# Existing Systems

## What is DDoS attack

Distributed Denial of Service (DDoS) attack is a way to jam a host network or its resources with a large number of data packets[[1]](#footnote-23) or connections, so that the host becomes disabled. There are different types of DDoS attacks such as : 1. Volume-based, e.g. SYN Flood Attacks, in which the victim is flooded with a high volume of Transmission Control Protocol (TCP)/ User Datagram Protocol (UDP) packets or connections. 2. Application-based, in which an application such as Domain Name Service (DNS), Voice over Internet Protocol (VoIP) or Hypertext Transfer Protocol (HTTP) are attacked. 3. Low-rate, in which the attacker exploits a vulnerability in the application design, e.g. Slowloris.

## Challenges in dealing with DDoS attacks

Many DDoS attacks happen every day, some come in the news while many remain unnoticed. The real challenge in detecting and defending the DDoS attack is because of its dynamic nature. The source[[2]](#footnote-25) is not a single node or a system on the Internet but there can be many systems participating in DDoS attack, and often these systems are distributed over different regions of the Internet. Also, the source of the packet is often spoofed[[3]](#footnote-26), which makes harder to know the actual IP address of the system from where attack is originated because original attack source is changed in spoofed data packet. In addition to that, many times the source system itself is not aware that it is compromised, and it is being used as a bot by an attacker to launch DDoS attack.

As the source address can’t be a reliable way to know the attack source (because of spoofing), detecting and mitigating attack at the destination[[4]](#footnote-27) is not very useful. Destination may know that the attack is happening but to stop it happening it will have to block all the incoming traffic including the legitimate traffic. To avoid this, companies such as Cisco, Netgear have come up with some solutions. Many of the solutions provided by those giants, or most of the research that is done in this field is focused on collecting network traffic flow information at routers(gateways). A flow consists of several Internet packets captured during a fixed time interval. Router send that flow information to the central system for the analysis. Central system is a hardware and software infrastructure which is capable of processing and analyzing large flow information.

Some of the major protocols which are widely used for flow collection and analysis are, Internet Protocol Flow Information Export (IPFIX) protocol created by the Internet Engineering Task Force (IETF), Ciscos NetFlow and Sflow (Sampled flow). These protocols have defined standard way to export the flow information from router and similar devices. All these flow monitoring protocols gather information and send the consolidated flow information to the centralized server where user can login and perform functions; such as Security Monitoring, Bandwidth monitoring, Resource Management, Traffic Analysis, Performance Management etc. On such systems, there are some modules which are specifically used for anomaly/DDoS detection.

E.g. Cisco netflow has flow exporter, collector and analysis modules. Flow exporter modules are installed on routers. The routers which are having flow exporter modules, send flow information to the collector module installed on the server. Along with the collector module, server also has analysis module which can be used to detect different patterns in the flow.

These technologies scales well and can be sufficient to indicate trends in network traffic but they have limitations. 1) They are not cross platform, e.g. router with Sflow protocol can’t work with Cisco routers. 2) They involve setting up expensive hardware which acts as collector server. 3) Source address is used for flow analysis which is not reliable due to IP spoofing in the case of DDoS attack.

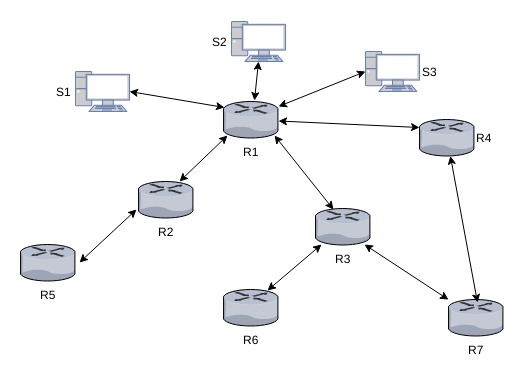
# DDoS Detection and Mitigation

DDoS attacks can be detected by checking if there is any anomalous behavior in the network traffic, such as, a sudden increase in the number of packets going to a destination. This can be done at the server by observing all the incoming traffic or it can be done by observing all the outgoing/incoming traffic at the ISP’s or at every router. Attack can be mitigated if the anomalous packets are blocked from reaching their destination.

From the previous section we know that the Internet device-based (i.e. the router-based) flow analysis can be useful for anomaly detection, but it has limitations. We don’t want to set up expensive hardware, we want to have a protocol or a system which is compatible with other routers. Also, we don’t want source IP address for detection analysis. So, if we can come up with a way by which we can detect anomaly in the traffic at the intermediate devices on the Internet such as gateway devices (routers) and create a communication protocol between such gateway devices and the destination server or network, then better decisions on regulating the packet flow can be taken.

# Network Functioning

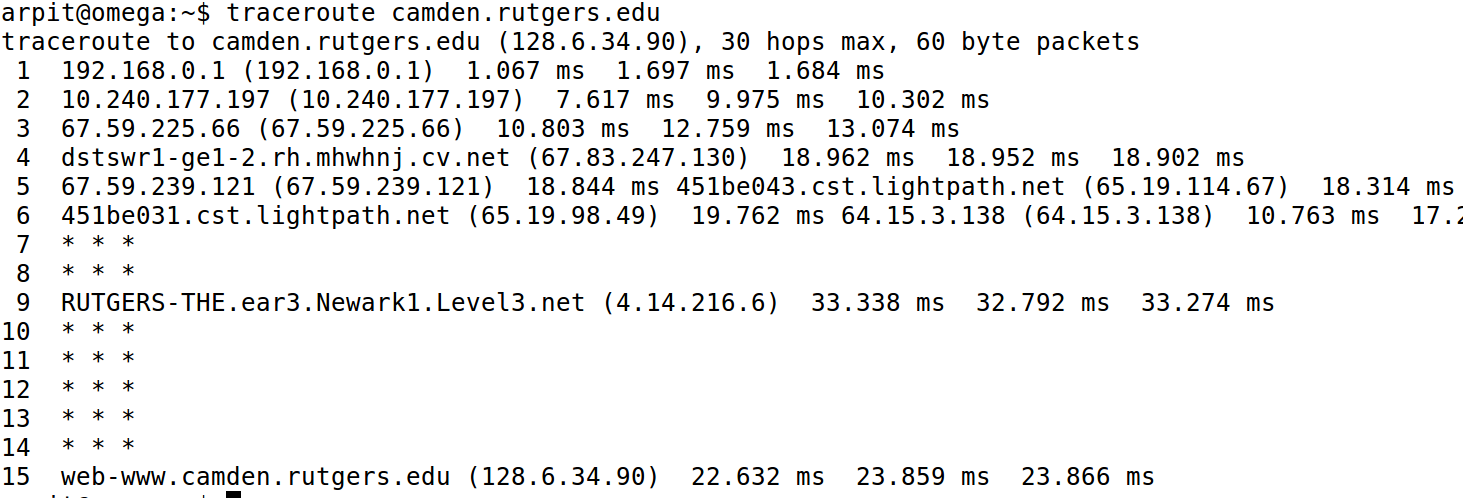
A switch creates a network and router connects those networks. A router links a computer to the Internet through other routers. Routers are the backbone of the network who helps to forward packet from one point to other point on the Internet. Every packet traveling on the Internet goes through a router. Router knows where the packet is destined, hence it could serve as first point of knowledge about the change in the packet flow information for a destine network. Each router has interfaces to which hosts, or other networks are connected. So, router is aware to whom it is connected. Router uses protocols to communicate between other networking devices and by that it gathers knowledge about other networks or routers on the Internet. Internet Control Message Protocol (ICMP) is one of the most frequently use protocol by routers for sharing operational information.



