Volume 8, No.1, January – February 2019

International Journal of Advanced Trends in Computer Science and Engineering

Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse12812019.pdf https://doi.org/10.30534/ijatcse/2019/12812019



Detection of HTTP Flooding DDoS Attack using Hadoop with MapReduce : A Survey

Ziyad R. Al Ashhab¹, Mohammed Anbar², Manmeet Mahinderjit Singh³, Kamal Alieyan⁴, Wajih I. Abu Ghazaleh⁵

¹National Advanced IPv6 Center (NAv6), University Sains Malaysia 11800 USM, Penang, Malaysia ziyadashhab@yahoo.com

²National Advanced IPv6 Center (NAv6), University Sains Malaysia 11800 USM, Penang, Malaysia anbar@nav6.usm.my

³School of Computer Sciences, University Sains Malaysia 11800 USM, Penang, Malaysia manmeet@usm.my

⁴ National Advanced IPv6 Center (NAv6), University Sains Malaysia 11800 USM, Penang, Malaysia Kamal_alian@nav6.usm.my

⁵ Servers Management Manager, Deanship of Information Technology, King Abdulaziz University,

Jeddah, KSA

wajiha@kau.edu.sa

ABSTRACT

These DDoS (Distributed Denial of Service) attacks have affected large cloud environments and are a huge threat for worldwide organizations. These attacks flood the target network with a large number of packets because of which network becomes incapable of providing the services to its legitimate users. The horrible destruction of DDoS attacks can be seen from the very first attack in 1999 to the recently publicized attack Ababil. Attackers find new ways to launch attacks irrespective of the standard DDoS defense mechanism. The existing DDoS detection technologies or methods requires to be improved in order to effectively deal with such attacks in reasonable response time. This survey paper emphasizes on DDoS attack mainly HTTP GET flooding attack and a detection technique that is based on MapReduce.

Key words : Cloud computing, DDoS (Denial of Service), flooding attack, HTTP attack, Big data, Hadoop, MapReduce.

1. INTRODUCTION

An appropriate platform is offered to the user through which they can access the various applications and resources as a service by cloud computing. The cloud computing technology uses utility-based computing that means enormous distributive data centers are used to store the large data. Cloud computing provides three types of services: software, platform, and infrastructure as a service and it can be used as private, public, hybrid, or community cloud model [6]. As this technology is proved to be a big benefit to all users but side by side there are many security challenges identity identification, traceability, availability, proof-of-ownership, access control, integrity, encryption, and key management. Existing cloud infrastructure faces some vulnerabilities that are used by adversaries to introduce some attacks [1]. Cloud computing is defined by Khorshed et al. [7] as "Cloud computing is a system of shared resources of a data center using virtualization technology. Such systems provide elastic on the basis of demand and ask for charges based on customer usage".

In a computing machine network mainly 3 types of intrusions can occur that are penetration, DoS and scanning [8]. Many security threats are being continually faced by the cloud-like hacking, DoS (Denial of Service), Cross Site Scripting (XSS), DDoS (Distributed Denial of Service) attacks and SQL injection [9]. The user information is not stored locally at the user's location but it is saved at the cloud provider's location. Because of this, it is understandable that the user may get concerned about their data and its security. Therefore, it is the responsibility of the cloud service provider to support data security of the user for their own good. As day by day data and information is being stored on the cloud environment because of which cloud security for information and data security is of greater concern. Because of this many distribute attacks like HTTP flood attack, protocol vulnerability exploitation, malformed packet attack, UDP flood attack, the slow Loris, the SYN flood attack, Ping of Death and ICMP flood attack is experienced by many networks.

This paper discussed the HTTP flooding attacks against the web servers on cloud computing and presents an organized survey concerning detection of HTTP GET flooding DDoS attack using MapReduce in cloud computing. Section 2 describes the DDoS attacks in cloud computing. Section 3

presents the techniques used to detect HTTP GET flooding using MapReduce and in section 4 current survey is compared with the existing surveys in particular field and finally, section 5 concludes this survey paper with possible future work.

2. DDOS ATTACKS IN CLOUD

In a cloud computing environment network security is one of the main challenges and among them, the DDoS (Distributed Denial of Service) attacks are of main concern. These DDoS attacks are distributed and coordinated on the large scale. In such attacks, the target network is flooded with immense data packet amount because of which network becomes incapable of handling such huge amount of data packets and therefore it cannot deliver the required services to its intended user and hence, network performance is degraded significantly [17][29]. The target network or machine resources becomes inaccessible by the users when the network is under DDoS attack. Though there are many reasons or ways to perform a DDoS attack on a variety of targets. Usually, it involves many attempts that can interrupt the target system either temporarily or for an indefinite time so that target system won't be able to provide its services to hosts that are connected to it via internet [2] [3] [4].

