IJCSE International Journal of Computer Sciences and Engineering Open Access

Research Paper Vol.-6, Issue-5, May 2018 E-ISSN: 2347-2693

Behaviour Analysis of DDoS Attack and Its Detection

Mahadev[1*], Vinod Kumar[2]

[1*]Dept. of Computer Science, Gurukula Kangri Vishwavidyalaya, Haridwar, India
[2]Dept. of Computer Science, Gurukula Kangri Vishwavidyalaya, Haridwar, India

[*]Corresponding Author: mahadev.agra@gmail.com, Tel.: +91-9760626562

Available online at: www.ijcseonline.org

Accepted: 19/May/2018, Published: 31/May/2018

Abstract— In recent times the internet is growing exponentially. Many important services and records are available on different websites of the government and as well as of private sectors. A valid user becomes irritated when websites become unavailable when needed. Human being accesses only those web pages in which they are interested in. Some flash crowd occurs on specific time or events. Attacker's main aim is not to browse specific web pages of valid users' interest but to fail the web server so that authentic users could not avail web services. The DDoS attack becomes difficult to detect when this attack imitates the behaviour of irritating and non-professional users. There is need to analyze the behaviour of sophisticated DDoS attacks using advanced tools of DDoS attack at Layer 7. This paper analyzes tool of DDoS attacks using their log records and checks behaviour of DDoS attacks and stores its pattern in ODAM (One Dimensional Access Matrix). It also proposes an efficient algorithm to detect DDoS attacks at the application layer.

Index Terms—DDOS (Distributed Denial of Service Attack), ODAM (One Dimensional Access Matrix), Layer 7, flash crowd, application layer

I. INTRODUCTION

Attacks of distributed denial of service (DDoS) raise an ever greater risk to the internet with increasing resources at the hands of attackers. Financial institutions of different types and sizes come into contact with DDoS attacks that put these institutes into a variety of risks which includes financial loss as well as brand image loss [1]. If the attacks come with fraud, a financial institute may face huge loss and may lose the faith of customers. DDoS techniques are actually used to stop working of the website by transferring heavy loads of messages or requests to servers and stop its services from running. As a result, attacked websites couldn't perform accordingly with respect to a standard set by web developers and owner of these websites. DoS attacks have widened their scope from one website to multiple websites at a time. DDoS is just a DoS attack executed in a distributed manner where attacks come from several compromised computers.

'Black Friday' or 'Cyber Monday' is a good example of the DDoS attack through which impact of DDoS attack can be understood well. Millions of internet users got affected by this incident where people bought gifts through a website but had error message without showing web page they wanted to see. A high amount of traffic to the server which couldn't be possible to handle for the server was the main reason for this incident. Two-way communication was compulsory to deal with the website, but services were not able to run because all the resources were consumed by the high amount of requests. Without common web user's interaction, the communication takes place between laptop and particular website. There are many ways to DDoS attacks on a particular website with marginally different objectives, but the main objective of this attack is to make the website unavailable to web users. In DDoS attack, a wicked actor takes control of multiple devices which are connected to internet and are used to attack. These malicious actors are perceived as attack vectors which are then used to decrease the performance of server of victim's website. This characteristic of DDoS attack becomes reasonable to choose for attacking site and make the important services (like an email application, registration, reservation, banking application, search engine etc.) of sites impossible to access.

DDoS attacks required the processing power of multiple computers for a successful attack on particular website. A master controller starts attack through handlers that install an attacking program on other compromised computers (around hundreds) called agents which start a DDoS attack after getting signal from the master controller via handlers. These compromised computers (agent, handlers) corrupted with the malicious program are remotely located and activated to start an attack without any consideration of the owner of computer (agents and handlers) that his computer is being used by DDoS attackers. An infected system with an installed DDoS attack is known as zombies and collection of zombies connected to agents and handler is known as botnet who launch attack after getting signal of the master controller.

