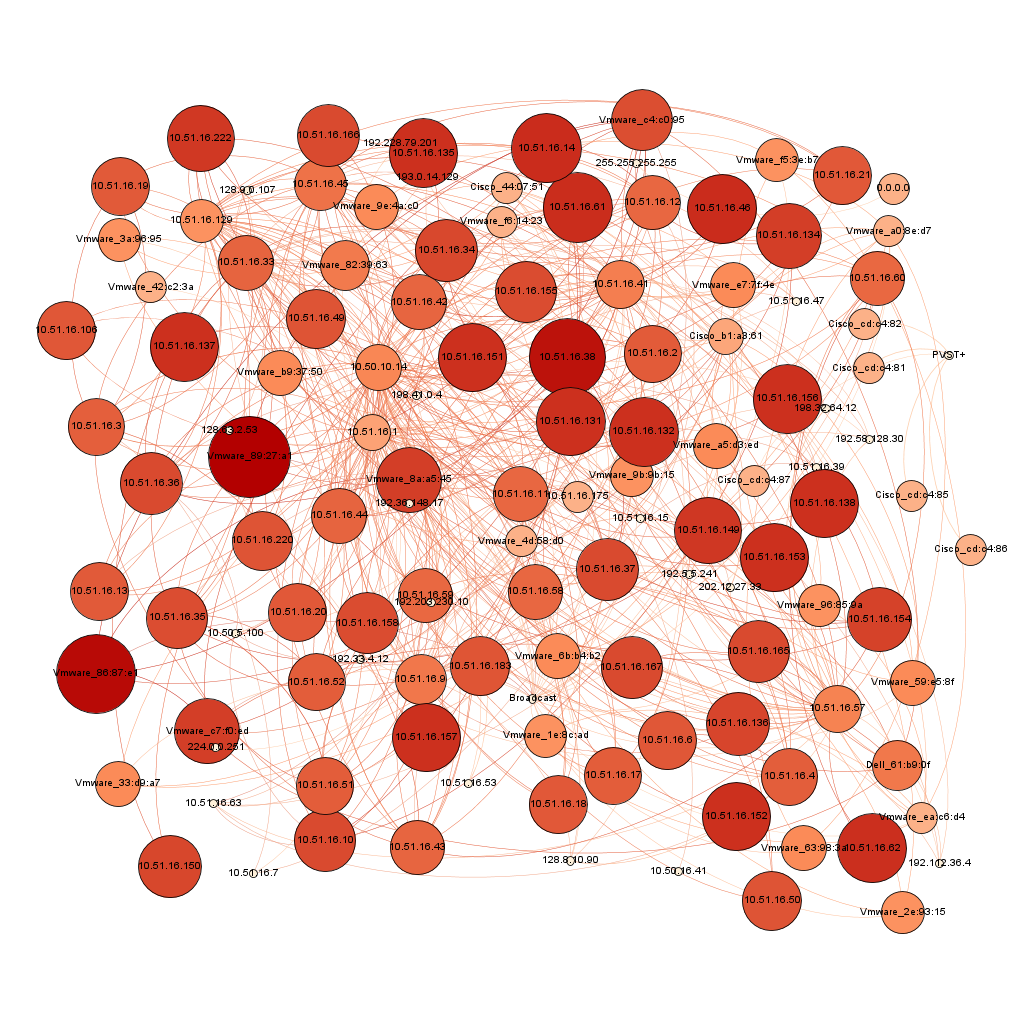
The above graph gives the picture on the degree of a particular node (Ip). The higher the degree(in-degree and out-degree), the larger the node and darker the node. We see **10.50.10.14, 10.51.16.1** almost with the highest number of degree in the network followed by **10.51.16.129, 10.51.16.57, 10.51.16.33, 10.51.16.9, 10.51.16.41, 10.51.16.45** approximately.