In the history of DDoS attacks, the very first DDoS attack was reported at the University of Minnesota in 1999. Many major and popular websites such as Amazin.com, CNN, eBay and Yahoo! Were affected by DDoS attacks in early 2000 [15]. Because of these attacks, these websites were not functioning for many hours and users were not able to access these websites [14]. As botnets were used by these attacks the target networks were affected greatly.

In 2014, the Arbor Networks [13] experienced the largest DDoS attack of that time of 400 Gbps. These DDoS attack incidents are increasing gradually with time [11] [12]. For example, public attention was drawn towards DDoS attacks when they affected the root DNS servers in the year 2003 and again in 2007 and large e-commerce websites were affected in February 2000. A new weapon of DDoS attack, called Mirai botnet a cyber-attack, was used in October 2016. Through Mirai botnet, most of America's Internet was brought down and it was considered as the largest in history. This DDoS attack was created using IoT i.e. Internet of Things which include digital video recorder (DVR) players and digital camera etc. whereas earlier botnets were simply made of computers. The first attack of Mirai botnet was made on servers of a company, Dyn, through which most of the Internet's Domain Name System (DNS) infrastructure was controlled. Because of this attack many sites like CNN, Reddit, Netflix, The Guardian, and Twitter were not working for many days in the US and Europe areas [10]. This all happened on 21 October 2016 and attack has a remarkable strength of 1.2 Tbps.

Many ways for detection, prevention, and mitigation of these Distributed DoS attack in a cloud environment was explained by many researchers. Two main detection techniques are used by all these techniques and they are signature or anomalies techniques. Either one or both the techniques can be used by a specific technique or new attacks can be learned by these techniques according to their set rules

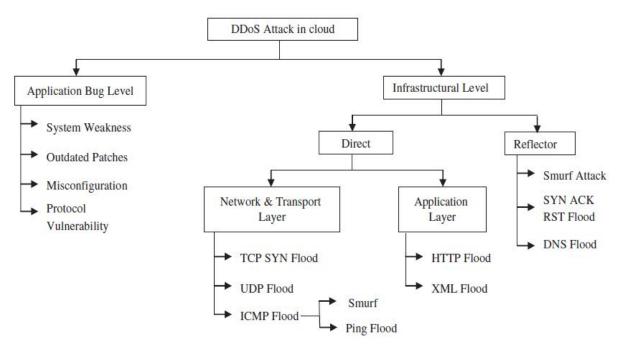


Figure 1: DDoS attack taxonomy in the cloud [1]

The DDoS taxonomy as shown in figure 1 was obtained from Aggarwal et al. (2017). The DDoS attacks in cloud computing mainly categorized into two categories: Application bug level and Infrastructure level. In Application bug level, the weaknesses of the system are exploited by the intruder so that users will not be able to access the cloud resources. These system weaknesses mainly include the misconfiguration, outdated patches, system weakness and protocol vulnerabilities etc. whereas in Infrastructure level, the cloud components are exploited by the intruders, for example, TCP buffers, CPU circles, network bandwidth, storage etc. It is also known as flooding attacks and these cloud components are flooded with packets so that the user will not use them any further. Intruders only need to know the IP address of the target system to attack it. These flood attacks are further categorized into direct attack and reflector attack.

In Direct attacks, the attacker makes use of some computer system to initiate the fraud packets which then exhaust the target systems resources and make them unavailable for the users. Further, these attacks are classified into application layer DDoS and network layer DDoS attacks. So in case of network layer DDoS attacks, the network and transport layer protocols are attacked or flooded to exhaust target system resources. ICMP flood, UDP flood, and TCP SYN flood come under network layer DDoS attack. In the application layer DDoS attacks the cloud services are targeted by the attackers through flood packets, especially HTTP flood packets. These packets are sent at a very high rate so that the target web server is overwhelmed at the cloud. These attacks mainly affect cloud providers' revenue, reputation, experience quality, service quality, and productivity. HTTP flood attack and XML flood attack are common examples of such attacks.

