1 Introduction

In the year 1988 Morris worm or the Internet worm was distributed via the Internet to ascertain the spread of the Internet and in the year 1999 Kevin Mitnick was tried in the US court for largest computer-related fraud committed in the U.S. history. At that time the loss of intellectual property was estimated at eighty million. Since then, Internet security has now become *numero uno* research area for the networking fraternity.

The first system security laid out to protect the computers was *firewall*. The firewalls were very simple in their approach– allow the incoming traffic by exception. These exceptions were chosen by the network administrators. Firewalls were good as long as the *World Wide Web* was in its infancy. It's humongous success led to the growth of simpler user interfaces for application developments. When the applications became popular firewalls seemed an obsolete tool in handling the incoming network traffic. Their failure sowed the seeds of *signature model* – match with the predetermined pattern, if matched allow it, discard otherwise.

Signature-based intrusion detection remained quite popular for over a decade but then came popular Internet applications like Napster, YouTube, Skype, Facebook, Twitter, WhatsApp, etc. and it seemed if almost everyone was logged into the Internet. High speed communications became need of the hour. It was not only the number of naive users of the net that increased, a new crop of hackers came into being and flourished. As a result the time consuming signature-based intrusion detection were no longer an iron gate in front of the high speed inflow of network packets and protocol vulnerability exploits.

In early 2010 Signature-based Anti-Virus failures were experienced by Stuxnet, Google, Flame and others. In September 2012, online banking sites of 9 major U.S. banks (i.e., Bank of America, Wells Fargo, Citigroup, U.S. Bancorp, Capital One, PNC, Fifth Third Bank, HSBC and BB&T) were continuously targeted by powerful DDoS flooding attacks and resulted into not only huge financial loss but cost them their goodwill as well. The rise of newer attacks like, *zero-day* attacks and *polymorphic malwares* have only worsened the things for Anti-Viruses.

As a consequence *anomaly-based* detection has emerged as a saviour to enforce network security and

solve potentially lethal problems. *Anomaly detection refers to the problem of finding patterns in data that do not conform to expected behavior*. Internet traffic anomalies are generally caused by the vulnerabilities that spruce up in new IP protocols, or due to changes in the Internet routing protocols, or new application vulnerabilites, or malicious activites like worm propagations, etc. In addition to the aforesaid anomalies, these days network attacks like spam, voice IP calls, Denial of Service (DoS), Distributed Denial of Service (DDoS), Flash Crowds (FCs), polymorphic attacks, etc. also account for anomalous traffic in the Internet. Even though there are many tools and techniques that assist in taking a decision, yet, to declare any out of the line activity as an anomaly is completely dependent on the network administrator. *In this thesis, we therefore, have short-listed DDoS attacks as a candidate for anomalies in computer network traffic*.

Abnormal Internet traffic conditions are broadly termed as DDoS attacks. The two necessary conditions for DDoS are – they have to be distributed and force the servers to fail in providing quality of service to its users. The internet infrastructure lacks in providing security against both of these necessary conditions. With the help of *bots* and *spoofed IP* addresses DDoS attacks can be launched in distributed manner and by exploiting the vulnerabilities of Internet Protocols the servers can be inundated with unwanted packets and hence degrading the services. The network administrators and engineers have round the clock job to keep looking for policies, techniques, resources, etc. to find and mitigate these attacks. Research community also tries to assist in tackling these attacks by working on better, fast and accurate algorithms. Conventional DDoS attacks that primarily used to consume the bandwidth of the network to force the denial of the services have now outgrown themselves.

These days instead of making the bandwidth unavailable, specific transport and application layer protocols like TCP, HTTP, DNS etc. are targeted to bring down the services of the Internet. Amongst the many types of DDoS attacks the strongest and hence most damaging ones are the *rate-based DDoS attacks*. Mainly characterized as either constant rate or variable rate DDoS attacks, they have popular ones, namely, Low-rate DoS attacks (LDoS), Pulsating DoS attacks (PDoS), TCP targeted Low-rate DoS attacks (TLDoS) and Reduction of Quality (RoQ) attacks. *Detection of rate based network-anomalies (i.e constant and variable rate DDoS attacks) thus became main aim of our thesis*.