Network Example

Let’s illustrate this using an example. In the above figure we can see that host S1, S2, S3 are connected to a router R1. Router R1 is connected to the Internet through router R2, R3 and R4, thus every packet reaching to the system S2 is coming from either of these three routers. All three routers are located in different geographical region. Most of the websites are regional, either county, state or national (If we leave out few global websites) and hence they are mostly accessed from those regions it is meant for, e.g. Rutgers University website is accessed mostly from the eastern region of United States and that too mostly from the New Jersey State or the Philadelphia region.

Using traceroute we can see how many hops[[5]](#footnote-31) away the destination is. Following is one of the captured traceroute for Rutgers University website.



Trace Route: All the routers in the path to destination

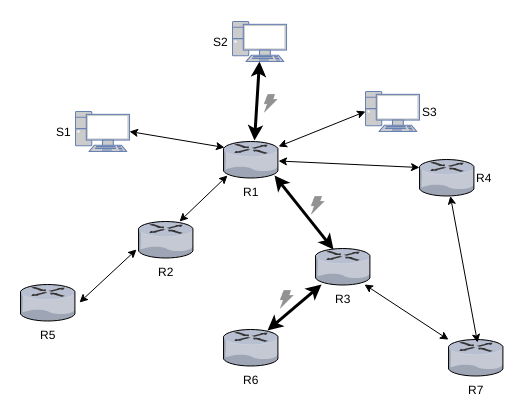
We can see that the packet traverse through the 15 routers to reach to camden.rutgers.edu server. This trace route is taken from a location in the New Jersey State.

# Our Approach

## The router as a point of analysis

From Figure [fig:routers] and [fig:traceroute] we know that routers are located at different geographical locations and also there are specific regions from which a given website or web server is accessed (except few). To know the locations from where the web server/site has been accessed we can use services, called as GeoIP services. These services can detect the geographical location of the system from which the IP packet has originated, but that is just an approximate, based on the source IP and not always correct. In case of DDoS attack, this information is unreliable because the packet source address is often spoofed, hence it is difficult to know the actual geographical location from which the packet has come, but if the router through which that packet has traveled can provide its own geographical information, then such information can be useful to understand the path through which a packet has traveled and thus we can know the region from which the packet was originated.

Also, in the normal scenario there is a consistency in which a website is accessed from different geographical regions, and this consistency can be found by measuring the number of requests or packets traveled from a router to a destination server. This behavior of accessing different websites from a router can be learned over the time. Thus, finding this behavior in a flow[[6]](#footnote-35) at the router can form the basis of analysis in this paper. If there is a deviation from the learned behavior of accessing a particular destination, then that change in behavior as well as router geographical region information can be communicated to the destination network. The destination network on receiving that information, can decide, whether it wants the reporting router to discard or forward the traffic for it. This is a selective process in which traffic from only specific router is blocked while traffic from other routers remain unaffected.



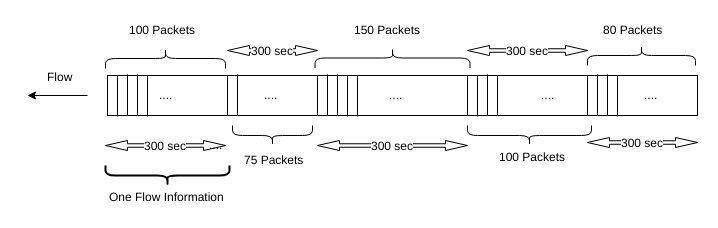
DDoS Attack path

In the above figure an attack is initiated from the region where router R6 is located and from router R6 data packets travel to the victim[[7]](#footnote-37) system S2. Attack packets traveled through router R3 and R1 to reach system S2. If we can detect an attack at router R6, then router R6 can discard all the packets heading towards system S2. In this process, only traffic from router R6 is affected but traffic from all other routers remain unaffected.

With the advance of electronics and the Internet of Things, processing and storage capacity of the electronic devices has increased. Routers are also not left behind, but storage capacity of routers is always very less compared to server which collect flow data for network traffic analysis. If we use learning techniques that don’t need much storage, then we don’t have to store large chunk of packets on the router. Instead of storing data packets for longer time for analysis, we can learn from a small number of packets and then discard packets once learning is done, leaving behind only the learned information on the router. This is necessary because the number of entries on the Internet routing table has steadily grown. Now that the table has passed 500,000 routes, so storing each and every flow information for these routes could be difficult.

## Internet packet flow capturing at the router

To do the analysis, we have first gather the flow information during a time window (e.g. 300 sec) whose size will be fixed at the beginning. Thus, a flow will contain all the packets that traveled from a router to different destinations during a time span of 300 seconds. We can also combine such flows to form a new flow with bigger time windows, e.g. if we combine two flows of 300 seconds then we will have flow of 600 seconds. This flow information can be capture during a particular period or throughout the day. Once we have flow information we can apply learning techniques on each flow iteratively to gain deeper knowledge about normal behavior of the flow, e.g. on average, how many packets of different protocols are destined for a given destination from a given router/region during a particular time of the day.



A 300 second time segment corresponds to a flow of information. The number of packets can vary in a flow.