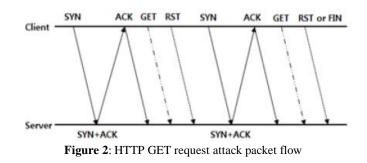
There exist many DDoS attacks in a cloud environment that are based on attack launch and implementation strategy, attack traffic rate, attack features. In this survey paper, we will be discussing HTTP GET flooding DDoS attack that mainly affects the application layer of the web servers on cloud computing.

2.1. HTTP GET flood attack in the cloud

This HTTP GET flooding attack is one of the usual DDoS attacks. It develops a normal TCP connection between the target network's web server and the client. Through this specific TCP, connection intruder tries to keep the server as busy as possible so that actual clients cannot get the required results. Intrusion Detection Systems (IDS) has the challenge to maintain both the accuracy and scalability of this large amount of data that is stored on cloud environment and which is rapidly increased year after year, from the DDoS attacks [19].

In HTTP flood attack web applications and servers are hacked. A legitimate TCP connection that consists of a pair of HTTP GET or POST session-based request messages is delivered to the target network's web server. These GET or POST request exhaust most of the resources of the server and are specially designed for this purpose so that server will not be able to respond to its actual user and resulted in Denial of Service situation because of high network traffic. To increase the power of overall attack these messages are sent via botnet in huge amount.

Whenever a TCP connection is set up the following HTTP GET request attack take place as shown in figure 1. The following figure shows the single TCP connection that processes HTTP GET request and the target server is affected because of low bandwidth is used



Intruder sends the HTTP GET requests continuously in the valid sessions to get the responses from the target. Such numerous HTTP GET requests are a new type of attack format that uses HTTP. Such single TCP connections that contain multiple HTTP GET requests are shown in figure 2

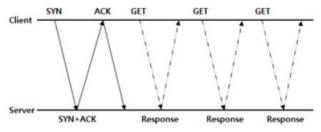


Figure 3: Single TCP connection containing multiple HTTP GET requests.

SYN rate limit detection method cannot detect this attack. This is because the rate based detection methods may have a higher threshold value than the traffic caused by HTTP floods. Therefore, many detection parameters should be used including rate-invariant and rate-based [21]. This attack can further avoid detection by requesting the server to alter a number of pages. This process continues so that database realizes the heavy load on the same page in a single connection through multiple HTTP GET requests.

The most non-vulnerable and advanced attack that is being faced by today's web server is the HTTP flood attack. The network security devices cannot easily distinguish between the malicious HTTP traffic and the legitimate HTTP traffic. This can result in numerous false-positive detection by the network security devices if it is not correctly handled.

3. TECHNIQUES USED TO DETECT HTTP FLOODING USING MAPREDUCE

Hadoop is an open-source distributed cluster platform that includes a distributed file system, HDFS, and the programming model, MapReduce. Lee et al. [18] extended their proposed P3 (Hadoop based Packet Processor) to develop a DDoS anomaly detection method on Hadoop by implementing detection algorithm that is based on MapReduce against the HTTP GET flooding attack. Two main MapReduce techniques that are used to detect HTTP GET flooding in the cloud are discussed as below.

The DDoS defense mechanism is categorized into 2 categories: Defense deployment and detection. Here, we will focus on the Detection of DDoS attacks in cloud computing. Further, the defense mechanism has two categories: Classification and Traceback. In Classification, there are 3 types of detection mechanism named: Signature-based, Anomaly-based and Hybrid mechanism for DDoS detection in cloud computing[22]. It is also noticed that earlier DDoS attacks were detected through Signature Based Classification method but it fails to detect recent DDoS attacks that are having unknown DDoS signatures. To overcome this situation the Anomaly Based Classification method proves to be more effective. These detection methods were able to detect unknown as well as known derivative attacks patterns. Furthermore, among anomaly based classification methods, the Support Vector Machine (SVM) method that is classified under machine learning methods is proved to be best for classifying the DDoS HTTP-GET flooding attack.

3.1 Access Pattern-based Method

It is assumed by the access pattern-based detection method that same behavior is possessed by the clients that are infected by the same botnet and normal clients can be easily differentiated by the attackers. Two or more than two MapReduce jobs are required by this method: the access sequence between the web server and client are obtained and byte count and spending time is calculated for every URL request by the first job, and these access sequence and spending time of all the clients are compared which are accessing the same server and then the infected hosts are found out by the second job. The huge computational complexity is required to identify the DDoS pattern is the only drawback of this method.