Although a number of tools and techniques have been proposed in the last two decades, it is still very

difficult to detect these attacks due to the large number of *bots* and *spoofed* IP addresses. Detecting DDoS attacks quickly and accurately in network traffic is therefore an important area of study in the current field of research. One particular challenge faced by the researchers in detecting the aforesaid attacks is that the illegitimate has very close resemblance to the legitimate *i.e. flash events or flash crowds*. *Flash crowds* occur when an application server faces sudden surge in the number of incoming legitimate user requests. It therefore is equally important that the algorithm should be able to not only detect various DDoS attacks but also should be able to distinguish between DDoS attacks and flash crowds as well.

Considered as an alternative to the traditional network anomaly detection approaches or a data preprocessing for conventional detection approaches, recently *signal processing techniques* like wavelet transforms , entropy measurements and spectral analysis, due to their abilities in *point change detection* and *data transformation*, have been tested for detecting few cases of DDoS attacks in the network traffic. In the early '90s, studies of network traffic in *Local Area Networks* (LAN) showed that the LAN traffic possessed the properties of *self-similarity* and *long-range dependence* (LRD). *Self-similarity is a scale invariance property. Mathematically, a self-similar pattern is either an exact or an approximate look-alike to a part of itself.* If we could prove that every day Internet traffic exhibits scale-invariant behaviour then techniques based on *self similarity principle* can be used for detection of the attacks. Firstly, in 1988 and later in 2007, Patrice Abry and Darryl Veitch and Will E. Leland et al. in their pioneer works have laid down the seeds for the presence of self-similar nature of computer network traffic. Since then a lot of studying has been done by mainly in queuing modeling and performance measurement. They have argued against the use of conventional Markov based models for traffic engineering and stress on the need to introduce traffic models based on fractional behaviour or self-similarity rather than Markovian models. Their studies were able to establish the fact that all network events were not Poisson and hence the burstiness of traffic does not smoothen out. As a consequence, building Poisson-based network models that could capture the characteristic of burstiness at different scales required many parameters. But due to large amounts of ingress and outgress data at the network hosts, we also needed models that could be worked on *fewer parameters*. The solution lied in using packet aggregation and looking out for a technique that could be worked on fewer parameters, was fast and could measure the burstiness of the network traffic. Hence, scale-invariance with self-similarity as one of it's property became an interesting area to work in. Table 1

shows comparison of some of the works where signal processing techniques have been used for detection of DDoS attacks.

Table 1: Comparison of Anomaly Detection Techniques based on Signal Processing

References	Detection Technique	Attacks Detected
A. Hussain	generate, compare fingerprints using pattern matching techniques & maximum-likelihood classifier	18 DoS attacks
P. Barford et. al.	measure shift in local variance using wavelets	>100 anomalies detected
A. Magnaghi et. al.	Using locality principle measure variations in TCP RTTs and identify misconfigurations	Anomalies with TCP were detected
X. He et. al.	Measure power spectral density, autocorrelation and normalized cumulative spectrum	UDP and TCP based anomalies detected in WAN
G. Bartlett et. al.	used Haar wavelet to do iterated filtering for full decomposition	low-rate periodicities were detected
G. Carl et. al.	Detect change points in the Cumulative Sum of captured traffic	DoS attacks were detected
M. Hamdi et al.	Decompose the state metrics using wavelets. Lipschitz singularities are considered as anomalies	DoS attacks were detected
R. Xunyi et al.	used wavelets to compute wavelet coefficients to find *self similarity*	TCP flooding attacks
W Lu. et. al.	Used 15 features using wavelet analysis, approximate autoregressive and outlier detection techniques	Attacks in the KDD Dataset

From the times when self-similar nature of LAN traffic was explored to the times of self-similar based applications in Internet the field of heavy-tailed distributions has evolved tremendously. However, the shift from measuring the impact, analyzing the origins, to the applications of self-similar has been gradual. Different areas ranging from modeling of aggregated network traffic in the Internet, queuing delays at routers, congestion delays in the links, protocol behaviors like TCP and all look out for self-similar properties. Now-a-days self-similarity is being applied and tested at three different levels in networks; network infrastructure and its protocols being the first level, nature of the data being transferred across intra and inter-continental links as second level and system end-users and end-applications behaviour being the third level. *In this thesis we have explored the usage of self-similarity as a single parameter in distinguishing*

the nature of normal network traffic from anomalous attack traffic.