## Analysis techniques to be implemented

In the proposed system, each router will itself act as an analyzer. Each packet will be analyzed, and a flow statistic will be created based on the destination IP address. Flow statistic will be used to create feature vectors and those feature vectors will be used for clustering based on destination IP address. Clustering is a process of examining the collection of “points,” and grouping the points into “clusters” according to some distance measure. The goal is to minimize the distance of every point in the cluster to other points in the same clusters..

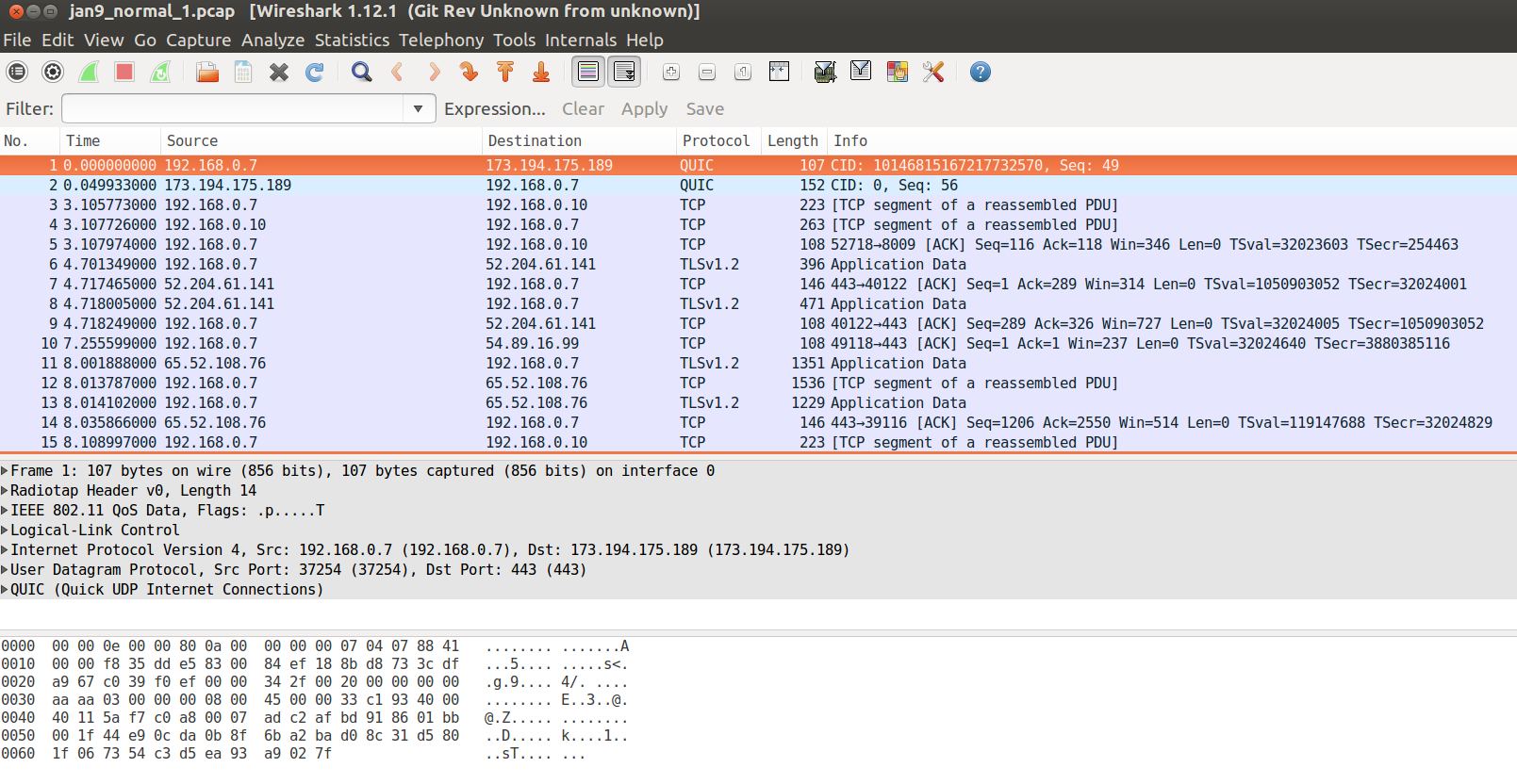
Once the clustering is learned (i.e. we know which destination IP address is tending to be in which cluster), the learned information will be used as a benchmark for all future flows. The routers will constantly keep clustering destination IP addresses and if there is deviation for the normal traffic at a router for any destination then that will affect the clustering and it will cause the destination IP address to be placed in different cluster. This change in the cluster for a given destination IP address can be marked as a change in the behavior of the traffic for that destination. Along with the cluster we will also use the Novelty Detection algorithm to achieve more accurate result. This change in traffic behavior will be reported to destination network, which then decides, buy doing its own analysis, on regulating the traffic coming to itself from the router which has sent the information.

# Implementation

As discussed previously, there are different types of DDoS attacks, such as volume-based, application-based and also low-rate DDoS attacks. Among these different types of DDoS attacks, the volume-based attacks are most common. In the volume-based attack, a victim is flooded with a high volume of Internet packets (TCP, UDP, HTTP or ICMP), which make it unable to serve the requests.

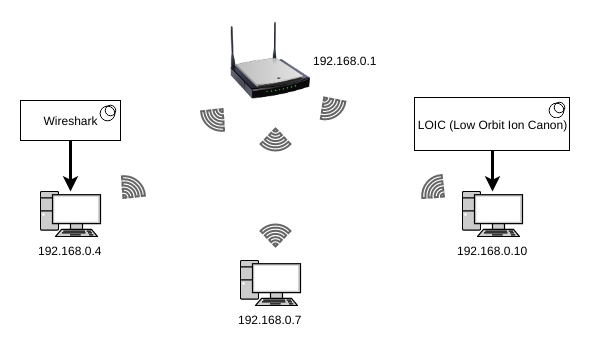
For the demonstration of the suggested approach we are simulating a volume-based bot attack. In the real world, the volume-based attacks are orchestrated using bots. Bots use compromised computer systems, controlled by an attacker for launching an attack. They are not bounded by geographical boundaries, so they can be anywhere in the Internet. Botnet (i.e. network of bots) are employed by an attacker to launch a DDoS attack. As we know that the Internet is connection of different computer system that communicate with each other, through different channels such as cables, satellite or radio device and these communication channels run throughout the glob; connecting different computer systems at different locations. For our simulation, instead of any bot program, we are using Low Orbit Ion Canon(LOIC) tool which is a free DDoS attack launching tool. This tool is even used by attackers in the real world to launch DDoS attacks.