3.2 Counter Based method

The simplest detection method that counts the web page request numbers or the total traffic is a counter-based method. As in HTTP GET request to attack the traffic volume is low therefore such attacks are affecting the web servers more and more so counting the page request frequency from users can be an effective factor to find such attacks.

Lee et al. [19] proposed a MapReduce algorithm to detect DDoS with URL counting. Three input parameters are used in this algorithm that is: unbalance ratio, threshold and time interval. In this proposed MapReduce algorithm, the non-HTTP GET packets are filtered by map function and key values for client IP address, masked timestamp, and server IP address are generated. A conversion to MapReduce implementation can be easily generated because of low computational complexity. But to know the threshold value it requires some historical monitoring data criteria.

4. DISCUSSION AND COMPARISON OF THIS SURVEY WITH EXISTING SURVEY

The analysis of scientific relevant literature and synthesizes parameters on packet and traffic flow levels applicable to the detection of DDoS attacks in the infrastructure layer. It was concluded that detection of packet level uses two or more parameters, whereas detection of traffic flow level often uses only one parameter to make DDoS easy and resource-efficient. The attacks to the TCP / IP layer infrastructure are more common, according to [13]. The proportion of attacks of this type is 99.43 %, while the proportion of request attacks is 0.57% [14]. The reason for the relationship presented is perhaps because of less complexity than the application-based implementation of infrastructure-based DDoS attacks.

In the earlier sections, we discussed the DDoS attacks and its categorization into the infrastructural level. Further, we mainly concern about the infrastructural level attacks in this paper. It is observed that in recent years the infrastructural level attacks are more common as compared to application-bug level attacks. This is because these attacks are easy to implement and that too at low bandwidth requirement. The intruder floods the target web server with flood packets to exhaust its resources so that the target won't be able to provide its services to its actual users. It is also noticed that earlier DDoS attacks were detected through signature-based classification method but it fails to detect recent DDoS attacks that are having unknown DDoS signatures. To overcome this situation the anomaly based classification method proves to be more effective. These detection methods were able to detect unknown as well as known derivative attacks patterns. This survey is compared with the existing work in the given

This survey is compared with the existing work in the given table 1 and table 2.

AUTHOR/DATE	MECHANISM/ DETECTION TECHNIQUE	TOOL USED	DISADVANTAGES
Karnwal et al. (2012)	Filter Tree Approach [23]	Not specified	Cannot detect unknown attacks
Korad et al. (2016)	To detect DDoS Hadoop is used on Live Network [28]	1) Wireshark 2) Hadoop	 Internal attacks like Memory corruption cannot be detected For combining multiple nodes high computational cost is needed
Chen et al. (2016)	MapReduce	 Hadoop Spark VMware ESXi 6.0 Cloudera CD5 5.4 	Reduced processing time and increased efficiency
Hameed Ali et al. (2015)	Live DDOS Detection with Hadoop [22]	 HADEC Apache Hadoop 	 For small log files, parallelism is not offered by Hadoop Most of the detection is consumed by the capturing.
Choi et al. (2014)	MapReduce [24]	1) SNORT 2) Hadoop	Reduced processing time
Csubak et al. (2016)	Bigdata Testbed for detection of Network Attack [21]	 Python-dpkt package Wireshark NS3 Snort 	As a packet rate threshold is set so the only packet below this will be detected
Vissers et al. (2014)	Gaussian model with Parametric technique	1)Eucalyptus 2.0.3 2)Xen hypervisor 2.6.33.x	Protects only cloud Broker
Singh et al. (2014)	For Big Data Analytics Random Forests is used [27] in the detection of Peer-to-Peer Botnet	 Tshark using Libpcap library MapReduce Mahout Hadoop 	 Because of MapReduce usage, the computational cost is high As JVM and data requires large space, therefore, non-distributed classifiers cannot be used
Chen Xu et al. (2016)	For critical infrastructure Cloud computing based threat Detection network and monitoring system [25]	 PHP with AJAX Mysql database Spark Hadoop 	 Data samples that are collected decides the accuracy level Dynamic attacks cannot be detected Extra monitoring agents are required for new components

Table 1: Summary of DDoS techniques,	tools, and disadvantages of proposed schemes by authors
--------------------------------------	---