In studying new classes of attacks on internet, we consider protected system that has defenses against both 1) attacks, i.e., attacks by intruder that exploit software vulnerabilities such as buffer overflows and 2) protocol attacks, i.e., attacks that exploit protocol inconsistencies to render servers unavailable[2] (e.g., take over DNS entries or modify routing). In such a situation, attackers may avoid detection by being non-intrusive and according to protocol, and yet disturb the system resources while posing as authenticate clients of the application. Hence, those system attributes and files stored on network and server available for the attacker to misuse are those available for the application workload. The App-DDoS attack comes into different form that makes it difficult to identify.

Application layer attacks are classified into three groups based on the nature of its attack on application layer [3]:

Lower Rate request Attacks

Each compromised computer sends lower rate requests at a higher rate of sessions than normal users and the rate of the session may vary randomly.

Lower Rate Session Attacks

Each illegally agreed computer sends higher rate requests at a lower rate of the session than normal sessions and the rate may change randomly to avoid detection.

Resource depletion Attacks

Each request specifically designed to consume resources at server side in a lower rate of requests and session attack.

In this paper, section I contains introduction of DDoS attack at the application layer, its classification and impact on internet users. Section II contains present work on this problem. Section III explains proposed work that solves the detection problem. Section IV describes experiments and results and Section V concludes research work with future direction.

II. Related Work

A partial solution is proposed for the problem of DDoS attack as trust management helmet (TMH) [4]. This scheme uses trust strategy to differentiate legitimate users from malicious requests and give priority to those nice users whose system is not used as an agent of DDoS attackers. This scheme protects rights of good users to avail service while the server is under a DDoS attack. Trust to users is calculated based on the history of users' browsing. The license of trust is given to users after verification from log records in the server to maintain the continuity in service. The cryptographically protected license is used assuming that this will continue the services of the client during an attack. The simulation result shows the effectiveness of TMH in mitigation of HTTP flooding attack. More than 99% of requests are accepted with TMH coming from authorized users and less than 18% of requests are accepted without this trust. This solution doesn't provide any suggestion when licensed and trusted users' system becomes agent or a node of a botnet after getting trusted license, the situation may be worse if licensed users start taking part in DDoS attack as a part of a botnet.

A clustering method [5] is used to analyze the app-DDoS attacks using users' browsing behaviour. Four parameters of the session as sessions' mean of the number of requested objects, average popularity of all web objects in the session, request rate and the average probability of transition of web pages are extracted. A browsing behaviour model is built taking a high amount of real sequence of normal users and validates this model with attack dataset. Dataset used in this method doesn't contain any DDoS attack. So the possibility of identification of DDoS attack from this real dataset is very low. The DDoS attack doesn't use transition of web pages but it directly sends a large quantity of requests for one particular page without transition.

A matrix for HTTP request transition [6] is proposed to understand browsing behaviour of users. A pattern transition probability from page to page is also considered assuming that legitimate users browse only interesting web pages and objects. But bot behaves differently from legitimate behaviour. It doesn't know the popularity of web pages and sends requests in a random manner with a sequence of pages that have very less transition probability with respect to legitimate sequences. If advanced bots are used that try to simulate behaviour of legitimate users, the likelihood of bots' behaviour is still much high. Likelihood interval can be used to recognize these bots.

A type I error (recognized as a false positive) and type II error (known as a false negative) will differ based on what threshold value is being taken. The complexity of attack strategies decided threshold and other parameters' value. If false positive and false negative are high, it is impossible to identify the legitimate user and DDoS attacks correctly.

The threshold value is determined by three parameters: frequency vector as objects' popularity, transition probability matrix as the probability of transition one page to another page, host request sequence probability as average probability of transition probability of the request sequence.

Smart DDoS attackers do not send a sequence of pages in the present scenario where the different type of puzzles are available to test the presence of human being and tools of DDoS attack that could use a sequence of pages, to the best our knowledge, is still not available.

A technique Discrete-time Markov Chains (DTMC) [7] is proposed to detect an App-DDoS attack through an anomaly in the user behaviour. This paper analyzes SQL queries made by users after that DTMC model is applied to make a comparison between legitimate users' behaviour and abnormal access with anomalies.