We used Wireshark, an open source tool, for capturing Internet packets. Wireshark can capture all digital information received or sent through different devices such as Ethernet or Wi-Fi devices, which connect computers to the Internet. It also helps identify different protocol packets (e.g. TCP, UDP) within the wrapper packet created at Data Link Layer packet.



Wireshark Tool: snippet of captured packets

To gather data for the demonstration, we have created a small network which has one router and couple of host machine. Each machine can be a victim of a DDoS attack. We have installed Wireshark tool on one of the machine in this network. For capturing traffic in the network, we are using the Promiscuous mode of Wireshark. In Promiscuous mode, a network interface can record not only the traffic that is intended to itself but all the traffic on the network, so we can see all in/out traffic in our setup network. This setup is similar to any router on the Internet which is connected with different routers and hosts.



Network for simulating DDoS detection

To collect packets traveling in our network, we start the Wireshark tool. We let it run for a while and then we orchestrate a DDoS attack on one of the host (e.g. 192.168.0.7 in Figure [fig:demonetwork]) in the network. This attack is engineered using the Low Orbit Ion Canon(LOIC) tool that is installed on one of the host (e.g. 192.168.0.10 in Figure [fig:demonetwork]) in our network. This tool allows us to launch TCP and UDP flood attack on any destination. In this paper we are analyzing the flow-based TCP and UDP attack, which is one of the most common type of DDoS attacks.

In the LOIC tool we need to give the IP address and the port number of the destination where we want to orchestrate attack while rest of the work is done by it. This tool floods destination with packets and if we choose TCP then it will try to create multiple connections and send packets over them. We start flood attack on one of the system (e.g. 192.168.0.7) in demo network and let it continue for few minutes. We launched such attacks few times in between the packet capturing session.

All the traffic, including the normal and the attack traffic, will get captured using Wireshark. Packets are captured for about five hours in the given network. Once the packets are captured they are saved as pcapng file, which is a Wireshark file format for captured packets. Captured packets during the normal operation and during the attacks are saved separately. The normal packet flow information is used for training and testing the learning algorithms (we will explain it in later sections) while attack packets flow information is used for detecting the attack. Wireshark captures every detail of the packets, but we don’t need all of the information, we are interested only in the IP layer information of the packet. Most of the routers analyze IP layer of the packet for routing, having said that, there is no reason that other layers of the packet cannot be analyzed, but for our demonstration purpose we are analyzing only IP packets.

A data extraction program is written in Python language to extract IP layer information from the captured packets. This program extracts address, port and time information from each of the captured IP packets[[8]](#footnote-44).

Sample file snippet (row represents an IP packet)

|  |  |  |  |
| --- | --- | --- | --- |
| Destination IP Address | Protocol | Time stamp(Sec.) | Sample Number |
| 52.6.129.72 | 6 | 1512094785.928596000 | 1 |
| 192.168.0.4 | 6 | 1512094785.946987000 | 1 |
| 192.168.0.4 | 17 | 1512094786.148488000 | 1 |

Our data extraction program also divides the captured data into 300 seconds capture window, thus creating a sample data which is a collection of IP packets captured over the time of 300 seconds. It then writes each sample in the separate file for further processing. This sampling of the packets is the continuous process. We then run another program which extracts the flow information from those sample files. One ‘flow’ contains the number of different packets capture for a given destination. We will store this flow information as sample flow. Then we will train learning algorithms using those sample flows. Once learning is done, those sample files will be discarded, and new samples will be created for further training. This training process has to be continuous process in order to correctly reflect the current status of the flow at given router. Whatever the new information learned, is augmented with the previous learning to have the correct understanding of the flow. This learning can be done for the time during a day or during a week of a year, e.g. We can have separate learning information for flow from morning 9 am to 12 pm and also can have information for evening 6 pm to 12 pm.

Flow based model is built, as it is more reliable and fast. Packet analyzing is often difficult due to the size and encryption. Also, destination port number is not a reliable information in detecting attacks because of the fact that attacker uses different ports during an attack.

Creating a training set for the learning algorithms is an intermediate step in which IP packet count for each destination for a given protocol (e.g. TCP, UDP) is calculated. The training set gives us the flow information for each destination (i.e. how many packets of a particular protocol are recorded for a given destination IP address during a time window, e.g. 300 seconds).

Training Set with three training examples

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  | ICMP | TCP | UDP |
| 172.217.10.134 | 0 | 8 | 12 |
| 65.19.96.252 | 5 | 0 | 192 |
| 68.67.178.134 | 0 | 78 | 0 |

We are using Python program to create training sets, as given in above table [table:feature], from the sample files. Each row in a training set is one training example with destination IP address as label. A training example is vector whose elements are the number of packets observed for a particular protocol for a destination during a fixed time (e.g. 300 second). There are around 150 protocols managed and assigned by the Internet Assigned Numbers Authority (IANA) but most commonly used protocols in the DDoS attack are ICMP, TCP and UDP protocols. For the training and analysis purpose we are using only these three protocols as the desired feature. In the larger system such as routers managed by ISP, other protocols can also be used as features if required.

Each sample file is processed, and corresponding train/test set file is created. Train and test sets files are stored on the file systems, so that those can be used by other programs for training and testing the learning algorithms.

## Machine Learning

According to Tom Mitchell (1998), a computer program is said to learn from an experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience.

A learning algorithm builds a hypothesis using a training set as input. Then that hypothesis is used to perform predictions. The most common categorization of machine learning algorithms is Supervised and Unsupervised.

Let be the function from to which we need to guess from an input vectors also called as ‘input variables’ or ‘feature vectors’. Let be the set of such input vectors. Let be the number of input vectors in a training set . Let be the set of some functions from to . Let be the hypothesis about function . We select based on a training set , of input vectors. In Supervised leaning we know the values of for samples in the training set . We assume that if we can find a hypothesis that closely agrees with for the members of , then this hypothesis will be a good guess for when is large. In Unsupervised learning, we simply have a training set of vectors without function values for them. The problem in this case, typically, is to partition the training set into subsets, , in some appropriate way.