As is known in the area of network anomaly detection the detection algorithms are applied to aggregated flow-based network traffic. In terms of network engineering, it frees us from time consuming, high cost packet analysis techniques that require reading of individual packet headers of IP protocols for data analysis. The high cost involved in the packet capturing and analysis further aggregates in the presence of sudden changes in the network traffic. Because of the flooding nature of DDoS attacks there is always a sudden increase in the number of packets at the time of attack and therefore the cost of per packet data collection becomes very high. Techniques of self-similarity estimation are of great support here. Because of low recording cost of the traffic intensity at a given frequency, with no need for per packet protocol reading, it is sufficient enough to estimate self-similarity of captured aggregated flow-based traffic. The *self-similar* behaviour of the traffic has the tendency to degrade in the presence of abnormal traffic conditions like DDoS. Variance in self-similarity with its low cost in collection of statistics is therefore a good alternative in measuring and locating the attack points.

The degree of self-similarity can be captured by the *Hurst parameter*. Various statistical estimators can be used to estimate *Hurst* parameter and consequently self-similarity. Such estimation is always based on observing the *second moment* of the network flow-based signal on various time scales. Most commonly used algorithms for the estimation of *Hurst* are – time-domain estimators like Rescaled (R/S) analysis and variance time(VT) plots; frequency- domain estimators like periodogram and wavelets-based estimators. *Wavelets* are a mathematical technique that can be used to observe an arbitrary signal on various time scales. The low computational cost and the scale invariance property makes wavelets an excellent tool for analysis of self-similarity in network traffic. *In this thesis, we apply wavelet-based estimation of Hurst to network traffic to measure variability in degree of self-similarity*.

The aim of this thesis, is therefore, to develop a collaborative methodology worthy of measuring the changes in the self-similar characteristics of the computer network traffic and signal the changes in order to detect the rate-based network-anomalies like DDoS attacks. Rest of the synopsis includes identified knowledge gaps in section 2, thesis objectives in section 3, proposed thesis organisation in section 4, and lastly the conclusions.

2 Identified Knowledge Gap

Based on the surveyed literature we identified following issues worth addressing:

ISSUE 1: In recent years *network anomaly detection* has become the most prevalent area of academic research. Events like network outages, crashes due to router misconfigurations, worm propagations, denial-of-service attacks, flash crowds, protocol-vulnerability exploits, etc. all account for anomalous traffic in the computer network. With time a wide range of anomaly detection algorithms have been proposed. Because of huge financial and good will loses that are incurred by organizations the top-ranking anomalies in network traffic are *DDoS* attacks. Nonetheless, the current *state-of-the-art* attack detection and mitigation techniques fail to provide a collaborative solution for detection and classification of these attacks.

ISSUE 2: Degradation of a network service can be a resultant of intentional DDoS attacks or unintentional increase in traffic volume due to Flash Crowds. Flash Crowds are a created as a result of sudden increase in network traffic due to increased user interest in a particular website for some time. Flash Crowds look-alike DDoS attacks but unlike them are traffic from legitimate users. These days, with the increase in number of Internet-based applications, the number of Flash Crowds is increasing. Any network administrotor may want to keep the Flash Crowds but stop the DDoS attacks. Majority of DDoS attack detection techniques that we reviewed either did not discuss it at all or failed to distinguish the Flash Crowds from DDoS attacks. Identification of legitimate Flash Crowds from illegitimate DDoS attacks was therefore one of our identified issue.

ISSUE 3: While studying the papers related to anomaly detection techniques we noticed that many studies have been conducted to show the presence of self-similarity in the network traffic. Different techniques and tools have been used to measure the degree of self-similarity in network traffic. It has been used to prove the nature of the traffic being self-similar, however, the applicability of the aforesaid property as an anomaly indicator has not been explored deeply. We therefore identified detection of DDoS attacks based on self-similar property as our next issue.

ISSUE 4: Network traffic visualizers enable the administrators in faster comprehension of large quantities of log files data. Rather than scrolling through tens of pages of data on intrusion alerts from a log

file a single graph can be successfully used to summarize the month long data. The current best practices for identifying and diagnosing traffic anomalies consist of visualizing traffic from different perspectives and identifying anomalies from prior experience or profiles. In today's dynamically changing traffic behaviours depending largely on the *profile based or signature based* detection of anomalies leads to failures in missing out the fresh ones. So, to be able to detect new and old type of anomalous patterns we need visualizers specifically designed for *signature free* detection of anomalies in network traffic. We have therefore, from our reviews, realized the need of having a good graphical output visualizer.