Table 2: Summary of data set and mechanism used by authors

Authors	DATASET USED	MECHANISM USED
Veetil and Gao (2013)	10% KDD intrusion detection dataset, Live	Packets per second, packets per minute
	network Stream packets as training data	
Cepheli et al. (2016)	DARPA 2000 dataset, Real	Protocol frequencies, packet sizes,
_	training data from a past	packet inter-arrival times
	penetration test of a commercial bank in	
	Turkey	
Csubak et al. (2016)	Simulated network traffic using NS3,	Packets per second rate
	Normal traffic data ranging	
	From MBs to GBs [21]	
Shamsolmoali et al. (2014)	CAIDA "DDoS Attack 2007"	Detection rate and false alarm rate.
Lonea et al. (2013)	DARPA 1999 Dataset	Detection rate and computational
Lo et al. (2013)	Snort	Detection rate and computation time

5. CONCLUSION AND FUTURE WORK

In this paper, we focused on the DDoS attacks and especially on HTTP GET flooding attack. The DDoS attacks are discussed with their history to have a proper insight into DDoS attacks. This paper also discussed the detection and prevention techniques for distributed denial of service attacks using the latest technique of Hadoop with MapReduce.

Regardless of the work that has already been done in this particular area, there is also a need to focus on the challenges that are still affecting the cloud services. A method should be developed that can help in detecting the application bug level as well as infrastructure level attacks. Efficient research should be performed for developing such detection solutions. Existing techniques should be modified as per the recent attacks and techniques or detection mechanism should be able to provide a good solution by detecting new attack patterns within the possible time to minimize its effect.

The future work of this particular survey includes the thorough analysis of DDoS attacks' characteristics, such as attack length, attack time period and attack magnitude and to design system for defending (detecting, preventing and mitigating) against all DDoS attacks on cloud computing.

REFERENCES

- 1. Agrawal, Neha, and Shashikala Tapaswi. "Defense schemes for variants of distributed denial-of-service (DDoS) attacks in cloud computing: A survey." Information Security Journal: A Global Perspective 26.2 (2017): 61-73.
- 2. B. S. Sumit kar. An anomaly detection system for DDoS attack in grid computing. International Journal of Computer Applications in Engineering, Technology, and Sciences (IJ-CA-ETS), 1(2):553–557, April-September 2009.
- 3. t. f. e. Wikipedia. http://en.wikipedia.org/wiki/ Denial-of-service_attack, Denial-of-service attack, 2013.
- 4. Y. FENG, R. GUO, D.WANG, and B. ZHANG. A comparative study of distributed denial of service attacks, intrusion tolerance and mitigation techniques, intrusion tolerance, and mitigation techniques. In Proc. of the 5th International Conference on Natural Computation (ICNC'09), Tianjian, China, pages 628–632. IEEE, August 2009.
- Choi, Junho, et al. "Detecting Web based DDoS Attack using MapReduce operations in Cloud Computing Environment." J. Internet Serv. Inf. Secur. 3.3/4 (2013): 28-37.
- Osanaiye, O., Choo, K.-K. R., & Dlodlo, M. (2016). Distributed denial of service (DDoS) resilience in the cloud: Review and conceptual cloud DDoS mitigation framework. Journal of Network and Computer Applications, 67, 147–165. doi:10.1016/j.jnca.2016.01.001
- 7. Md. Tanzim Khorshed, A. B. M Shawkat Ali, and Saleh A. Wasimi. A survey on gaps, threat remediation challenges and some thoughts for proactive attack

detection in cloud computing. Future Generation computer systems, 28(6):833–851, 2012. https://doi.org/10.1016/j.future.2012.01.006

- Rup K. Deka, Kausthav P. Kalita, Dhruba K. Bhattacharya, and Jugal K. Kalita. Network defense: Approaches, methods, and techniques. Journal of Network and Computer Applications, 57:71–84, 2015.
- 9. Deka, Rup Kumar, Dhruba Kumar Bhattacharyya, and Jugal Kumar Kalita. "DDoS Attacks: Tools, Mitigation Approaches, and Probable Impact on Private Cloud Environment." arXiv preprint arXiv: 1710.08628 (2017).
- 10. https://www.theguardian.com/technology/2016/oct/26/d dos-attack-dyn-mirai-botnet, Accessed: January 2019.
- 11. G. Gandhi and S. Srivatsa. An entropy architecture for defending distributed denial-of-service attacks. International Journal of Computer Science and Information Security (IJCSIS), 6(1):129–136, March 2009.
- S. Suriadi, D. Stebila, A. Clark, and H. Liu. Defending web services against denial of service attacks using client puzzle. In Proc. of the 9th International Conference on Web Services (ICWS'11), Washington DC, USA, pages 25–32. IEEE, July 2011.
 - https://doi.org/10.1109/ICWS.2011.22
- 13. http://www.arbornetworks.com, Accessed: December 2018.
- 14. Tao Peng, Christopher Leckie, and Kotagiri Ramamohanarao. Survey of network-based defense mechanisms countering the dos and ddos problems. ACM Computing Surveys (CSUR), 39(1):3, 2007. https://doi.org/10.1145/1216370.1216373
- Felix Lau, Stuart H. Rubin, Michael H. Smith, and Ljiljana Trajkovic. Distributed denial of service attacks. In: International Conference on Systems, Man, and Cybernetics, IEEE, Nashville, Tennessee, 3:2275–2280, 2000.