A supervised algorithm such as One Class Support Vector Machine (SVM) is efficient at identifying the anomalies in the data but this algorithm is process and memory intensive, so training the algorithm for each and every IP address is very costly in terms of resources. Because of the resource constraints of the router, our approach is to first cluster the IP addresses based on the features using Unsupervised learning algorithms such as k-means and then apply One Class SVM on the clusters to decide on the boundaries of those clusters. The k-means algorithm is fast and consumes less resource compared to One Class SVM, that makes it good choice to be used on the devices such as routers which have less processing power and less memory.

### Feature Scaling

Before feeding training data, which was acquired in an earlier stage, to a learning algorithm, we have to do feature scaling, also called Standardizing. Feature scaling is done by subtracting the mean and scaling the feature to a unit variance value. It is necessary because different features which are at different scales could cause one feature dominating the others in the algorithm output result, e.g. consider two vectors (1, 2, 3000) (1, 3, 2000). If we calculate the Euclidean distance between these two vectors using the formula , then the distance will be . Form this, it is evident that the larger term is dominating the result.

First, we will convert training examples into vectors. Each vector is one training example and each coordinate in the vector is the feature. To standardize the input vector, the mean and the standard deviation are calculated for the set of input vectors. Then a new vector is created by subtracting the mean from each feature vector and dividing that feature vector by its standard deviation. The new vector created after this step is standardized vector. A set of such standardized vectors is used as input to the learning algorithms.

Standardization formula:

Where is a feature vector, is the mean and is its standard deviation.

### Clustering

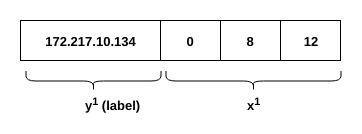
k-means clustering is one of the most efficient algorithms for creating clusters. This algorithm takes any randomly chosen points as centroids (also called, cluster centers) as input from the training set , , .

The following is Lloyd’s algorithm which is most popular heuristic algorithm for k-means clustering. The clustering that we will be doing is of destination IP addresses, such that each cluster will have some number of destination IP addresses.

For this paper we will be using the k-means++ algorithm which is an improvement of k-means, where an arbitrarily initialization step is replaced by simple, randomized seeding technique. If is the shortest distance from a data point to the closest centroid we have already chosen, then the following is the k-means++ algorithm.

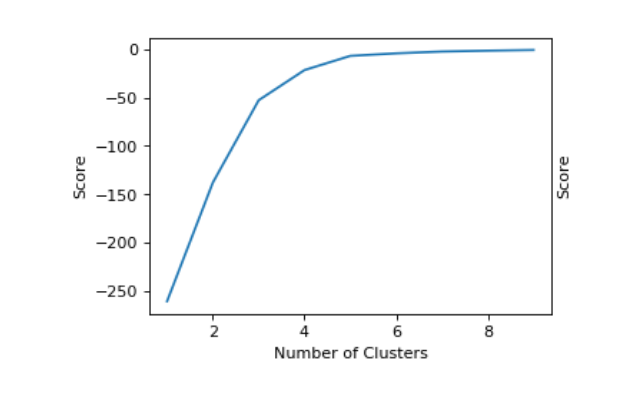
For our DDoS attack detection program, we will be using the Scikit-learn libraries. Scikit-learn is the most popular and rich open-source machine learning software library for the Python programming language and it has implementation for both k-means++ and One Class SVM machine learning algorithms, also it has data preprocessing programs such as feature scaling.

We will take a training set (in the format given in figure [table:feature]) and we will cluster training examples from that training set to using the k-means++ clustering algorithm from Scikit-learn library.



One training example

Before doing the clustering, we have to first determine the number of clusters and centroids.[[9]](#footnote-49) Deciding on the number of clusters is important, because randomly choosing the number of clusters will not have correct clustering. So, we will use the Elbow method to find the optimal number of clusters in our training set. The Elbow method checks the portion of the variance explained by function of the number of clusters. Following is the example of a Elbow diagram.



Elbow Method for cluster count detection

Using the elbow method, the variance for each cluster number is calculated and the cluster number which produces less variance for the next cluster number is selected as best choice for the give training set.

To have correct clustering we first run the k-means++ algorithm to determine the central vectors called centroids. Clusters are represented by these centroids. As we have multiple training sets (also called ‘samples’), we have cluster centroids for each sample. But in the future, we will require only one set of centroids which would be the more accurate clustering for all our training sets. To achieve this, we will save all the centroids, and then find the median of those centroids after removing the outliers, and that will give us a good estimate of centroids. These estimated centroids are used for clustering the samples in the future. Following is the example of centroid vector, where a row represents a cluster and a column represents a feature.

Example of cluster centroids: A row is a cluster and a column is a feature

|  |  |  |  |
| --- | --- | --- | --- |
| Cluster | ICMP | TCP | UDP |
| 0 | -0.16815612 | -0.14928111 | -0.16948046 |
| 1 | -0.18181818 | 5.13527652 | 5.68956244 |

Once we have centroids we can use them to cluster training set again and get the correct clustering based on the centroid.

After clustering is done, it needs to be tested for the accuracy. I.e. how similar the test set clustering is with training set clustering. This can be done by checking, whether an IP address has the same cluster number assigned in both the training and the test sets or it differs. For measuring this similarity, Rand Index(RI) is used. RI is a measure of how many percent does the test clustering matches with the trained model.

where and are the number of true positives, true negatives, false positives and false negatives respectively.

Because of the training sets contain the flow information for different IP addresses on a router during the window of 300 seconds, there is a very high possibility that the same destination IP address is captured in multiple training set. Our goal is to find the correct cluster for the destination IP address, and to achieve this goal, we have labeled IP address with cluster in which it is found most of the time in all our training sets, e.g. If an IP address is assigned cluster ‘0’ in five training sets and cluster ‘1’ in two training sets, then that IP address would be assigned cluster ‘0’ in the result, that we store for future referencing.

Learned information after clustering

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Destination IP Address | Cluster | Packet Count |  |  |
| 74.125.141.106 | 1 | 113 |  |  |
| 72.30.2.182 | 0 | 16 |  |  |
| 64.94.191.14 | 0 | 22 |  |  |

The clustering information tells us the normal behavior of the packets traveling from the router to a given destination. As we have fixed the centroids for the clustering, every time in the future we should expect an IP address to be found in the same cluster mentioned in the clustering information table. But If there is a flooding DDoS attack on a destination, then we can expect to see that destination IP address assigned to a different cluster.