While studying the research work we also found out that most of the results are presented for simulations carried out on *NS2* simulator. Validation using real time captured datasets has been done and presented very scarcely. KDD Dataset is publicly available dataset that can be used for validation of detection methodolgy and therefore it was our complementary identified issue that we looked out to resolve.

We now know that occurence of self-similar behaviour and long-range dependency in a small network was established by Willinger and his group in the late 90's. It was a pioneer work but limited to basic application services of the network. Then in early years of two thousand, Barford and his team mates in their work, applied wavelet tool to organize the network traffic data into strata and were able to detect four common anomalies in the local area network. In the later years, concept of self similarity was applied and tested on varied types of application traffic in the internet traffic. More recently, due to the ever-changing face of network attacks and inability of signature-based detection techniques in successfully detecting the upcoming new breeds of attacks, the study & research on anomaly-based detection methods has gained momentum. The techniques based on aggregated network traffic data are therefore a good alternative. DDoS attacks and it's variants are one such set of omnipresent anomalies in the traffic data. *The novelty of our approach is that we have tried to establish that self-similar behaviour is not only present in the deterministic, periodic internet traffic but the anomalous traffic also exhibits the aforesaid behaviour in constant and variable rate, periodic and aperiodic denial-of-service attacks. In doing so, we have tried to establish the differences between traditional DDoS attacks, newly discovered PDoS attacks and Flash-Crowds and detect them.*

3 Thesis Objectives

- To resolve issues 1 and 2, our *first* thesis objective was to work on detection of most prominent anomalies in the network i.e. DDoS attacks and, secondly, distinguish DDoS attacks from Flash Crowds and develop a single detection scheme for characterization.

- To address issue 3, our *second* thesis objective was to explore the applicability of property of self-similarity in characterization of DDoS attacks and Flash Crowds.

- To resolve issue 4, our *third* identified thesis objective was to provide a graphical tool which enables the user in monitoring the regular and anomalous network traffic and in identifying anomalies in network traffic.

4 Thesis Organisation

The thesis has been organized into eight chapters. The necessary details pertaining to fundamental concepts, state-of-the-art research done in the area, proposed scheme, received results, analysis and conclusions have been covered and chapterized as follows:

The first chapter is the introduction chapter and fundamentals of the Internet and its vulnerabilities have been explained along with definitions of anomalies, anomaly detection and how DDoS attacks fit in as network-traffic based anomalies. Brief explanation on various types of DDoS attacks, commonly used tools to automatically generate the attacks as well as detection techniques has also been given. Since self-similarity is locus of our detection methodology therefore the relation between self-similarity and DDoS attacks has also been explained. The chapter concludes with a note on identified research problems, contributions and organization of the thesis.

In Chapter 2, we first provide necessary background on *periodic* and *non-periodic* nature of the DDoS attacks, primarily focussing on current and traditional varieties of DDoS attacks alongwith attack generation tools and subsequently give an overview of DDoS attack taxonomy. The chapter provides *state-of-the-art* research that has been conducted in the field of DDoS attack detection and mitigation. The comprehen-

sive review of supporting literature on self-similar aspects of the computer network traffic and detection of DDoS attacks based on self-similarity estimation techniques is given. In the chapter we highlight the limitations of each of the defense techniques and in accordance to the identified shortcomings, proper justifications have been provided that associate to our research issues. The chapter also draws attention to the role of visualization techniques in detection of anomalous events in the network traffic and presents summary of various visualization tools that are being used by network engineers and administrators for network traffic capturing and monitoring. The work done has been published in survey paper on *Network Anomaly Detection and Role of Wavelets*.

Chapter 3 is a prelude to the later chapters. It explains the concepts of self-similarity, presence of self-similarity in the computer network traffic and key characteristics of self-similar processes. The chapter also presents briefing on *Hurst* estimation to measure self-similarity, various *Hurst* estimation techniques and details of wavelets based estimation of *Hurst* parameter. Descriptions of *fast-pyramid algorithm* and role of *multiresolution analysis* in diagnosis of scale-invariant behaviour has also been provided.

Chapter 4 presents our proposed detection methodology: Multi Scale Network Anomaly Detection (MS-NAD) and pre-test experiments. There are four phases of MS-NAD, namely, filtering of the computer network traffic, computation of wavelet coefficients, *Hurst* estimation and plotting of Multi Resolution Outlier (MRO) map. Filtering of the network traffic is an aggregation process where the values are processed with respect to timestamps and individual data bytes are summed irrespective of the kind of protocol. We have used wavelet coefficients in our self-similarity computations. *Fast-pyramidal algorithm* computes wavelet coefficients using Discrete Wavelet Transform (DWT). The computed values are used to calculate *Hurst* estimates and henceforth a test on self-similarity of the incoming network traffic is done. In the last phase of the detection visual *multi-resolution outlier map* for different scales is plotted.