III. PROPOSED MODEL

1. PROBLEM SCENARIO:

A. *Study of Behaviour of AppDDoS Attacks*

There is certainly a difference between request behaviour sent under DDoS attacks and those sent by legitimate users. A large number of bots participate in a DDoS attack at the application layer. If the DDoS attack of a low rate per bot is sent to victim there will be a large volume of requests to the targeted server side. There is a need for detection of a DDoS attack and recognize the behaviour of attack through recognizing characteristics per bot. Detection of DDoS attack has become very complex in present scenario with sophisticated attack tools. So, reducing false positive and false negative rate is a big issue.

B. *Assumptions:*

i. Attackers use services of botnet to attack application layer.

ii. Modern attackers use original IP address to receive acknowledgement.

iii. Server becomes incapable to serve the authentic user requests during DDoS attack.

iv. Attackers' tool directly access web pages mimicking that referred page is from other websites.

2. Detection Approach: methodology

A. *Dataset preprocessing*

Analysis of DDoS attack required attack data as well as data of normal user when users access web pages. A new website is designed for the collection of these types of data, for log data of a DDoS attacked file is not available. Caida dataset and clarknet dataset have only real-world log data of legitimate access. It is necessary to understand the structure of website before understanding the behaviour of DDoS attack pattern. Zambie toolkit is used to launch DDoS attack installed on Kali Linux that uses TCP connection flood, HTTP attack which is amplification attack type. HTTP attack included tools like slowloris, tors hammer, slowhttptest with rudy in that tool. The structure of a website is considered to understand the behaviour of web access. One characteristic of the log record is byte size. A web developer should design each web page of a website of different size so that each web page refers to a different size of a web page. It could be useful in pattern recognition and analysis of each request that comes to the website. Normal log records are collected through several systems having DDoS attack tool installed on LAN. The graph displays actual behaviour of legitimate user as displayed in Figure.1. Attacked data is also collected using zambie toolkit to analyze the behaviour of DDoS attack which is shown in Figure 2

