Volume 4, Issue 11, November 2014 ISSN: 2277 128X

International Journal of Advanced Research in Computer Science and Software Engineering

Research Paper

Available online at: www.ijarcsse.com

Proficient Network Traffic Classification Method with Cluster Labelling in CDN

V. Muthu
Assistant Prof. Dept. of Information Technology
Apollo Engineering College
Chennai, India

M. Subala
Assistant Prof. Dept. of Information Technology
VelTech HighTech Dr.RR. Dr.SR Engg college
Chennai, India

Abstract*: *Content Distribution Network is a large distributed system of servers deployed in multiple data centres across the internet. The main goal of CDN is to provide contents to end-users with high availability and high performance. But, nowadays there are many problems in achieving high performance due to heavy network traffic and security issues. The consequence of above problems leads to increased upload and download time of files, high time consumption in finding the nearest server etc. In order to address the above problems, we are proposing a cluster labelling algorithm which will reduce the network traffic and reduces the time consumption in finding the nearest server. In addition to this algorithm we are addressing the security issues like presence of malwares in requested files and DOS attacks by providing some methodologies. Also, we are introducing an IP Trace back system [1] which is used to trace the blacklisted domains that are accessing the CDN.

Keywords: Network Traffic, Cluster Labelling, DOS Attack, Upload and Download speed, IP Trace back system

I. INTRODUCTION

CDN is a network in which the servers are deployed worldwide and the clients can access the network anywhere anytime [2]. This network is mainly used for web objects like text, graphics and many applications like portals and live streaming media etc. The CDN operator gets paid by the content providers such as media companies for delivering their contents to their audience. In turn CDN pays ISPs, carriers and network operators for hosting its servers. In this network, the contents may exist on several servers. When a user makes a request to a CDN host name, DNS will resolve to an optimized server and that server will handle the request. Therefore, the advantage of CDN is that, the DNS itself chooses an optimized server and routes the request to that server. But, there are many problems in finding the nearest server to handle the request. When we are not able to find the nearest server in a limited amount of time, the network traffic increases and it leads to the denial of client's request. So we are proposing a cluster labelling algorithm which is used to cluster the servers based on their IP Address and latitude longitude position for finding the nearest server in very short period of time thereby reducing the network traffic and decreasing the upload and download time of files. There are many security issues related to CDN like presence of malicious software in requested files, DOS Attack etc. Malicious software's are those which will make our system to malfunction. DOS Attack is an attempt to make a system unavailable to its intended users. Usually, the attackers will flood the request to the server in order to make it unavailable to the clients who are all accessing that particular server. In section II we are going to see about the existing works related to Traffic classification in CDN. Section III deals with the proposed system. Section IV deals with the experimental results of our algorithm. Section V deals with the conclusion.

II. RELATED WORKS

The goal of the network traffic classification is to reduce the traffic between the client and the servers when they communicate with each other. The current research on classification concentrates on the methods to overcome the traffic and to enhance the security in the CDN network. This classification is considered as the superficial another approaches such as Port-Based Classification, Payload-Based Classification, Statistical-Based Classification, Machine Learning-Based Classification, and Manual Searching of Nearest Server.

- **Port-Based Classification:** This method is one of the traditional methods which are used to classify the Internet traffic by using UDP or TCP port numbers. Some traffic uses well known port numbers, and the port numbers can be found on Internet Assigned Numbers Authority (IANA) [3] .Therefore the traffic deduction is more since the same port number is used many times and the process is likely to give inaccurate estimates if the port number is dynamic.
- **Payload-Based Classification:** The packets are classified based on the payload of IP packets and the signature in the packet payload [4]. The main problem in this approach is if the data is encrypted then the payload is hard to identify. Users may encrypt the payload to avoid detection, and some countries forbid doing payload inspection to protect user information privacy. It shows only 50 percent to 70 percent efficiency. Also the classifier experiences heavy operational load in the applications.