Two pre-test experiments were done, firstly, pre-test experiment-I for measuring presence of self-similarity in synthetically generated traces of web-traffic using PackMIME and secondly, pre-test experiment-II for measuring the self-similarity on a publicly available dataset. The chapter provides details on synthetic generation of web-based HTTP traffic on *client* and *server* clouds. Recently PackMIME tool has been developed for synthesizing HTTP 1.0 as well as HTTP 1.1 persistent, non-pipelined Full-TCP connec-

Table 2: Comparison of Different Techniques

Reference	N/W Size	Success Rate	Obs. Period	Parameter Count
L. Li et. al. [2005]	NS2, 200 nodes	all 4 attacks detected	10ms-1000ms	5 (IP packet attributes)
X. Luo et. al. [2006]	NS2, Dumbell	all PDoS in mid and high freq. detected	30s	3 (incoming TCP, Outgoing ACK packets)
Z. Xia et. al. [2010]	NS2, 100, 100 nodes	83 % DDoS attacks were detected	3-stage	1 (no. of bytes)
H. Hu et. al. [2010]	NS2, 50 nodes	all simulated LDoS detected	100-250ms	3 (rate, length, time of burst)
Z. Sheng et. al. [2010]	NS2, Dumbell	med & high LDoS detected	-	4 (IP packet attributes)
Q. Zhu et. al. [2011]	NS2, Dumbell	Single LDoS detected	-	4 (TCP ACK, SYN attributes)
Proposed	NS2, 200 nodes, Dumbell	all DDos, all PDoS in mid and high freq. detected	10ms-1000ms,	1 (aggregated byte count)

tions. As described earlier, detection of recently discovered pulsating DoS attacks is one of our focus areas and TCP is the affected protocol, therefore it is critical to understand the impact of PDoS attacks on TCP supported HTTP based web traffic. Self-similar nature of the background network traffic is necessary condition for our framework and we therefore have firstly checked for its presence in PackMIME generated traffic traces. Chapter 4 covers all the details related to that accompanied by discussion on results. Wavelets based fast-pyramid algorithm has been tested for *Hurst* measurements and degree of self-similarity has been observed for aggregated traffic of varying block sizes. *All the generated traces were successfully tested for presence of self-similarity in simulated web-traffic.*

In pre-test experiment-II, work was carried out in order to select an appropriate wavelet for computing self-similarity. We chose the publicly available *Lawrence Berkeley National Laboratory Datasets* (LBNL) to select our wavelet. LBNL traces have been proven to have self-similar nature and therefore are suited for our pre-test experiments. Although there are other tools and techniques available for measuring scale-invariance but wavelets are best suited for the task because the wavelet basis themselves possess a scaling property, and therefore generate a matched 'co-ordinate system' naturally suited to study scaling.

Extensive tests were done with wavelets family of Coiflet, Mexlet, Haar, Symlet and Daubechies. *The computed results for presence of self-similarity in the LBNL data set were found to be in close agreement to results. Results showed that PackMIME simulated traffic was self-similar in nature and Daubechies-6 (Db6) wavelet was best suited for our detection work.* It must be remembered that hurst index measured the degree of self-similarity which has been calculated using wavelets and wavelets are dependent upon vanishing moments (*N*) of the mother wavelet. Therefore, the wavelets with vanishing moments perform better compared to the ones without vanishing moments, and the wavelets with higher vanishing moments perform better amongst family of same wavelets. The increase in number of vanishing moments increases the efficiency of the estimator but comes at a cost of reduction in the number of coefficients at each scale. A tradeoff was therefore required and thus, Db6 was selected for experiments. The work done has been published as *Characteristics Analysis of Web Traffic with Hurst Index*.

In order to develop our detection technique we studied a number of other techniques. Some of them have been tabulated in Table 2. We present comparison with these techniques here due to two important reasons. Firstly, since our work was confined to wavelets based detection of traditional flooding-based DDoS attacks, PDoS attacks and their variants and flash crowds, we choose techniques used for detection of these irregularities only. Secondly, we didn't have active network to capture the real network traffic. In the absence of real network traffic traces we simulated our network using NS2. We referred to the benchmarks- referred by research groups world-wide laid out by renowned J.Mirkovic and her group, in the field of DDoS attacks. Out of those metrics we have done performance evaluation using success rate, detection/observation time and true positive and false negative rates.