### Anomaly Detection using One Class Support Vector Machine

From the experiments on different data sets, it is found that the destination under attack is labeled with the same cluster number. This happened because there is no other cluster it can be assigned to, so it gets assigned to the cluster whose centroid is nearest.

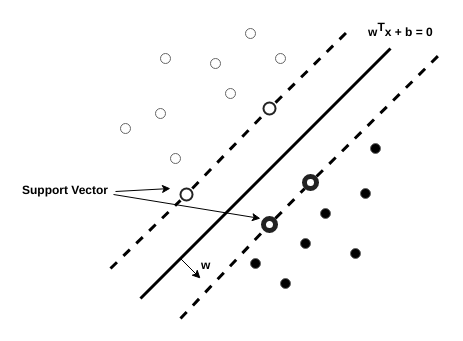
To avoid such a situation, we will have to create boundaries for the clusters, and if any IP address is outside of the cluster boundaries then that can be considered suspicious. This provision will make sure that the attack on the destination IP address will be detected even if the destination is assigned to the same cluster it was assigned in the past.

To create cluster boundaries, we have used One Class SVM classifier. One Class SVM is a type of Support Vector Machine (SVM). SVM is a supervised machine learning algorithm, which means, the training example will have label. I our case, a cluster number will be treated as label for a training example (see table:labeled-set).

#### Support Vector Machine (SVM):

SVM is a supervised learning algorithm which tries to classify training examples into two distinct classes. Classification is based on the labels of the training set. Consider the training set , where is the training example and is a label (or a class) for . SVM will separates those classes using a line or a curve.

SVM tries to create a non-linear separation boundary by projecting data points through a non-linear function (kernel function) to higher dimension space. The data points in space which can’t be separated by a line are projected to the feature space where there can be a hyperplane that separates data points of one class form another. If that separating hyperplane is projected back on the original space then we get a non-linear curve.



Linear Separation using Support Vector

The hyperplane is represented with the equation , where , and . This hyperplane separates the training example labeled with and . The position of the hyperplane is such that the distance from the closest point from each class to the hyperplane is same. To avoid the over-fitting, slack variables are introduced. Over-fitting happens because the learned hypothesis fits training examples so well that it fails to generalize the new examples. The SVM classification is an optimization problem which is stated as follows.

where , , and

is a kernel function and constant is the regularization parameter. If is chosen large, miss-classification of training examples can be avoided. If chosen small, then we may miss-classify few examples, but the margin will be large, so most of the points will be far away from the decision boundary. If this minimization problem is solved using Lagrange multipliers, then the classification function is:

is the Lagrange multipliers. called the support vectors. The function is the kernel and is the intersection.

#### One Class Support Vector Machine:

One Class Support Vector Machine (SVM) is the extension of SVM which detects boundaries of the training set so that every new training example will be classifies as belong to training set or not. It separates all the training set data point from feature space and maximizes the distance of hyperplane from . This creates a binary function which returns +1 for the training example that fits in the trained set region, otherwise it will return -1.

The minimization function of One Class SVM is slightly different than the SVM.

The new parameter introduced in place of in previous SVM equation is used to set upper bound on outliers/anomalies and lower bound on the number of training examples.

Because SVM solves a quadratic programming problem (QP), computing and storage requirements of SVM increases rapidly with the number of training vectors. The approach presented in this paper is more efficient because the number of cluster are limited and always be far less than the number of destination IP address. Training one class SVM on clusters will be far less process and memory intensive than training on massive number of IP addresses, e.g. In our simulation training sets we have 4 four clusters, while the unique number of destination IP addresses in all the training sets is 268, which is more than 60 times the count of clusters we have found.

To classify whether a test example can be a part of a cluster, we need to have classifier for each cluster. I.e. If we have four clusters, we need to train four One Class SVM classifiers, one for each cluster.

# Detection

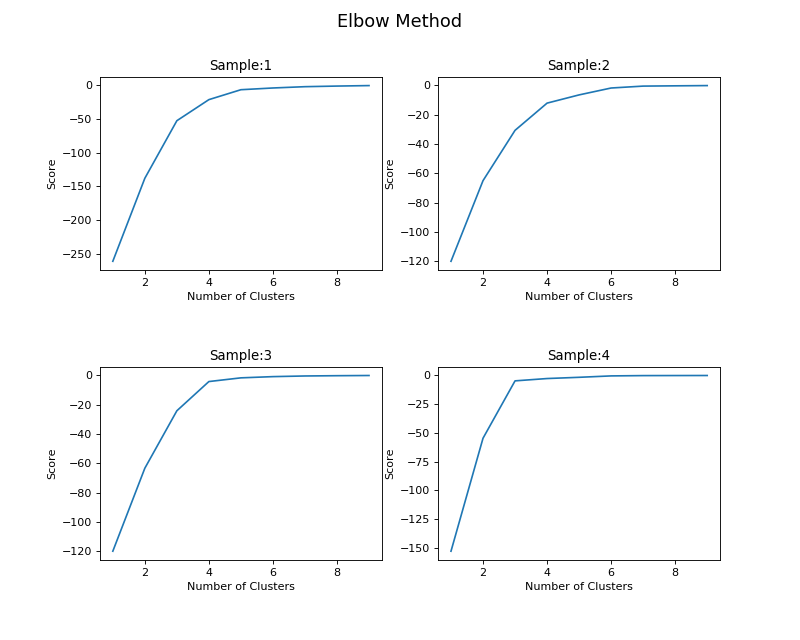
To detect an attack, we have to capture the packet flow at the router during a time interval with which we have created our trained set (e.g. 300 second window). We have to transform the captured flow into a test set using same programs we have used to create training sets from the sample files. The created test set is then clustered using the k-means++ algorithm, using the centroids, that were found after clustering the training sets. We then check the cluster assignment for each example in the test set, and if there is any destination IP address[[10]](#footnote-56) for which the new cluster assignment does not match with the already known cluster for that destination IP address, then that destination IP address is suspected to be under the DDoS attack and so it is added to the suspect list. The suspect list will contain the IP addresses which are probably under DDoS attack because they are not in the cluster they are supposed to.