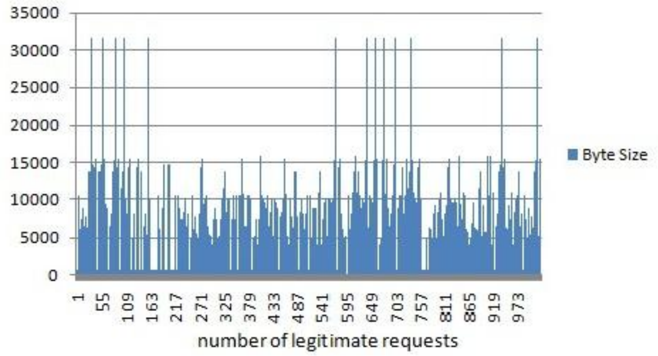


Figure 1: Legitimate user's access behaviour w.r.t byte size

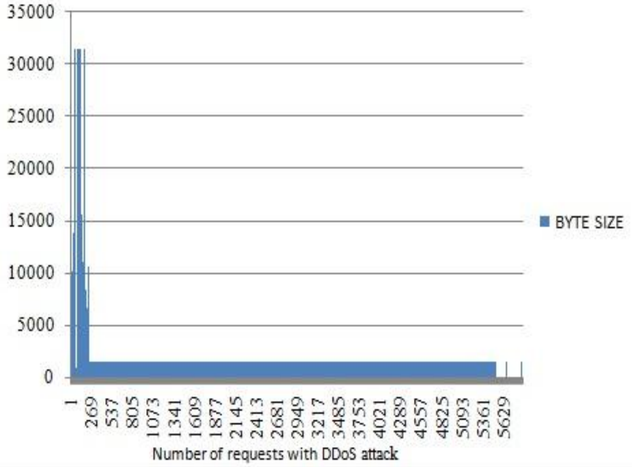


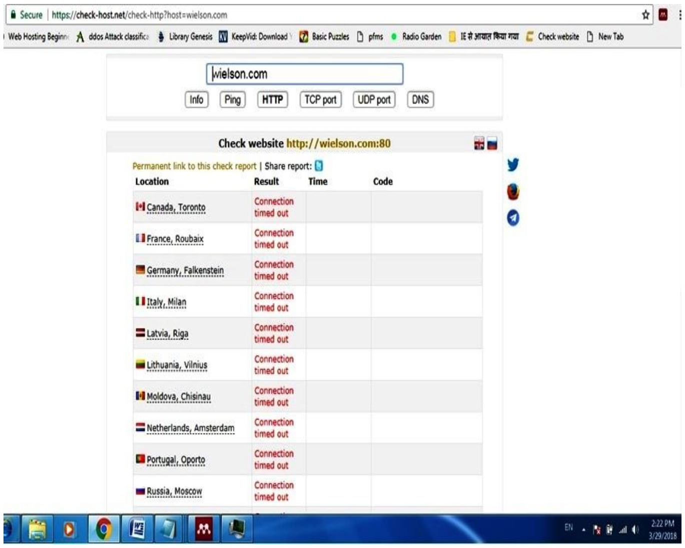
Figure 2: DDoS attack behaviour w.r.t byte size

Log records are collected from a web server of website wielson.com. For actual analysis of the dataset, there will be only small preprocessing in dataset because, during the time of attack, agent directly accesses the objects (jpg file, webpage, audio file, video file, high workload file) and repeat this action continuously in a small fraction of time until services go down.

B. *Description of log Record:*

Fields are separated according to values in a log file and stored in ascending order on the field of IP address. Time is defined in a table of a database of MySQL.

The log records are collected on different dates. Log records of normal data on 29th Nov 2018, 20th March 2018, 23rd March 2018 and 4th April 2018 are collected, while log data of attacks are collected on 30th Oct 2017, 20th Feb 2018, 27th March, 29th March 2018. The DDoS attack is performed on my website that is created with a simple structure using HTML and CSS. Figure 3 shows that DDoS attacks make the website unavailable all over the world as depicted by website check-host.net that site checks the performance of websites for the different protocol.



Check website http://wielson.com:80

Location	Result	Time	Code
Canada, Toronto	Connection timed out		
France, Roubaix	Connection timed out		
Germany, Falkenstein	Connection timed out		
Italy, Milan	Connection timed out		
Latvia, Riga	Connection timed out		
Lithuania, Vilnius	Connection timed out		
Moldova, Chisinau	Connection timed out		
Netherlands, Amsterdam	Connection timed out		
Portugal, Oporto	Connection timed out		
Russia, Moscow	Connection timed out		

Figure 3: Unavailability of the website due to DDoS attack

Preprocessing of log data is performed using java program and results are stored in MySQL database. Apache 2 server is installed on the space provided by AWS Server. Description of log records stored on Apache 2 server is as follows:

14.139.238.206 - - [29/Mar/2018:06:25:06 +0000] "GET /clothes/clothes.html HTTP/1.1" 200 806 "http://wielson.com/" "Mozilla/5.0 (Windows NT 6.1) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/65.0.3325.181 Safari/537.36"

Definition, as given by Apache organization for access log records in Apache server, is given below:

- IP address: 14.139.238.206
- TIME: [day/month/year:hour:minute:second zone]: [29/Mar/2018:06:25:06 +0000]
- Method: GET/POST
- Requested page with protocol: /clothes/clothes.html HTTP/1.1
- Status: 200
- Byte send: 806
- Referer: http://wielson.com/
- User-agent: Mozilla/5.0 (Windows NT 6.1) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/65.0.3325.181 Safari/537.36

C. *Feature extraction:*

Some features are identified for analysis of log records that contains normal as well as attacked data. The threshold value is also set using these features as described below.

i. Frequency vector at first access.
ii. First Access Probability of historical data
iii. First Access Probability of current data per session

Frequency Vector: $p_{0i} = \frac{r_{0i}}{\sum_{i=0}^{n} r_{0i}}$ i=1 …..*n*, ……(1)

p_{0i} is the probability to access ith page from 0th page of log record of the web server to be stored in ODAM. In p_{0i}, 0 means that referer ∉ W (set of the webpage of the website) because the referred page may be a web page of another web. r_{0i} is the number of requested web page stored in ODAM that directly comes from URL or through some another web site. In r_{0i}, 0 means that referred page is from any website but does not belong to this paper's experimental website wielson.com that is referer ∉ W (set of the webpage of the website) because the referred page may be the web page of another website. Table1 shows the normal behaviour of the legitimate user when a website is less popular.