In chapter 5 our focus was to study the behaviour of traditional *flooding-based* DDoS attacks, distinguishing them from *flash crowds* as well as *PDoS* attacks. DoS attacks and DDoS attacks have been largely studied through computer network simulations and testbed experiments. Few of them are TCP SYN flooding, SYN/ACK flooding, UDP flooding, IP spoofing, ICMP flooding etc. Now a days we encounter less of network layer based DDoS attacks because of strong network layer defense mechanisms. Application layer-based DDoS attacks like HTTP-based attacks, DNS amplification attacks, VoIP attacks etc combined with Pulsating DoS attacks are the new-breed. These attacks are stealthy and bear high resemblance to

the regular traffic behaviour like flash crowds. The timely detection of these attacks is therefore a primary concern.

Accurately generating the simulation topology, traffic variables, protocols and measuring a DDoS attack is very important to study these attacks. We therefore have used hierarchical graphs *(Transit-Stubs)* of topology generator GT-ITM. There are total of 200 nodes following three level hierarchies: domain, cluster, and nodes. J. Mirkovic and her group has done and published a detailed study on the simulation environments for all kinds of DDoS attacks. We therefore have referred to the works by Mirkovic for setting the network traffic and protocol variables in our simulation environment. Different network traces have been generated and analyzed for DDoS flows ranging from low to high intensity in scenarios consisting of protocols varying from UDP, TCP, HTTP to VoIP, TELNET and FTP over TCP. *All the generated attack traces were successfully detected within one observation time period.* The implemented work was published as *DDoS Detection with Daubechies*.

The second important part of chapter 5 was our second identified issue, i.e. to distinguish between flooding-based DDoS attacks and flash crowds. Flash Crowds are sudden increase in traffic due to spontaneously created curosity in the web users. The sources of flash crowds are spread across Tier-3 networks, send legitimate HTTP Get requests and the flash crowds converge to one point because of the users interest in one web page only. We generated a three-phase topology for Flash Crowds. The shock level parameter has been used to determine the length of each of the phases.

It was observed that Flash Crowds led to highly self-similar network traffic with degree of self-similarity close to one whereas in the case of flooding-based DDoS attacks the degree of self-similarity was widely distributed between 0.5 and 1.0. Secondly, the surge in flash packets was highly concentrated in a short period of time and it could be seen that Flash Crowds were present in the middle order scales of 6 to 10 whereas flooding attacks were evenly distributed in all scales. Therefore, with combined monitoring of *MRO map* and *hurst* values the Flash Crowds could be smoothly distinguished from DDoS attacks. As a subsidiary study we also tested for variabilities in Flash Crowds and TCP variant PDoS attacks and observed that the TCP supported HTTP based Flash Crowds were altogether different from TCP supported FTP and other data carrying protocols. The results were supported by *hurst* values, log-scale

plots and *MRO maps*. The received results and analysis have been published as *A Multi Scale Approach to Distinguish Flash Crowds from PDDoS Attacks*.

In chapter 6, we address the identified issues I, II and IV. The chapter, firstly, provides background on PDoS attacks, different types of PDoS attacks and how they are different from each other followed by the basic congestion control mechanism of TCP and how exploitation of the *TCP's Timeout Phase* can be done by PDoS attacks. The chapter covers in detail *modeling* of PDoS attacks. Aforesaid attacks are primarily dependent upon three variables, namely, length of the attacking pulse, rate of the attack and duration of the attack. The strength and stealthiness of the attacks is largely dependent on the choice of the variables mentioned above. Smaller attack pulses can be equally effective as longer attack pulses if the number of attack bots is large. Similarly, pulses with smaller time period can be stealthier and highly damaging to the TCP servers if the attacking bots are closer to the TCP servers and hence can map with TCP timeout phases more closely. It was therefore required to test all the possible scenarios of