- **Statistical-Based Classification**: This method we focus on the use of transport and flow layer behaviour statistics for packet classification [5]**.** This approach uses a set of sample traffic trace to train the classification engine to identify future traffic based on the application flow behaviours, such as packet length, inter-packet arrival time, TCP and IP flags, and checksum. The main disadvantage is the accuracy of classifying the encrypted traffic using this approach is low varying from 70 to 80 percent.
- **Machine Learning-Based Classification:** It is one of the efficient methods which follow two main approaches; they are supervised [6] and unsupervised traffic classification [7]. In both the approaches the traffic is classified only when the information of the traffic is given, if the information of the traffic is not known it is hard to classify the traffic. It shows poor performance because of the unknown flows. Some of the other existing methodologies are Bayesian Techniques [8] and etc.,

These are the drawbacks in our existing approach which is avoided by our proposed approach known as cluster labelling in CDN network

III. PROPOSED WORK

In our proposed work, we are going to see in detail about the cluster labeling algorithm. At first, we will study the architecture of our proposed work. Figure 1 explains about the Architecture diagram

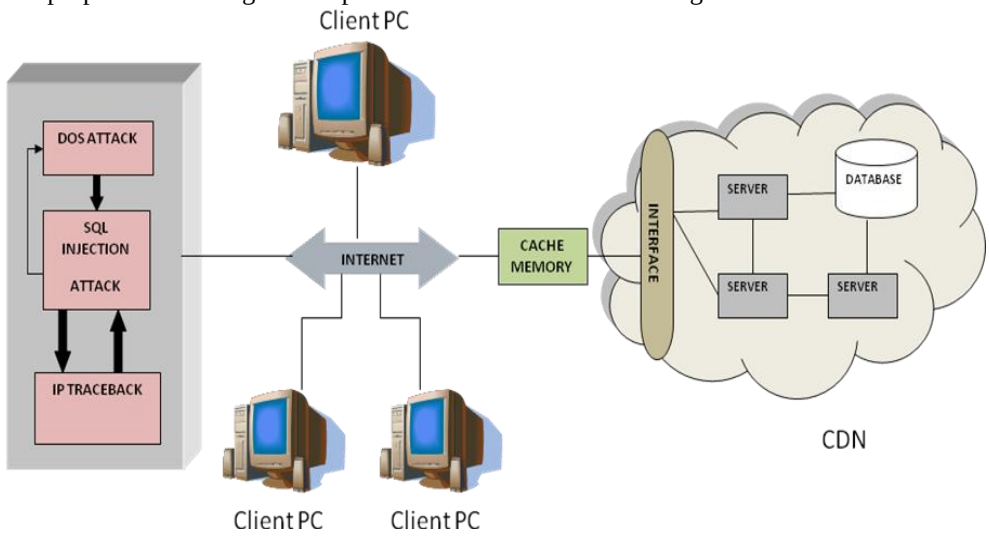

Figure 1: Architecture Diagram

- **CLUSTER ALGORITHM:**

Cluster algorithm is used to cluster the servers based on the latitude, longitude position and the IP address of the servers. The clustering depends on a factor called the exact position of the servers. When two servers are separated by a distance of less than 10 degree, then they both are clustered together and they form a cluster [9]. When the distances between the servers are more than 10 degree, then they won't be clustered together. These are the ways to cluster the servers. Whenever the client sends a request, our algorithm will first traces out the client's IP Address and client's latitude-longitude position. Then according to it, our algorithm searches for the nearest cluster present in the network. Once the nearest cluster is found out, our algorithm searches for nearest server that is present in the cluster and routes the client's request for that server. In case, if that server is busy at that time, then the request will be rerouted to the next nearest server that is present in the cluster. As a result of this, we can reduce the network traffic thereby, reducing the upload and download time of files and increasing the upload and download speed of the files. Figure 2 shows that how we have clustered the servers depending upon its position and IP Address.

- **ONTOLOGY:**

Ontology is an XML-based representation that is used to find the malwares and the malicious domains that are accessing the CDN. So whenever the user from a malicious domain enters the CDN network, they are blocked from using this network. It is also used to trace the malicious contents that are present in the malicious domains. Our ontology is created with the help of an open source website called "phishtank". This website periodically updates the malicious patterns and the malicious contents that are present in those domains. It also updates the malware information. So whenever a client uploads a file into the CDN network, the contents of the file will be checked for malwares. If any malwares present in that file, then it will be blocked from uploading and the client can't upload that particular file.

- **DOS ATTACKS:**

It is an attempt to make a network resource or machine unavailable to the users. One common method of attack involves saturating the target machine

Fig:2 Clustering of servers

With external communications requests, so much so that it cannot respond to legitimate traffic or responds so slowly as to be rendered essentially unavailable. Such attacks usually lead to a server overload. In general terms, DoS attacks are implemented by either forcing the targeted computer(s) to reset, or consuming its resources so that it can no longer provide its intended service or obstructing the communication media between the intended users and the victim so that they can no longer communicate adequately.To avoid dos attack we are going to set a parameter for the server to accept only possible number of clients.If the number of request exceeds the parameter value then the request will be redirected to the next server present in the cluster.

