DDOS ATTACK DETECTION STRATEGIES IN CLOUD
A COMPARATIVE STUDY

M. ALYAS¹, MAZHAR IQBAL NOOR², HAMID HASSAN³
Department of Computer Sciences, School of Systems and Technology, University of
Management & Technology
Lahore, Pakistan

ABSTRACT. Cloud is known as highly-available platform that has become most
popular among businesses for all information technology needs. Being widely used
platform, it’s also a hot target to cyber-attacks. Distributed Denial of Services
(DDoS) is a great threat to cloud in which cloud bandwidth, resources and
applications are attacked to cause service unavailability. In a DDoS attack,
multiple botnets attack on victim using spoofed IPs with huge number of requests to
server. Since its discovery in 1980, numerous methods has been proposed for
detection and prevention of network anomalies. This study provides a background
of DDoS attack detection methods in past decade and a survey of some of the latest
proposed strategies to detect DDoS attacks in the cloud, the methods are further
compared for their detection accuracy.

Keywords: DDoS attacks; cloud security; network security; entropy; machine
learning; arima.

1. Introduction & Background: Cloud Computing has become the most convenient and innovative architecture for
businesses globally for information technology (IT) needs. It’s a practice to use remotely hosted servers over the
internet instead of local computer for daily tasks e.g. Data storage, information processing, applications hosting and
usage. Cloud advantages includes but not limited to on-demand self-services, highly available network, location
independent availability, highly scalable, risk-free transfers and pay as you go pricing model [1]. The common cloud
service models are Software as a Service (saas) where a provider provides application services to its consumers,
Platform as a Service (paas) enable its users to use cloud platform for hosting their applications and Infrastructure as
a Service (iaas) gets basic bare computing resources and utilize according to needs [2]. The cloud offers private,
community, public and hybrid deployment models [2].

Cloud is known for highly available infrastructure and is widely used but it is also a hot target for attackers
and hackers. There has been a lot of research in past decade on cloud security for the measures to protect cloud data
and resources and fight against attackers. There are several threats that cloud including denial-of-service (DoS), data
breach, unauthorized access, insecure APIs, vulnerable applications, account hacking, malicious content, loss of data,
and abuse of services [3]. Each threat to cloud either data theft/loss or services unavailability may cause severe loss to
the businesses. Denial of Service is a common threat to cloud services, during a DoS attack, the legitimate users are
denied to access and use cloud network resources [4] and it has been a network research trend since 1980 [4].
Distributed DoS (DDoS) is a DoS in which multiple systems called botnets or zombies are used to form an attack on
cloud network target [5]. The first DDoS attack was reported in 1999 by Computer Incident Advisory Capability
(CIAC) [4]. A DDoS attack can target either a network or an application, two methods are used for DDoS attacks: (i)
the malformed packets are sent to the target to baffle network protocol or a service running on it for vulnerability (ii)
to upset legit users’ connectivity to cloud network by overwhelming network capacity/bandwidth or by depleting
server resources e.g. CPU, Memory, I/O etc. by flooding it with requests [4]. The DDoS attacks are rising lately, and
are reported to be over a terabyte per second in 2016 [6].

DDoS attacks are usually launched by botnets that continuously and concurrently flood the target with huge
number of requests to bring it down and exhaust its resources. A botnet is a network of machines/devices/computers
(often compromised) widely dispersed and controlled by attackers to form the distributed attacks, which are often
done by automated applications [4,7]. A bot can find vulnerabilities in entire network by scanning it, and can install binaries in compromised machines which stays active to execute future instructions commanded by bot-master [8] and each new compromised machine becomes part of botnet, that’s how bots network keeps growing. In a DDoS attack multiple IP addresses, sometime thousands, are used to originate the cyber-attack [6] which also makes it difficult to mitigate. The symptoms of a DoS attack are: poor network performance, dead slow web site/application response, unavailability of cloud service, sudden rise in bandwidth consumption, and sudden increase in server resources utilization, network connection instability, malformed response, and denial of service requests. Figure 1 illustrates the legitimate users and attackers interacting to cloud network, where they send requests to establish connection to cloud servers.

