www.ijecs.in
*International Journal Of Engineering And Computer Science ISSN:2319-7242*
*Volume 3 Issue 10 October, 2014 Page No. 8686-8689*

# Detecting Client Based HTTP Attacks on Web Proxy by Temporal and Spatial Locality Behavior and Protocol Modification

***Jabin A[1]***

[1] P G Scholar,
Department of Computer Science and Engineering,
Government Engineering College , Idukki , Kerala,
India .
jabin.alif@gmail.com

**Abstract**: ***Distributed Denial-of-Service (DDoS) attack is a major threat for Internet applications. DDoS attacks relay on interrupting the services of a host connected to the Internet. In computer networks, a proxy server act as an intermediate server for bypassing the HTTP requests from clients to web servers. Client can access web server through different proxies. A novel attack detection scheme is proposed to prevent http attacks in web proxy based network. Here the attack detection is carried out at server. Accurate attack detection at server is difficult if proxy hide the information of clients. In the proposed scheme the detection is client based rather than proxy based. Here the client information is also passed along with the request so the server can easily identify the client and the attack detection become more accurate. This scheme utilizes locality behaviours such as Temporal and Spatial Localities(TSL) of proxy to server traffic.***

**Keywords:** DDoS, HTTP request, Hidden semi Markov model, TSL, soft control.

## 1. Introduction

In computer networks, a proxy server is a server that acts as an intermediary for requests from clients seeking resources from other servers. A client usually connects to the proxy server, requesting some service, such as a file, connection, web page, or other resource available from a different server and the proxy server evaluates the request as a way to simplify and control its complexity across the network. A Web proxy may be turned easily into an attacker by two steps: In the first step the attacker sends attack requests to a Web proxy and forces it to forward the attack requests to the origin server. In the second step the attacker disconnects connections between itself and with the proxy.

In DDoS attacks, attacker generates a huge amount of traffic which causes significant damage to the victim's network. The overlay proxy network being used to protect applications against DDoS attacks .Client can access web server through different web proxies which facilitates access to content on the World Wide Web. Web servers have no technique for identifying malicious client. In such cases, the innocent web proxies may get blocked. To avoid this problem, a client based approach is employed for detecting spatial and temporal attacks. This paper proposes a protocol modification for detecting actual attacking client rather than blocking the innocent proxy by adding custom headers in HTTP protocol.
In the final aggregated proxy-to-server traffic, it is very difficult to identify the difference between the normal traffic and the attack traffic. Thus, it is difficult for the victim server to identify and filter the attack requests. The proxy based HTTP attack is more flexible and covert than most of the existing DDoS attacks. The challenges of detection of attacks lies in three aspects which can be identified as follows: i) Real attacking hosts are unobservable to the origin server since they are shielded by the hierarchical Web proxies; ii) A Web proxy may be passively involved in an attack event and may act as an attacker without any awareness; iii) Observed from the victim server, both normal and abnormal traffic comes from the same sources (i.e., Web proxies). Although most of the large-scale official proxies are usually configured to be secure, they cannot avoid being abused for the proxy based attacks. This type of attacks may bring new challenges to existing network security systems [8]. Motivated by these issues, a novel resisting scheme is proposed to protect the origin server from Web proxy based HTTP attacks.

The scheme uses properties of network behavior analysis [1]. Here a Hidden semi-Markov Model is used to obtain the access behavior of the proxy server[3],[4].HsMM is a double stochastic process model. The hidden semi-Markov chain of an HsMM describes the transformation of a proxy's internal behavior states that can be considered as the intrinsic driving mechanism of a proxy to server traffic. In such behavior model, detecting the abnormality of a Web proxy can be achieved by measuring the deviation between an observed behavior and the Web proxy's historical behavior [9] profile. Long term and short term behavior assessment methods are proposed. Long term behavior assessment provides warnings on a large scale whereas the short term behavior assessment locates abnormal request sequences embedded in the proxy to server traffic.

Proposed scheme uses TSL based behavior analysis with IP based filtering to accurately detect the particular attacker client. The scheme also proposes a new soft control for attack response. A soft control scheme is used as an attack response. This approach block the server for providing service to the malicious sequences and allow server for providing services to the normal users.