Table 2 shows the behaviour in log records while DDoS attacks occur. Zombie of botnet requests specific web page repeatedly so the value of r becomes very high abnormally that try to imitate behaviour of flash crowds when we see in just log file without analysis.

Table 1: Normal user behaviour (duration 1 min)

r	1	2	. . . . . .	n
0	300	90	. . . . . .	75

Table 2: Behaviour while DDoS attack occurs (duration 1 min)

r	1	2	. . . . . .	n
0	4000	30	. . . . . .	20
0	2000	3200		15
:	:	:	:	:
0	2500	50	. . . . . .	20

D. *DDoS attack Detection Technique:*

Basically, Euclidean distance is used to recognize similarity and dissimilarity among the values of two variables. Here two vectors of probability values of web page access of a website in two scenarios are given for detection DDoS attack.

For each session of each IP address:

Euclidean distance $\sqrt{\sum_{1}^{n}(x_{0j} - y_{0j})^2}$ > threshold ………(2)

x_{0j} is the probability of requested page in historical log data metrics, y_{0j} is the probability of requested page in current log data metrics, *j*=1…*n*, *n* is the total number of direct access pages per thousand requests. Threshold value depends on the average of the maximum Euclidean distance of normal user and minimum Euclidean distance of DDoS attack using testing data of both types of requests.

Threshold =

$$mean\left(max_{1}^{q}\left(\sqrt{\textstyle\sum_{1}^{n}(x_{0j} - y_{0j})^2}\right), min_{1}^{s}\left(\sqrt{\textstyle\sum_{1}^{n}(x_{0j} - y_{0j})^2}\right)\right)$$

Where *q* represents series of requests of normal user and *s* represents series that contain DDoS attacks of historical data.

If the Euclidean distance is greater than the threshold value, higher are the chances of a website being under DDoS attack. IP addresses that take part in the DDoS attack may be a zombie which should be blacklisted by network devices and server. Mitigation process should be started after recognizing the DDoS attack. This detection process minimizes false positive rate in recognizing real culprit of the DDoS attacker.

Proposed Algorithm for DDoS Detection:
Step 1: Input log records of victim's access server.
Step 2: Partition of each log entry using the regular expression.
Step 3: Make parametric calculations for feature extraction.
Step 4: Remove those rows if the referred column contains the name of the website.
Step 5: Calculate frequency vector of the first access.
Step 6: Calculate Probability x per web page using historical data and the threshold value.
Step 7: Set threshold value using Euclidean distance formula using test data.
Step 8: Calculate probability y of the requested page per session.
Step 9: Calculate the Euclidean distance
Step 10: If Euclidean distance > threshold value, it results in the higher possibility of DDoS attack.
Step 11: End.

IV. Experiments and Results

The proposed algorithm successfully detected that given log records are infected with the DDoS attack. The complete log records are separated into series of thousand records. Only those log records are collected that have direct access to web pages. For this purpose, only those records are deleted that contain website URL in the referer column. This is because DDoS attack tools access web page directly [8]. The first row of Table 3 represents the Euclidean distance between probabilities of the directly accessed web page of historical log records (x_{0j}) and log records of normal and attack behaviour (y_{0j}). Training log data of a normal user is used to decide x_{0j} of the particular web page. The threshold value is decided by historical data of a normal user and DDoS attack. This pattern recognition method clearly shows the DDoS attack in given log records with the recognizable difference [9]. The second row shows the number of requests for direct access to the web page in each one thousand log records. This method also differentiates between a flash crowd and attack behaviour. The seventh column in test data of normal user shows its clear difference with threshold value 0.42477. Figure 4 and Figure 5 show the difference between legitimate behaviour and attack behaviour. Y-axis represents the square of the difference of probabilities of accessing the web page (di) and X-axis represents web pages of websites through numbers in the dataset.