In another scenario, If the test cluster label for an IP address matches with the trained cluster label, then the feature vector for that destination IP address is passed to the cluster classifier to check if it is inside the boundary of that cluster. As explained in previous section, we will use One Class SVM classifier to decide the boundaries, e.g. if an IP address is labeled with cluster number 0, then we will use One Class SVM classifier that was trained on cluster 0. If the output of the classifier is -1 then that destination IP address is added to the suspect list. Failing to be in the same cluster boundary, that it was in the past, is a sign of significant change in the behavior of the traffic for a given destination IP address on the router where this analysis is done. For every destination from the compiled suspect list, the total number of packets observed in the test set are also recorded. The packet count will act as a filter, to eliminate any misclassification. If there is a significant difference between the number of captured packets in the training and in the test, then that destination IP address will be considered under DDoS attack.

# Detection Simulation

We have created a simulation for detecting DDoS attack using the approach proposed in this paper. As explained in section [sec:Implementation] on page , we have a small network of three computers and a router. We captured the normal as well as orchestrated DDoS attack traffic in the network. We processed the capture data, converted it into samples and use those samples to create training sets. We are using training sets created from the normal traffic for labeling IP addresses with cluster numbers, and then training the classifiers.

We have first used the elbow method explained in section [subsec:Clustering] on page to detect the optimal cluster count. Using that method, we were able to detect 4 clusters in our training sets.



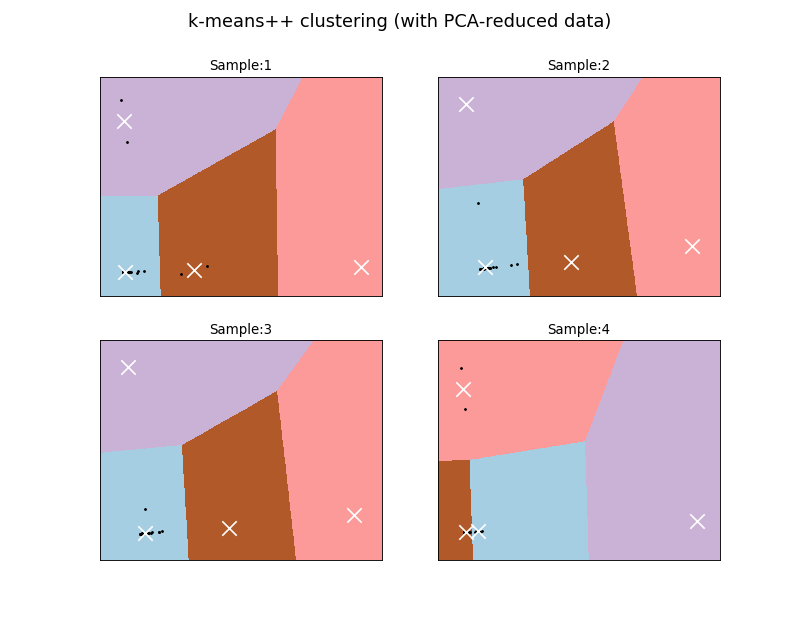
Elbow method to find number of clusters

We then applied k-means++ clustering, as explained in section [subsec:Clustering] on page , to find the correct centroids for our training examples. We found following centroids for our training sets.

Cluster centroids for our training examples

|  |  |  |  |
| --- | --- | --- | --- |
| Cluster | ICMP | TCP | UDP |
| 0 | -0.16815612 | -0.14928111 | -0.16948046 |
| 1 | -0.18181818 | 5.13527652 | 5.68956244 |
| 2 | 5.08663322 | -0.27110845 | -0.099885 |
| 3 | -0.18670401 | -0.18804342 | -0.018538 |

We then used these centroids to cluster our training sets. To draw the clusters in two dimensional space we had to reduce the three- dimensional training example into two-dimensions without losing much information. We have achieved this by using Principal-Component Analysis (PCA). PCA is a technique for taking a dataset consisting of a set of tuples representing points in a high-dimensional space and finding the directions along which the tuples line up best.. We have used the Scikit-learn PCA module for this purpose.



Clustering using k-means++ algorithm

k-means++ algorithm gives label to each training example. This label is the cluster number to which that training example has been assigned to. So, the new labeled data look like below (table [table:labeled-set]).

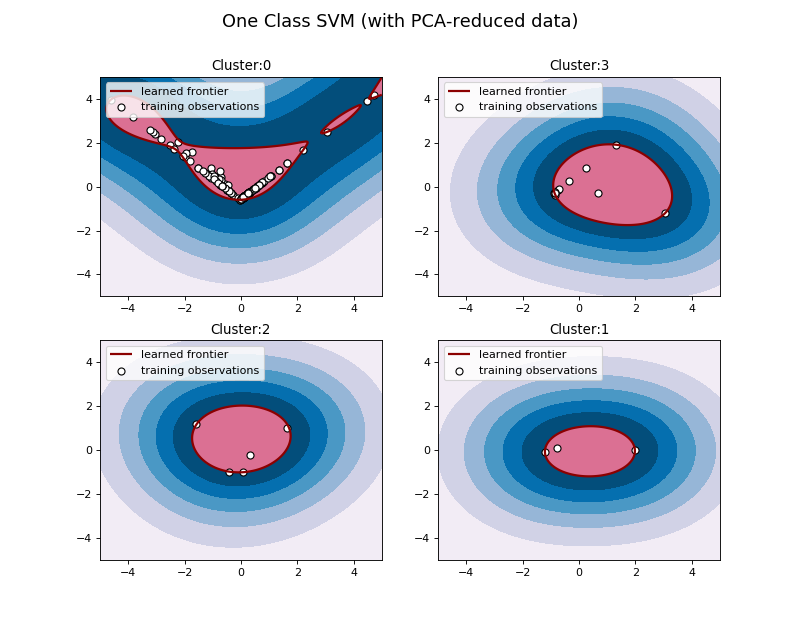
Labeled Training Set (with cluster number)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Destination IP Address | ICMP | TCP | UDP | Cluster |
| 172.217.10.134 | 0 | 8 | 12 | 1 |
| 65.19.96.252 | 5 | 0 | 192 | 0 |
| 68.67.178.134 | 0 | 78 | 0 | 2 |

We also checked the clustering accuracy using the Rand Index explained in the section [subsec:Clustering] on page . We observed that, with more training sets, the RI index improves. A tag file which contains IP addresses, their cluster labels and average packet count is created. This tag file will be used for detecting anomaly in the network traffic.

As explained in the section [subsec:Anomaly\_Detection] on page , we used the training sets, that are labeled with cluster number, to get classifiers. We are using Scikit-learn’s ‘OneClassSVM’ library to train the model and create classifier for each cluster. The input vector to the classifier is the set of all the training examples belonging to the same cluster. Thus, for our four clusters, we have four classifiers. The input vector to the classifier will be of the form shown in table [table:labeled-set] on page .