The scheme reshapes the suspicious sequences according to certain prefixed criteria. It converts a suspicious sequence into a relatively normal one by partly discarding its most likely malicious requests instead of denying the entire sequence. Hence a behavior reshaping occurs each time before discarding an entire request sequence. This server side technology can be made implemented in large proxy-based live networks. The proposed system will be able to be used as an efficient defending mechanism in the server side against http attacks such as application layer DDoS attacks.

## 2. Related Work

A low-rate distributed denial of service (DDoS) attack has significant ability of concealing its traffic because it is very much like normal traffic. It has the capacity to elude the current anomaly-based detection schemes. In [5], an information metric based scheme is proposed. Here the metrics quantifies the differences of network traffic with various probability distributions. In this paper, two new information metrics are presented to detect low-rate DDoS attacks by measuring the difference between legitimate traffic and attack traffic. The proposed generalized entropy metric can detect attacks several hops earlier than the traditional Shannon metric. The proposed information distance metric outperforms the popular Kullback–Leibler divergence approach as it can clearly enlarge the adjudication distance and then obtain the optimal detection sensitivity.

In [10],a sequence order independent scheme is introduced for detecting application layer DDoS attacks. It includes the profiling of web browsing behavior. The sequence order may be more harmful than helpful in the profiling of web browsing behaviors because it varies significantly for different individuals and different browsing behaviors. In this scheme, some attributes are extracted from web page request sequences without consideration of the sequence order of requested pages. A model based on the multiple principal component analysis is proposed for the profiling of normal web browsing behaviors, and its reconstruction error is used as a criterion for detecting DDoS attacks.

Distributed Denial of Service (DDoS) attack is a critical threat to the Internet, and botnets are usually the engines behind them. Sophisticated botmasters attempt to disable detectors by mimicking the traffic patterns of flash crowds. This poses a critical challenge to those who defend against DDoS attacks. In the study of the size and organization of current botnets, it is found that the current attack flows are usually more similar to each other compared to the flows of flash crowds[6]. Based on this, a discrimination algorithm using the flow correlation coefficient is proposed The extensive experiments confirmed the theoretical analysis and demonstrated the effectiveness of the proposed method in practice. The flow correlation coefficient was used to measure the similarity among suspicious flows, and then the HTTP based DDoS attacks from normal flash crowds were discriminated by the results of measurement. A traceback method was explored for the DDoS attacks based on entropy variations between normal and DDoS attack traffic, which is fundamentally different from commonly used packet marking techniques.

Many methods designed to defense distributed denial of service (DDoS) attacks are focused on the IP and TCP layers instead of the high layer. They are not suitable for handling the new type of attack which is based on the application layer. In [7], a new scheme is introduced to achieve early attack detection and filtering for the application-layer DDoS attack. An extended hidden semi-Markov model is proposed to describe the browsing behaviors of web surfers. In order to reduce the computational amount introduced by the model's large state space, a novel forward algorithm is derived for the online implementation of the model based on the M-algorithm. Entropy of the user's HTTP request sequence fitting to the model is used as a criterion to measure the user's normality. Hidden Markov Model is defined as double stochastic process. The upper layer is a Markov process whose states are not observable. The lower layer is normal Markov process where emitted outputs can be observed. HMM is a powerful tool for modeling and analyzing proxy's access behavior. An extension of HMM is a Hidden semi Markov chain and each state has variable duration . The important difference between HMM and HsMM is that, in HMM each state has one observation while in HsMM each state can emit sequence of observations. The existing system [2], uses Gaussian distribution Gamma distribution HsMM because it require s fewer parameters to be estimated than the discrete HsMM which reduce computational complexity.