Some common DDoS attack techniques are application-layer floods, degradation-of-service, HTTP Post, Internet Control Message Protocol (ICMP) floods, peer-to-peer, spoofed, amplification, R-U-Dead-Yet (RUDY), Slow Read, and SYN floods [6]. Application-layer attack, flood the requests to application and ultimately exhausts the resources. Degradation-of-service doesn’t flood but makes overall service response slower. Peer-to-peer attack exploits the bugs in peer-to-peer servers, while ICMP attack involves very large number of IP packets sent to target with faked source. Spoofed attack consists of sending engineered packets to victim for a reply, and amplification attack is to magnify the bandwidth to overwhelm the target. RUDY attack makes web sessions starve at server. Slow read involves in sending legit requests but it reads responses very slow that ultimately keeps the victim busy, and finally, upon SYN attack, the target receive huge number of SYN/TCP packets from fake senders. All these techniques ultimately cause the DoS to legitimate users.

To fight against DDoS attacks, different solutions and systems are used at application and network level. Firewalls (FW) and Intrusion detection system (IDS) are used to detect the attacks which are based on different set of algorithms and standards defined [10], and Intrusion prevention systems (IPS) are used to prevent such attacks.

Since the discovery of DDoS attacks, it has been a hot research topic and numerous methods has been proposed to reduce and prevent the attacks. ALPi method introduced by P.E. Ayres et al. [22] which used packet scoring concept that increased DDoS attack detection accuracy. Secure Overlay Services (SOS) [23], a DDoS prevention system was introduced by A.D. Keromytis et al. that claims to decrease attack chances and is consisted of three features: secure overlay tunneling, filtering and routing [11]. Another method WRAPS [24] was proposed by X. Wang, followed the concept of web graph to fight against DDoS attack where it requires legit users to authenticate using trusted referral link. Dou [22] introduced confidence-based filtering (CBF) method to detect attacks in cloud, CBF extracts the features of packets during non-attack period and use packet-scoring mechanism to decide whether to accept or reject packets during attack. A fast and elegant DDoS attack detection policy SNMP-MIB [26] was introduced by J. Yu et al. [26] which use support-vector machine (SVM) classifier for classification between legit and malicious traffic. Cloud Trace Back (CTB) was presented by K. Lee et al. [27] that finds the source of attacks and use clustering technique to classify the attacks. Wang et al. [28] proposed a multi-stage framework LR-DDoS to detect DDoS attacks that uses multiple features in packet headers to set network state and joint deviation rate (JDR) for anomaly detection in network flow. C. Jie-Hao et al. [29] proposed ANN for attacks detection and compared the results with decision-trees, entropy and bayesian classifiers, Carl et al. [30] proposed activity-profile, change-point and wavelet-based signal analysis for attack detection and proved the possibility to successfully detect the DDoS floods. G. Thatte et al. [31] introduced parametric approach that used statistical analysis that relies on headers of IP packet but it’s difficult to detect application-layer DDoS attacks with this method. P. Sangkatsanee et al. [32] used machine learning method involved C4.5 decision trees and demonstrated successful detection of DDoS attacks. K-nearest neighbor (KNN) classifier was used by Y. Liao et al. [33] to classify program behavior as normal or malformed, this was proved to be excellent method for DDoS attack detection but have high computational complexity in case of larger number of instances. Chaos-theory based system was proposed by A. Chonka et al. [34] to differentiate legit and malicious traffic.
and introduced neural network based model to detect DDoS attacks [12]. Generalized entropy (GE) method was used by I. Basicevic et al. [35] for HR-DDoS attack detection and compared results with Shannon entropy based model, and claimed better detection and low false positives. The above cited work highlights the methods proposed in/around last decade while this study aims to discuss in detail some of the lately introduced variant strategic DDoS attack detection systems in the cloud.

2. DdosAttack Detection Strategies: This section provides a detailed insight about the latest proposed methods to detect ddosattacks.

2.1. Entropy based Models: Entropy based models, to detect network traffic anomalies, analyze the changes in flow to determine if system is under attack or not.

2.1.1. Xinlei Ma et al. [18] presented a system that run chaos analysis on traffic entropy to conclude about system state by using Tsallis Entropy and Lyapunov exponent [43], as DDoS attacks on network will change behavior of traffic flow and structure of the packets and will be more spread when sender addresses are faked and entropy can catch this behavioral change [18]. In [36], GE and information distance specs are used to determine the DDoS attack, but additionally, for rate of separation among source & destination IPs, [18] proposed Lyapunov exponent variation for the entropies to detect attacks. Lyapunov exponent is measure that characterize the rate of separation among two paths [18].

$$e^{\Delta \tilde{X}_t} \approx \Delta \tilde{X}_0$$

EQ. 1 -Lyapunov exponent equation [18].