IV. EXPERIMENTAL RESULTS

Thus our experimental results shows that the upload and download speed of the files Gets increased when we use our clustering algorithm.

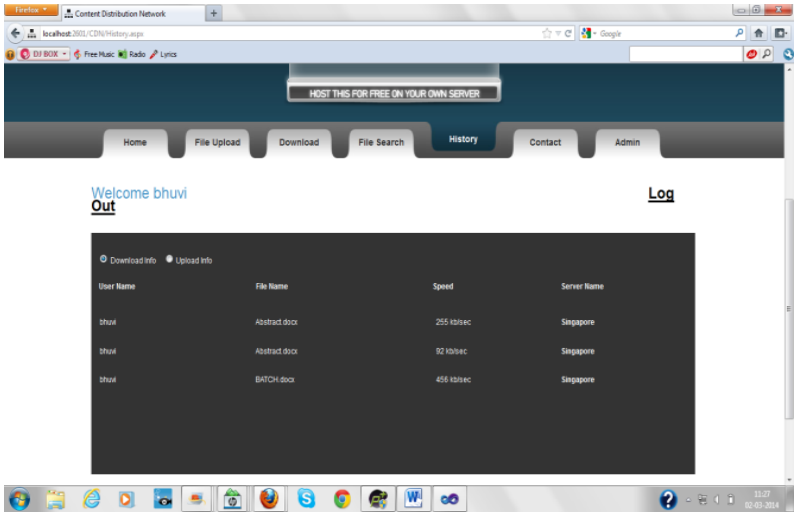

Figure 3: Download speed of files

V. CONCLUSION

This paper focuses on the concept of Cluster Labelling and on traffic issues and security problems in the CDN network. Traffic classification encounters more critical problems current advanced network and system, especially in CDN Environment. To address those problems, we have proposed cluster algorithm for servers. This will reduce the network traffic as well as increase the upload and download time of files. Also we have proposed some methodologies to overcome the security issues by creating an XML based representation called Ontology. Thus our result shows that our algorithm is better than the existing techniques and methodologies

REFERENCES

[1] Alex C. Snoeren, Craig Partridge, Luis A. Sanchez, Christine E. Jones, FabriceTchakountio, Stephen T. Kent, and W. Timothy Strayer, "Hashbased IP Traceback," in *Proceedings of ACM SIGCOMM*, August 2001.

[2] Akamai. Content Delivery Network. http://www.akamai.com.

[3] Internet assigned numbers authority (IANA), http://www.iana.org/assignments/port-number (last accessed October, 2009)

[4] A. Madhukar, C. Williamson, *A longitudinal study of p2 ptraffic classification*, in: MASCOTS '06: Proceedings of the 14th IEEE International Symposium on Modeling, Analysis,and Simulation, IEEE

Computer Society, Washington, DC, USA, 2006, pp. 179–188. doi:http://dx.doi.org/10.1109/MASCOTS.2006.6.

[5] T . Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. Springer, 2001.

[6] R. Alshammari and A. N. Zincir-Heywood, "*Machine learning based encrypted traffic classification: Identifyingssh and skype*," in Computational Intelligence for Security and Defense Applications, 2009. CISDA 2009. IEEE Symposium on, July 2009, pp. 1-8.

[7] Thuy T.T. Nguyen and Grenville Armitage. "*A Survey of Techniques for Internet Traffic Classification using Machine Learning*," IEEE Communications Survey & tutorials, Vol.10, No. 4, pp. 56-76, Fourth Quarter 2008.

[8] A. Moore and D. Zuev, "*Internet traffic classification using Bayesian analysis techniques*," in ACM International Conference on Measurement and Modeling of Computer Systems (SIGMETRICS) 2005, Banff, Alberta, Canada, June 2005.

[9] J. Erman, M. Arlitt, A. Mahanti, *Traffic classification using clustering algorithms*, in: MineNet '06: Proceedings of the 2006 SIGCOMM workshop on Mining network data, ACM Press, New York, NY, USA, 2006, pp. 281–286.

ABOUT AUTHOR

Mr. V. Muthu received his B.E(CSE) degree from JayaMatha Engineering College Nagerkoil in 2001 and M.Tech(IT) degree from Sathyabama University chennai, in 2009 .At present he is working as an Assistant professor with the department of IT at Apollo Engineering college, Chennai. He has the teaching experience of more than twelve years. His current research interests are in the fields of Network security