**Betweenness Centrality: (Fig: 8)**



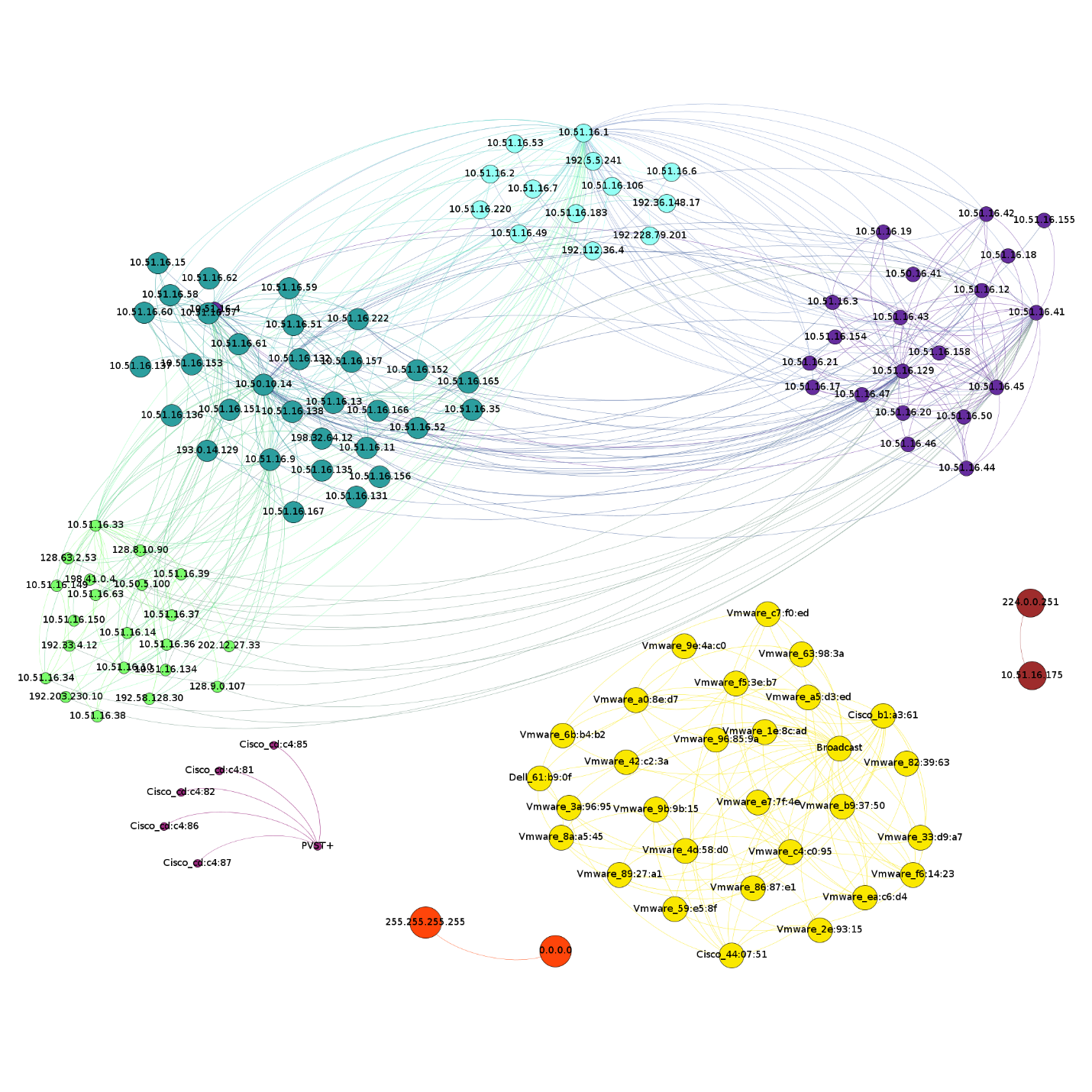
The above graph gives information on Betweeness Centrality. It indicates node’s centrality in a network. It is equal to number of shortest paths from all vertices to all others that pass through that node. This measures how critical a node is to information flow in the network. So if this node is attacked or brought down, then there is a serious concern. In the above graph, we see Ips **(10.50.10.14, 10.51.16.1, 10.51.16.57, 10.51.16.33, 10.51.16.41, 10.51.16.9)** in decreasing order of their impact to the network.

**Closeness Centrality: (Fig: 9)**



The above graph gives the information on closeness centrality. This is the average distance from a given node to all the other nodes in the network. This determines the speed with which a randomly walking message can reach a node from elsewhere in the network. **The above graph shows that most of the nodes are strongly connected with their average path length 2.**

**Cluster: (Fig: 10)**



In total the given network can be broken into 8 clusters. There are 5 independent networks in the data given as shown above. The clustering is done using the modularity class in Gephi Visualization tool. The different colors represent different clusters and the metric for the cluster is the weight of the graph.

**Conclusion:**

The analysis has been carried out both at the Ip level and at the network level. The behavior of the network and possible limitations of the network has been brought to notice by finding the weak links (Susceptible nodes). Also possible suspicious attacks on the network has been determined on the given data. Finally the clustering technique showed the possible variety in the network. So all the possible analysis of a network has been carried out. These observations give us a way to substantially improve the network depending on the topology and prevent loopholes.