The exponent separation consists of five steps: (i) collect network packets in real-time (ii) calculate entropies of source & destination IP (iii) pre-process entropy-time using AR model (iv) analyze rate of exponent separation (v) set classification threshold for DDoS attack detection [18]. The experiments conducted on widely researched MIT [37] DDoS dataset by simulating exponent separation detection five steps process [18]. Multiple tests were conducted with variant attack intervals ranging from 5 seconds to 100 seconds and each test confirmed 98.56% true-positive ratio with 100% sensitivity and specificity rate [18].

2.1.2. Another entropy based model presented by Sunny Behal [12] use information theory metrics $\varphi$-Entropy & $\varphi$-Divergence to detect Fast Events (FE) and DDoS attacks in cloud, it assumes that all attack sources works in coordination and have similar logic toward victim. As compared to GE and Generalized Information Divergence (GID) the [12] model is more sensitive to detect variations and can better distinguish between legit and malformed traffic. It use generalized $\varphi$-Entropy & $\varphi$-Divergence to sense the traffic state and is better in anomaly detection when compared to GE & GID [12], and it can detect FE with 100% accuracy [12] and is independent of any tool that makes it better choice for future [12]. The attack is detected by analyzing size of time-window and packet header features. The algorithm follows: (i) set sampling parameters (ii) keeping analyzing traffic (iii) eliciting features from packet headers (iv) calculate probabilities (v) compute $\varphi$-Entropy & $\varphi$-Divergence (vi) compute information distance and (vii) apply logic to distinguish between normal and malformed traffic [12]. The method was heavily tested using MIT [37], CAIDA [38], FIFA and self-created datasets, and concluded that algorithm can accurately detect malicious traffic from normal traffic.

2.2. Learning based Models: The learning based models use knowledge to classify the incoming traffic as legit or an attack, the learning can be either supervised or unsupervised. In supervised learning the system is first trained then tested against similar structured unseen data, while unsupervised learning involves partitioning the data based on dissimilar attributes. This sub-section discuss few latest learning based algorithms proposed to detect DDoS attacks in cloud.

2.2.1. A classifier system for DDoS (CS_DDoS) to detect TCP Flood DDoS attacks in cloud was proposed by A. Sahi et al. [11], which use least-square support vector machine (LS-SVM) as a classifier [11]. The CS_DDoS system is consisted of detection & prevention modules. It assumes that sender IPs are not faked, The detection module is responsible to detect the anomaly by following the steps: (i) gather the network packets for a certain time period (ii) cross-check the collected packets for sender IP blacklist (iii) if found in blacklist, then prevent them without further investigation (iv) if not found in blacklist, then sent to classifier to further decide on its state (v) a packet is considered as malformed if sender IPs tries to connect to same destination frequently and is prevented (vi) if packet is considered legit, then it’s forwarded to destination to establish connection [11]. The experiments were done using different learning models including LS-SVM, Naive Bayes, K-nearest and Multiplayer perceptron but LS-SVM found the best in performance and accuracy. The CS_DDoS performance experiments were conducted using K-fold cross-validation model where the dataset was split into six equal sets, five sets were used for training and sixth to test the classification. The results shows that the model using LS-SVM as classifier was able to achieve 97% accuracy, sensitivity and specificity for DDoS attack detection [11].

Table 1 lists the average results obtained by [11] using different classifiers in CS_DDoS.
2.2.2. Another learning based detection system, Ensemble-based Multi-filter Feature Selection (EMFFS) method was proposed by O. Osanaiye et al. [13] that use four different feature selection algorithms and combine their result to attain ideal selection [14].

Table 1 – Average performance results of classifiers [11].