Table 3: Pattern recognition for different series of normal user and DDoS attack

	training data of normal user					test data of normal user		data of DDoS attack for threshold calculation		test data of DDoS attack
a.	0.1661	0.1281	0.1612	0.0817	0.1049	0.1095	0.0958	0.6834	0.8565	0.6971
b.	61	51	62	132	99	859	172	811	466	1000

a. Indicate Euclidean distance $\sqrt{\sum_{1}^{n}(x_{0j} - y_{0j})^2}$ for different series with respect to historical data.
b. Indicate the total number of requests for direct access to webpage per thousand log records.

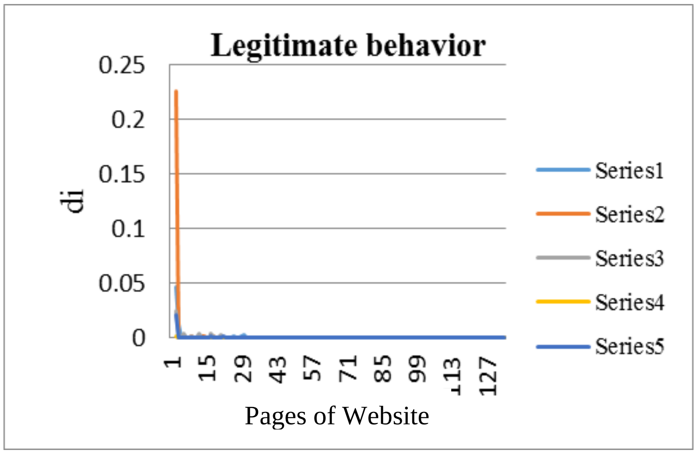


Figure 4: Legitimate behaviour

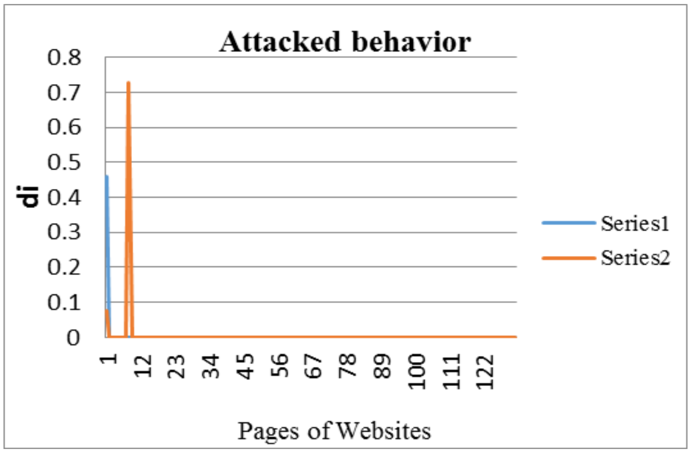


Figure 5: Attacked behaviour

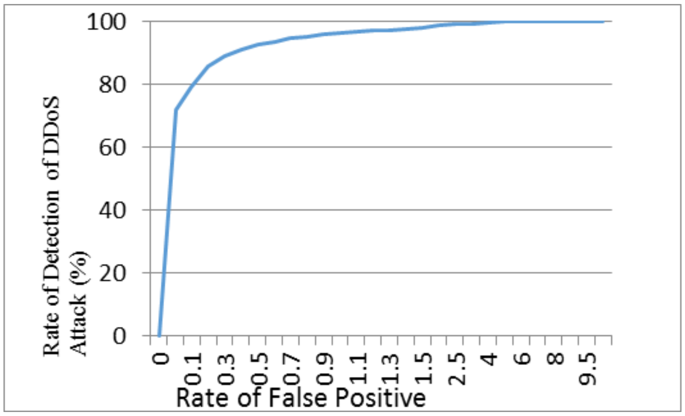


Figure 6: ROC curve for false positive rate

ROC (Receiver operative characteristics) depicting the performance of DDoS attack detecting algorithm on application layer is shown in Figure 6.