https://doi.org/10.1109/ICSMC.2000.886455

- Alzahrani, S. and Hong, L. (2018) A Survey of Cloud Computing Detection Techniques against DDoS Attacks. Journal of Information Security, 9, 45-69. https://doi.org/10.4236/jis.2018.91005.
- B. M. Leiner, V. G. Cerf, D. D. Clark, R. E. Kahn, L. Kleinrock, D. C. Lynch, J. Postel, L. G. Roberts, and S. Wolff, "A Brief History of the Internet," 2000. http://www.isoc.org/internet/history/brief.shtml
- Kene, S. G., & Theng, D. P. (2015). A review of intrusion detection techniques for cloud computing and security challenges. 2015 2nd International Conference on Electronics and Communication Systems (ICECS). doi:10.1109/ecs.2015.7124898
- Lee, Yeonhee, and Youngseok Lee. "Detecting DDoS attacks with Hadoop." Proceedings of the ACM CoNEXT Student Workshop. ACM, 2011. https://doi.org/10.1145/2079327.2079334

- 20. R. Ltd. DDoSPedia HTTP Flood. http://security.radware.com/knowledge-center/DDoSPed ia/http-flood/, 2013.
- Csubak, D., Szucs, K., Voros, P. and Kiss, A. (2016) Big Data Testbed for Network Attack Detection. Acta Polytechnica Hungarica, 13, 47-57.
- 22. Hameed, S. and Ali, U. (2016) Efficacy of Live DDoS Detection with Hadoop. IEEE/IFIP Network Operations and Management Symposium, Istanbul, 25-29 April 2016. https://arxiv.org/pdf/1506.08953.pdf
- 23. Karnwal, T., Thandapanii, S., & Gnanasekaran, A. (2012). A filter tree approach to protect cloud computing against XML DDoS and HTTP DDoS attack. In Proceedings of the International Symposium on Intelligent Informatics (ISI 2012), Vol. 182 of Advances in Intelligent Systems and Computing (pp. 459–469). Berlin/Heidelberg: Springer.
- Choi, J., Choi, C., Ko, B., & Kim, P. (2014). A method of DDoS attack detection using an HTTP packet pattern and rule engine in a cloud computing environment. Journal of Soft Computing, Springer, 18(9), 1697–1703. doi:10.1007/s00500-014-1250-8
- 25. Chen, Z., Xu, G., Mahalingam, V., Ge, L., Nguyen, J., Yu, W. and Lu, C. (2016) A Cloud Computing Based Network Monitoring and Threat Detection System for Critical Infrastructures. Big Data Research, 3, 10-23. https://doi.org/10.1016/j.bdr.2015.11.002
- Vissers, T., Somasundaram, T. S., Pieters, L., Govindarajan, K., & Hellinckx, P. (2014). DDoS defense system for web services in a cloud environment. Journal of Future Generation Computer Systems, Elsevier, 37, 37–45. doi:10.1016/j.future.2014.03.003
- Singh, K., Guntuku, S.C., Thakur, A. and Hota, C. (2014) Big Data Analytics Framework for Peer-to-Peer Botnet Detection using Random Forests. Information Sciences, 278, 488-497. https://doi.org/10.1016/j.ins.2014.03.066
- Korad, S., Kadam, S., Deore, P., Jadhav, M. and Patil, R. (2016) Detection of Distributed Denial of Service Attack with Hadoop on Live Network. International Journal of Innovative Research in Computer and Communication Engineering, 4, 92-98.
- M. Ezhilarasi V. Krishnaveni "A Survey on Wireless Sensor Network: Energy and Lifetime Perspective" Taga Journal vol. 14 pp. 3099-3113 ISSN 1748-0345