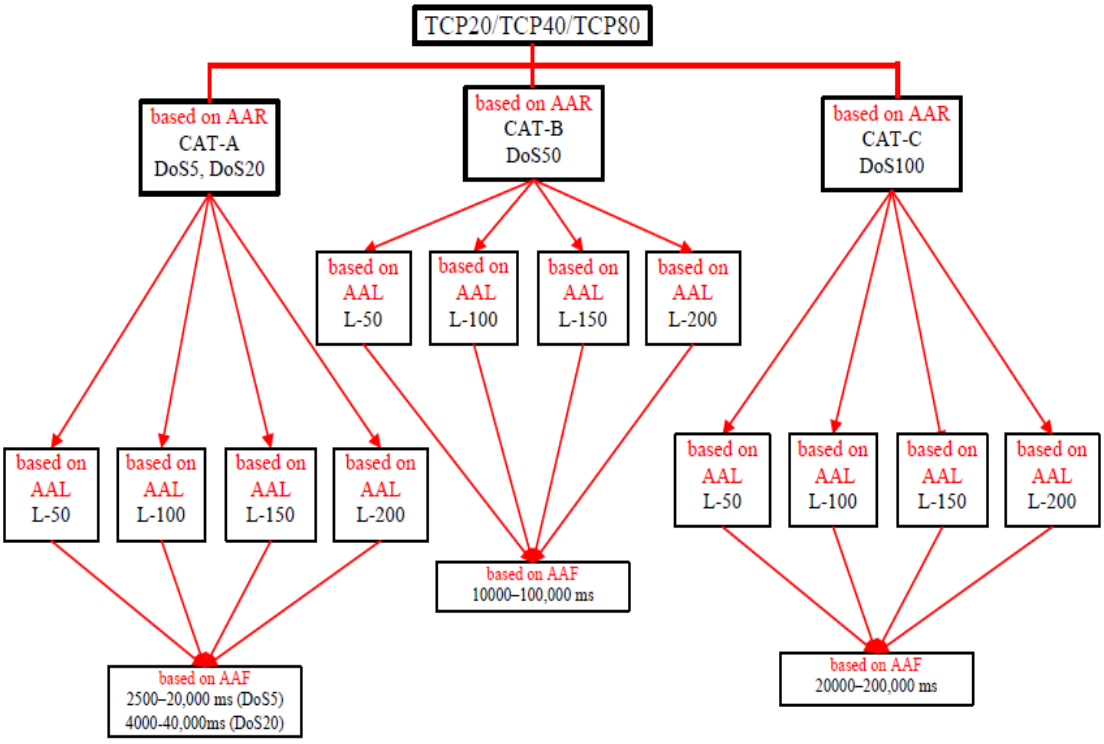

Figure 1: Categorization of Distributed PDoS Attacks

PDoS with our detection framework. Henceforth, to do a detailed study of the impact of PDoS attacks on

legitimate TCP flows we divided them into three categories namely Category A (*DoS5,DoS20*), Category B (*DoS50*) and Category C (*DoS100*) where (*DoSx*) is DoS attack of *x* number of flows. Each category had low(*TCP20*), medium(*TCP40*) and high(*TCP80*) legitimate flows where (*TCPy*) is TCP traffic of *y* number of flows. The pulse lengths were taken as short (*50*)ms, middle (*100,150*) ms and long (*200*) ms and time period was grouped as high frequency (4×10^3)ms to (12×10^3)ms, moderate frequency (16×10^3)ms to (24×10^3)ms and low frequency (28×10^4)ms to (4×10^4)ms. The details of all the scenarios studied in the chapter are shown in Figure 1.

The network traffic trace weres also checked for presence of *self-similar* behaviour using the *logscale plots*. *We observed that whenever the bottleneck links are underutilized before the attack and the attack is not very strong then the hurst deviations are small and cannot be successfully used to detect PDoS attacks. It was also observed that when the attack rate is medium-scale then the deviations in hurst values were noticeable for low, medium and heavy TCP flows. It was also noted that longer the pulses smaller the time required to degrade the service of the TCP server but detection of longer pulses was also easier.* The slopes were steeper for heavy DoS attacks and smaller, stealthier pulses could be used to launch the attack. The link utilization was to the maximum and logscale plots were overlapping for all the blocks. *The experiment results showed that the detection scheme can effectively detect PDoS attacks. The detection delay of one observation period was found to be in accordance with work done by*

The chapter covers visuals generated by the *MRO map* as well. We observed that for the attacks where the deviations in *hurst* values were very small, the *MRO map* visuals were quite effective in locating the start of the attacks. The implemented work got published as *A Novel Multi Scale Approach for Detecting High Bandwidth Aggregates in Network Traffic*.