V. CONCLUSION AND FUTURE WORK

The proposed algorithm is based on pattern recognition method Euclidean distance using ODAM can detect DDoS attack with high accuracy. Dataset related to DDoS attack is recorded through the website using attack tool. Data of normal user is also recorded on different dates. Advanced online tool for website monitoring for checking the availability of website during the DDoS attack is used during the experiment. This algorithm is capable to differentiate flash crowd from DDoS attack.

The proposed technique reduces the false positive rate and the false negative rate at a significant level. IP addresses of bots, taking parts in the DDoS attack, may be identified. This method can be useful to detect a DDoS attack in real traffic data. Mitigation of DDoS can be achieved by understanding the characteristics of this attack.

REFERENCES

[1] R. Kroszner and J. Munn, "*Federal Financial Institutions Examination Council*" *Distrib. denial Serv.*, no. 703, pp. 2–4, 2008.

[2] S. Ranjan, R. Swaminathan, M. Uysal, A. Nucci, and E. Knightly, *"DDoS-shield: DDoS-resilient scheduling to counter application layer attacks" IEEE/ACM Trans. Netw.*, vol. 17, no. 1, pp. 26–39, 2009.

[3] J. Yu, Z. Li, H. Chen, and X. Chen, *"A detection and offense mechanism to defend against application layer DDoS attacks,"* *3rd Int. Conf. Netw. Serv. 2007*, 2007.

[4] J. Yu, C. Fang, L. Lu, and Z. Li, *"Mitigating application layer distributed denial of service attacks via effective trust management," IET Commun.*, vol. 4, no. 16, pp. 1952–1962, 2010.

[5] C. Ye, K. Zheng, and C. She, *"Application layer DDoS detection using clustering analysis" Proc. 2012 2nd Int. Conf. Comput. Sci. Netw. Technol.*, pp. 1038–1041, 2012.

[6] C. Ye and K. Zheng, *"Detection of application layer distributed denial of service," Proc. 2011 Int. Conf. Comput. Sci. Netw. Technol.*, pp. 310–314, 2011.

[7] B. Meng, W. Andi, X. Jian, *"DDOS Attack Detection System Based on Analysis of Users' Behaviours for Application Layer",* Computational Science and Engineering (CSE) and Embedded and Ubiquitous Computing (EUC), 2017 IEEE International Conference on, pp. 596-599, July 2017.

[8] M. M. Najafabadi, T. M. Khoshgoftaar, C. Calvert, C. Kemp ,*"User Behaviour Anomaly Detection for Application Layer DDoS Attacks"* , IEEE International Conference on Information Reuse and Integration (IRI), pp. 154-161, 2017.

[9] Mahadev, V. Kumar, K. Kumar,*"Classification of DDoS Attack Tool and Its Handling Techniques and Strategy at Application Layer"*, IEEE International Conference on Advances in Computing Communication and Automation (ICACCA), Oct 2016.

Authors Profile

Mr. Mahadev received MCA degree in 2001 and qualified UGC Net exam in 2012. His area of specialization is in web technology and system analysis and design. He has working experience of 7 years as asst. professor in Invertis University, Bareilly. He is currently pursuing Ph.D. from Gurukula Kangri Vishwavidyalaya, Haridwar under the guidance of Professor Vinod Kumar getting junior research fellowship and senior research fellowship from UGC.

Dr. Vinod Kumar received Ph.D. degree in 1987. His area of specialization is in Distributed Computing, Network Security, and Reliability Engineering. He has total working experience of 19 years as a professor in Department of Computer Science, Gurukula Kangri Vishwavidyalaya, Haridwar and Computer Science Dept of Vidya College of Engineering. He is currently working as dean and registrar of Gurukula Kangri Viswavidyalaya, Haridwar. He has approximately 38 years of research experience, published 83 research papers in journal and completed 2 research projects. 14 research scholars completed their Ph.D. under his guidance. He has been a member of 8 educational bodies including IEEE USA, ACM USA, senior life member of Computer Society of India etc.