# Distributed Denial of Service Attack Detection

Travis Blue and Hossain Shahriar

Department of Information Technology
Kennesaw State University, USA

tblue5@students.kennesaw.edu
hshahria@kennesaw.edu

**Abstract.** Distributed Denial of Service (DDoS) attacks has been a persistent threat for network and applications. Successful attacks can lead to inaccessible service to legitimate users in time and loss of business reputation. In this paper, we explore DDoS attack detection using Term Frequency (TF)-Inverse Document Frequency (IDF) and Latent Semantic Indexing (LSI). We analyzed web server log data generated in a distributed environment.

**Keywords:** Denial of Service, Latent Semantic Index, Term-Frequency, Inverse Document Frequency.

## 1 Introduction

Distributed Denial of Service (DDoS) attacks occur by issuing a large number of requests to a target web server. Existing network layer approaches are not applicable for detecting DDoS attacks. A number of bots are available in the market that can automate application layer DDoS attacks. DDoS as a service is now currently available to mount attacks on legitimate entities [1, 2]. DDoS attacks have been mounted against various websites such as game (Sony Play Station) [3] and bitcoin [4]. DDoS attacks can lead to loss in revenue ($5600 per minute), productivity, and reputation [5]. In this paper, we develop a mitigation approach for DDoS.

We first identify page ranking using a popular Term Frequency (TF)-Inverse Document Frequency (IDF) from web server logs. Then, we identify ranking of resources that are accessed most to build normal profile. For a given web session, we form a query of accessed resources and find how close it is with respect to the normal profile using Latent Semantic Indexing (LSI). If a large deviation is identified, the session is identified as part of DDoS attack.

## 2 TF-IDF and LSI Approach

TF-IDF is a computation approach to find the importance of word in a set of documents. It is composed by computing two terms: TF and IDF as discussed below. Term Frequency (TF) measures how frequently a term occurs in a document.

*TF(t) = (# of times t appears in a document) / (Total # of terms in the document) … … … (i)*

Inverse Document Frequency (IDF) measures how important a term is in all document. TF assumes that each is equally important. However, certain term may be common (e.g., article in sentence). Thus, IDF consider the frequent term as rare and less occurring term as important.

*IDF(t) =* $log_e$*(Total # of documents /#of documents having t) … … (ii)*

We apply Latent Semantic Indexing (LSI) technique to perform query and obtain the close similarity of documents for decision making. We represent access pattern from sample data in a matrix form to apply LSI, where each row represents specific resource access for a certain day, and column represents specific words found in log line with their TF-IDF values obtained from previous step. The columns would not only contain web page name, but also specific resources such as images, amount of bytes, status code, browser name.

For DDoS attack detection, we consider building a query obtained from an ongoing session data. We extract the resources (term) of interests from the log. Log files (documents) obtained for legitimate traffic for various days are used form Term-Document matrix. Terms are defined based on words representing resources (php files, image files) having non-zero TF-IDF value (discussed in previous section). Queries obtained for given sessions are used to find how close a given day's log represent for an ongoing session. Similarity measures are based on cosine metrics. The closer a query vector and a document is, the higher the distance. Hence, if distance is above certain threshold (d) level, it is not considered an attack. If the similarity is less than a threshold value, then the ongoing session is considered as an attack.

## References

1. J. Cheng, J. Yin, Y. Liu, Z. Cai, and C. Wu, "DDoS Attack Detection Using IP Address Feature Interaction," Proc. of 2009 International Conference on Intelligent Networking and Collaborative Systems, pp. 113-118.
2. C. Wueest, The continued rise of DDoS attacks, Symantec Technical Report, October 2014, Accessed from https://www.symantec.com/content/en/us/enterprise/media/security_response/whitepapers/the-continued-rise-of-ddos-attacks.pdf
3. Sony Playstation Hack, 2016, Accessed from http://www.scmagazine.com/sony-psn-downed-hacking-group-claims-ddos-attack/article/463065/
4. T. Bienkowski, Your Network or Your Bitcoins: Three Rules for Dealing with DDoS Extortion Threats, February 2016, https://www.arbornetworks.com/blog/insight/your-network-or-your-bitcoins-three-rules-for-dealing-with-ddos-extortion-threats/
5. What is a DDoS Attack? 2020, Accessed from https://www.verisign.com/en_US/security-services/ddos-protection/what-is-a-ddos-attack/index.xhtml