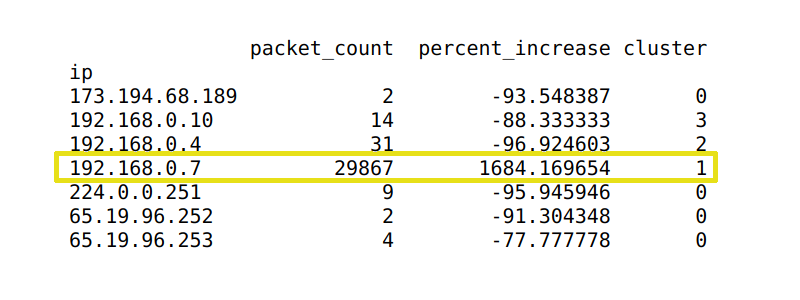
Following is the result of modeling on the training data sets. Each cluster has its own model. We used PCA component analysis to draw the model.



One Class SVM

At the end of the training, we have a file containing the IP addresses labeled with the cluster numbers and four classifiers for four clusters.

We took the test set which was created by capturing packets during the orchestrated DDoS attack and used detection techniques explained in section [sec:Detection] on page . Using the detection techniques on our data, we have successfully detected an attack on the destination IP address 92.168.0.7 in our modeled network.



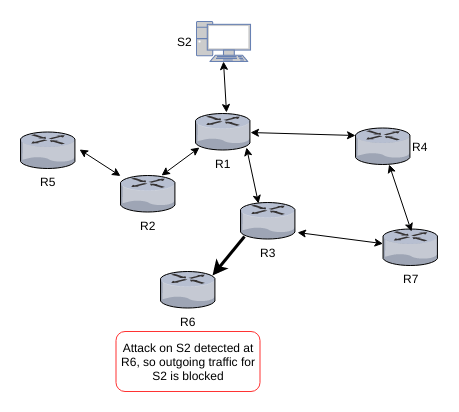
Program output

# Mitigation

By now we have a computer program (consist of few modules, e.g. clustering, classification, detection) which can be run on a router and can be used to detect a DDoS attack. As there are many routers through which a destination can be reached, so to correctly detect DDoS attack we should have this programs on every router in the Internet. One of the way to achieve this is by using Network Function Virtualization (NFV). NFV is an advanced technology which allows us to manage network devices remotely and also allows to add functionality to those networking devices. Because of this virtual technology, we don’t have to manually install computer programs on the routers rather we can do it anytime from anywhere. Although this technology is new, its spreading very fast and soon will be available for every commercial router.

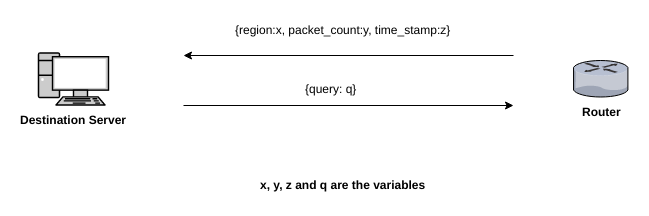
After detecting an attack, a router can choose to block all the packets going to the destination under attack, or it can communicate the attack information to the destination. A destination can be a server or a router. The attack information can contain region name and packet count. Region is the geographical location where the router is situated, and packet count is the average number of packets observed at the router for the destination during attack. Communicating the attack information to the destination can be more helpful because the destination will be aware of the nature of the attack, i.e. for the information provided by router it will know from where attack is originated and how intense the attack is. The destination can also query router for more information such as IP source of the attack packets.

The control of blocking the traffic at the router can also be given to the destination. In this case, destination can perform its own analysis, understand the nature of the attack and then decide whether the incoming packets from the reporting router should be block or not. There can be different parameter based on which the destination server can decide to block the flow and those can be provided by the router as additional information. Many questions, such as, how many routers have reported the change in traffic? Is the attack information coming from the region which never had traffic in the past? can be asked before making the decision at a destination.



Blocking traffic at router: The out-bound traffic for S2 is blocked from router R6, because attack has been detected. While traffic from all other routers is unaffected

To Communicate with a destination network about the attack, router can use the existing ICMP protocol. ICMP protocol is used to provide feedback about the problems in the communication environment. ICMP messages are sent in several situations such as: 1) when a datagram cannot reach its destination. 2) when the gateway does not have the buffering capacity to forward a datagram. 3) when the gateway can direct the host to send traffic on a shorter route. Similarly we can use ICMP protocol to inform destination about anomalies in the traffic. ICMP protocol has many unused type code (there can be 0-255 types but as of now only 0-41 are in use). We can create a new ICMP ‘type’ to send DDoS detection information from the router to the destination server and then destination server can send mitigation instructions to the routers.



Communication between a router and a destination server

# Conclusion

The DDoS attacks are real threat to everyone on the Internet, and its detection is difficult because of the spoofing techniques employed by the attackers. We discussed how we can combine two machine learning methods (clustering and classification) to efficiently detect DDoS attacks at the router. We also discussed that, the program which uses these algorithms can be installed on the router, thus eliminate the need for extra infrastructure for detection of a DDoS attack. With the advancement of Network Function Virtualization (NFV), it is also easy to push the learning programs on the routers. We have also discussed how a router can stop an attack and also how it can communicate the attack information with the stake holders (e.g. server) for better decisions on blocking the attack traffic.

1. Messages that are sent on the Internet are broken into shorter messages for transmission. These short messages are called packets. Term coined by Donald Watts Davies. [↑](#footnote-ref-23)
2. It is a system/device on the Internet that has an IP address, and which is involved in DDoS attack [↑](#footnote-ref-25)
3. spoofing is the way to change the source IP address of the message. This is a known issue in the protocol itself not in the implementation [↑](#footnote-ref-26)
4. System under DDoS attack [↑](#footnote-ref-27)
5. hops are intermediate routers in the communication channel [↑](#footnote-ref-31)
6. In this paper we will consider flow only in the sense of destination addresses [↑](#footnote-ref-35)
7. Victim is a computer system which is under DDoS attack [↑](#footnote-ref-37)
8. packets containing IP information [↑](#footnote-ref-44)
9. Centroid is the vectors which is arithmetic mean of all the point in the cluster [↑](#footnote-ref-49)
10. a destination IP address is a label of a test example [↑](#footnote-ref-56)