Chapter 7 presents an evaluation of our algorithm on real traces of KDD Dataset. The KDD Dataset is one of the oldest and most referenced reliably-labelled dataset available in the public domain. Most of the researchers who have worked in the area of intrusion as well as anomaly detection have used KDD Cup 99 dataset for evaluation of their detection algorithms. It has labeled attack samples which were obtained by passive monitoring, rather than by inserting the attack packets into the normal traces. It has been used to validate the algorithms based on entropy computations, fuzzy logic calculations, neural algorithms

Table 3: Comparison of Different Techniques with respect to KDD Dataset

Reference	Technique	Parameter Count	Results	Weakness
W. Lu et. al. [2009]	Wavelets based detection	15	detected DoS	15 parameters, no fixed wavelet
S. Rajasegarar et. al. [2010]	multiclass conic segmentation SVM	NM	99.2 % for 2 DoS	Training & normalization required
Lu et. al. [2010]	conditional random fields	41	99.97 %	Requires prior testing for deciding dynamic weights
J. Yan et. al. [2010]	Weighted Ensemble model	7	94.04 %	Requires prior training
D. Ippoliti et. al. [2010]	growing hierarchical SOM , pattern analysis	NM	75.8-99.7 %	dependent on no. of clusters Requires normalization and offline training
L. F. Lu et. al. [2010]	Wavelets based detection	6	33 %	High false detection rate
X. Wang et. al.[2011]	LBG algorithm	10	98.75 %	Requires data normalization
Z. A. Baig et. al.[2011]	AODE algorithm	NM	99.7 %	Requires learning
X. Xu et. al. [2011]	KPC Analysis and PSO, SVM	NM	96.5 %	results dependent upon weights, high dimension space
R. Vijayasarathy et. al. [2011]	Naive Bayesian classifier	7	98.3-99.5 %	only works for TCP and UDP packets
R. Karbaschian et. al. [2012]	Similarity based alert correlation	NM	98.0484 %	learning based
Ali et. al. [2012]	Resilience Strategy, co-clustering algorithms	15	79.4231 %	large number of input parameters
R. P. Palnaty et. al. [2013]	Jaccard Similarity Coefficient	39	67 %	Depends heavily on protocol and services attributes
S. Novakov et. al. [2013]	PCA, clustering and spectral analysis	14	71-77 %	PCA and Haar wavelet methods failure to pick all the infected bins
A. Aborujilah et. al. [2013]	remove correlated attributes algorithm	NM	84-99.47 %	Training required, only for flooding based DDoS
Z. Tan et. al [2014]	multivariate correlation analysis	32	95.20 %	does not work well in identifying Land, Neptune, and Teardrop attacks
Proposed	Self-Similarity based detection of anomalies	1	99.79 %	

and various other statistical and signal processing techniques. The chapter begins with discussion of some research works where KDD has been used for anomaly detection. The chapter highlights the characteristics of KDD Dataset that has been used for testing and validation of the detection method. The KDD dataset has attacks divided into four different categories with DoS attacks being one such category. There are six types of attacks under DoS namely, *Smurf, Neptune, Teardrop, Ping-of-Death, Back* and *Local Area Network Denial*. *We were able to detect all the six DoS attacks with accuracy of 99.79%. The Hurst values and MRO map have been explored further.* Table 3 provides brief comparison of detection accuracy between various techniques that have been tested on KDD Dataset and our proposed methodology. The work has been published as *A Novel Statistical Technique for Detection of DDoS Attacks in KDD Dataset*.

Chapter 8 concludes the thesis work by summarising the results and proposing future research directions. In this thesis we have presented novel single-parameter anomaly detection techniques using self-similarity. Wavelets with their inherited benefits of scaling were tested and db6 was shortlisted to carry out our work. Both the traditional (flooding-DDoS) and recently discovered (PDDoS) attacks were studied and detected. We were also able to distinguish between flooding-DDoS attacks as well as flash crowds. The methodology was validated using the KDD Dataset for DoS attacks. The scheme showed accuracy of 99.70 %.

In the future work, next step would be to explore the possibility of other input parameters in addition to self-similarity. We also observed that the detection of DDoS attacks based on self-similarity was significantly affected by network congestion. An in-depth study of the effects of rate limiting and queue length on self-similar traffic would be useful to establish a relation between level of congestion and reliability of DDoS detection. Look out for applicability of property of self-similarity for detection of other type of anomalies in the network traffic like worms, outages, port scans, DNS amplification attacks as well can also be done. A study in wireless networks could also be next avenue.