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Accuracy</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Kappa coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>LS-SVM</td>
<td>97%</td>
<td>97%</td>
<td>97%</td>
<td>0.8875</td>
</tr>
<tr>
<td>Naïve Bayes</td>
<td>88%</td>
<td>94%</td>
<td>94%</td>
<td>0.765</td>
</tr>
<tr>
<td>K-Nearest</td>
<td>81%</td>
<td>96%</td>
<td>95%</td>
<td>0.7275</td>
</tr>
<tr>
<td>Multiplayer perceptron</td>
<td>93%</td>
<td>98%</td>
<td>97%</td>
<td>0.69</td>
</tr>
</tbody>
</table>

Figure 2 illustrates the four feature selection methods used: Information Gain (IG), Gain Ratio, Chi-squared and ReliefF. IG is filter feature selection method that can determine the related attributes from a features set and it reduces the uncertainty for class attribute selection [13]. Gain ratio improves the weight of IG for diverse valued features, it gives higher value when data is spread almost evenly but lowest when all data belongs to one class [13]. Chi-squared (X2) method is used to measure the separation of two variables by score computation, the higher the score the higher the relationship dependency and vice versa [13]. ReliefF feature selection method is heuristic independent, noise tolerant and have lower computational complexity [39]. The process consists of: (i) for all four feature method rank and sort the features (ii) combine resultant features (iii) select common features from all methods up to the set threshold and ensemble [13]. The EMFFS method was validated using NSL-KDD [40] dataset & decision trees and system was able to detect DDoS attacks 99.56% accurately [13].

2.2.3. Peng Xiao et al. [16] presented another learning based DDoS attack detection model CKNN based on correlation analysis of network traffic and K-nearest neighbor algorithm [16]. It can detect traffic anomalies with high efficiency and lower cost [16]. It used r-polling model to reduce calculation of training data by using correlation information of training data that reduce overhead of density of training data [16]. In correlation analysis, flow is captured and analyzed by same application that share the similar parameters i.e. dest_ip, dest_port and protocol. For example an attack to web server tries to target http port 80. The proposed model CKNN use grid approach to reduce overheads and spatial index is used for training data partition [16]. The process consists of: (i) r-polling method to reduce overheads (ii) CKNN algorithm is used to classify the test data that scans the input data, apply r-model and kNN. The experiments were conducted using WEKA [41] API and Hadoop cluster with Hadoop and web services. Multiple tests were conducted with different K value ranging from 2 to 100 [16], also CKNN and KNN were compared for performance and accuracy, CKNN clearly is better choice for DDoS attack detection [16]. The tests shows that CKNN system was able to gain 96.3% accuracy and is better choice over KNN based models [16]. Figure. 3 illustrate the performance and accuracy comparison between CKNN and KNN models.
2.2.4. Another approach to detect DDoS attacks in real-time using Artificial Neural Network (ANN) was presented by Alan Saied et al. [5] that can detect TCP, UDP and ICMP attacks. The training pattern includes source IP address, ID and a sequence of source & destination ports. The motto assumes that attackers usually use their own programmed system to generate the malicious network packets rather than using operating system mechanism that makes it easier to distinguish between genuine and malformed traffic [5]. The solution continuously monitors the network traffic and analyze packet headers using trained ANN classifier to determine the network state i.e. normal or under attack. The detection process involves: (i) the DDoS detection modules are installed on networks (ii) when a detector finds a malicious packet, it informs its neighboring detectors about attack (iii) detectors keeps monitoring their networks (iv) if number of malicious packets exceeds the set threshold, then an organizer identifies the target IP and information is sent to ANN engine to determine if it’s attack or normal [5]. The experiment process included training the system with old and new datasets first and then test the real-time traffic using the trained system, the results shows that ANN based system can correctly classify 98% traffic in new datasets [5].

2.3. Time-Series based / Other Models
This sub-section enlighten two DDoS detection approaches other than learning and entropy methods.

2.3.1. SeyyedMeysam et al. [14] proposed a model that use Autoregressive Integrated Moving Average (ARIMA) [42] time series and chaotic system to detect DDoS attacks on a network. It use two traffic features; number of packets and number of source IPs at per minute interval, and form a time-series according to number of packets. It use ARIMA model to forecast the packets in future minutes. The abnormal behavior is predicted by calculating the maximum Lyapunov exponent. The proposed model TNA [14] requires number of packets and sender IPs as input, and classification process consists of (i) Features selection (ii) pre-processing data (iii) predictions using ARIMA (iv) Analyzing chaotic errors (v) classification of chaotic and non-chaotic (vi) and detection of system state i.e. normal or under attack. The experiments were conducted using R environment and DARPA 1998 [44] dataset is used for training and test purposes. The results shows that this novel system can classify 99.5% network traffic accurately in order to determine the DDoS attacks in real-time.

2.3.2. Another model Network Anomaly Detection Algorithm (NADA) was proposed by Yonghong Chen et al. [20] that use Autoregressive (AR) model and chaos theory to detect network anomalies. AR model is used to predict network traffic, system assume that prediction errors have chaotic behavior, thus chaos theory is used to detect abnormalities and a Neural Network (NN) is trained from abnormal traffic to classify DDoS attacks [20]. Because of its simplicity and lower complexity, AR model is preferred over other ARIMA and ARMA models for prediction [20]. The DDoS attack detection consists of: (i) real-time collection of network traffic (ii) preprocess and make predictions using AR model (iii) apply chaos theory to analyze the prediction error rate and abnormalities in traffic (iv) train neural network with malicious traffic to find the network state [20]. The experiments were conducted on DARPA 1998 [44] & DARPA 1999 [45] and determined that system was able to correctly classify 93.75% traffic, but assume that with more training data it can gain more efficiency [20].

<table>
<thead>
<tr>
<th>Ref#</th>
<th>Algorithm</th>
<th>Detection Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1.1</td>
<td>Xinlei Ma et al. [18]</td>
<td>98.56</td>
</tr>
<tr>
<td>2.1.2</td>
<td>Sunny Behal et al. [12]</td>
<td>φ-Entropy 94% φ-Divergence 100%</td>
</tr>
<tr>
<td>2.2.1</td>
<td>AqeelSahi et al. [11]</td>
<td>97%</td>
</tr>
<tr>
<td>2.2.2</td>
<td>O. Osanaiye et al. [13]</td>
<td>99.56%</td>
</tr>
<tr>
<td>2.2.3</td>
<td>Peng Xiao et al. [16]</td>
<td>96.3%</td>
</tr>
<tr>
<td>2.2.4</td>
<td>Alan Saied et al. [5]</td>
<td>98%</td>
</tr>
<tr>
<td>2.3.1</td>
<td>SeyyedMeysam et al. [14]</td>
<td>99.5%</td>
</tr>
<tr>
<td>2.3.2</td>
<td>Yonghong Chen et al. [20]</td>
<td>93.75%</td>
</tr>
</tbody>
</table>
Results Comparison: This section provides summary about the results achieved by different DoS attack detection methods discussed in last section for their detection rate accuracy. Table 2 lists the detection accuracy rate determined by attack detection algorithms discussed in this study. The results shows that O. Osanaiye et al. [13] has achieved the highest detection rate 99.56% by utilizing the learning based model and ensemble-based feature selection methods, which is the result of features selected by four different algorithms. Another closer model proposed by SeyyedMeysam et al. [14] which detects the 99.5% attacks accurately by using time-series model to analyze the cloud traffic. Xinlei Ma et al. [18] achieved 98.56% accuracy to malicious traffic by analyzing the traffic behavior using entropy-based model. These results shows that each type of model i.e. entropy, learning and time-series is capable to detect DoS attacks in the cloud with similar detection rate.

Conclusions & Future Work: Cloud as a highly-available and scalable platform, which is being used for different services including SaaS, PaaS and IaaS. Being popular solution for IT needs, it’s also a hot target for cyber threats including DoS, in which the attackers target a network/server in cloud with huge number of requests from different faked sources i.e. botnets to cause the denial of service. During a DoS attack, the network performance becomes very poor or completely out of service. Since businesses relies on services provided by cloud, so a DoS attack affects the businesses badly, thus the measures be taken to detect and mitigate such attacks. In this study some latest DoS attack detection has been comparatively discussed for their methodologies, algorithms and accuracy. The different models discussed are entropy-based, learning-based and time-series based. Almost all of them relies on network packets features set, extracts features (e.g. source IP, destination IP, source/destination ports, number of packets per interval etc.) and run a test to determine health of traffic flow. The comparison chart Table 2 shows that accuracy rate of all of them is over 90% that makes them quite useful. The learning-based models (section 2.2) reflects higher accuracy rate but could require memory to store the training features set. The two methods with highest accuracy are learning-based 2.2.2 and time-series based 2.3.1 that provides accuracy of 99.56% and 99.5% respectively.

Future work may involve a survey of DoS attack detection methods in software-defined network (SDN) as well as review of DoS attack mitigation methods along detection. The simulation of these algorithms implementation to obtain benchmarks can also be considered.

REFERENCES