## 3. PROPOSED SYSTEM

The proposed system protects the victim server from web proxy based http attacks. In the existing system the attack traffic is assumed to begin from web proxies instead of its real sources and the victim server can only observe the intermediate proxy .Web proxy's access behavior depends on locality behavior such as temporal locality and spatial locality. This can be directly mapped into an HsMM. The attack detection is based on the abnormality in the requests. Attacking requests will be blocked and others will be served.
Temporal and spatial locality analysis used to extract the proxy to server behavior. Temporal locality of reference has been widely applied in many fields including program behavior reference pattern of Web access and Web proxy cache replacement strategy. Temporal locality refers to the property that a referencing behavior, pattern or web request in the recent past will be referenced again in the near future. The resource request popularity metric represents the frequency of the requests without indicating the correlation between document reference and the time since it was last accessed. Spatial locality refers to the behavioral property that frequently accessed objects and its neighboring objects in the past are likely to be accessed in the future. Spatial locality indicates relation among a group of HTTP request patterns.

Like most of other physical processes in nature, a Web-proxy's access behavior can be regarded as a combination of external manifestations (e.g., temporal and spatial locality) and intrinsic driving mechanisms (e.g., normality or abnormality). The

external manifestations are observable and usually controlled by the intrinsic driving mechanisms that cannot be accurately obtained by the origin server but can be estimated by the observable features of proxy-to server traffic.
In order to accomplish a healthy rule base against HTTP flood attacks , the initial step is to define the normal traffic. A basic abnormal traffic rule based on these baseline values could be sampled as 10 requests in 0.1 seconds .According to the normal traffic baseline values 1 request in 40.000 microseconds from a single IP address cannot be considered as a HTTP flood attack. The abnormal traffic rule allows 1 request in 10.000 microseconds at 10 times from a single IP address. Based on the rule creation concept, the rule also has a tolerance factor pointed out by 10 times description. Thus the tolerance factor (the request count) can give an opportunity to mitigate false positive .The detected IP addresses shared with other security components would also provide an opportunity to block attacker's access to the web application. In addition to this traffic rule, a different mechanism is implemented for detecting attacks such as unsupported HTTP method, oversized Header and Body data size, large or small time out interval, minimum incoming data and SQL Injection.

## 4. System Modules

The architecture includes five modules. They are attack class creation, training, detection, behavior reshaping and protocol extension.

### 4.1 Attack Class Creation

Initial work deals with identifying and creating the attack classes. Based on the attack history and proxy, origin server can predict certain attack classes. The different attack classes are the following:

- Unsupported HTTP method
- Oversized Header and Body data size
- Minimum incoming data
- SQL injection
- Large or small timeout intervals
- Path traversal

The above list shows some unavoidable attack classes. Unsupported http method sometimes leads to http flood attacks. Oversized header and body may lead to request abnormality. SQL injection is another kind of attack which uses code injection technique and the attacker gets unauthorized access .Thus all these kind of attacks must be given high consideration. As a result here we give top priority for creating the given attack classes.

### 4.2 Training Phase

The main step involved in the training phase is the attack learning. The attack learning is a complex process in which the request sequence characteristics are studied. The main characteristics here we take into account is the TSL or the Temporal and Spatial Locality behaviors. The incoming requests are treated as queries which are then compared against each attack classes. The request is then assigned with the attack type that provides the best match using probability distribution. Then identify the temporal and spatial sequence and their distance.

### 4.3 Detection Phase

During the comparison high priority is given to TSL behavior rather than normal behaviors. If any sequence found within the distance, that pattern is identified as an attack. If attack is found in the incoming request, then perform Temporal and Spatial Behavior Pattern Identifier analysis and organize incoming request as valid and invalid sequences.

### 4.4 Behavior Reshaping

This includes a soft control scheme which reshapes suspicious request sequence according to normal behavior. This process is done by partly discarding most likely malicious request instead of denying entire request sequence. This partly discarding is done based on a threshold value. The request is cut into parts based on this threshold value in order to attain reshaping.

### 4.5 Http Protocol Extension

Design and implement a new HTTP protocol for detecting client based attack instead of proxy based. Modify existing HTTP protocol by adding custom headers in HTTP protocol. These custom headers contain the IP of the client when forwarding requests from proxy server to server. So web server can group request from each client separately and easily detect attack based on client IP which is mandatory along with the request sequence.

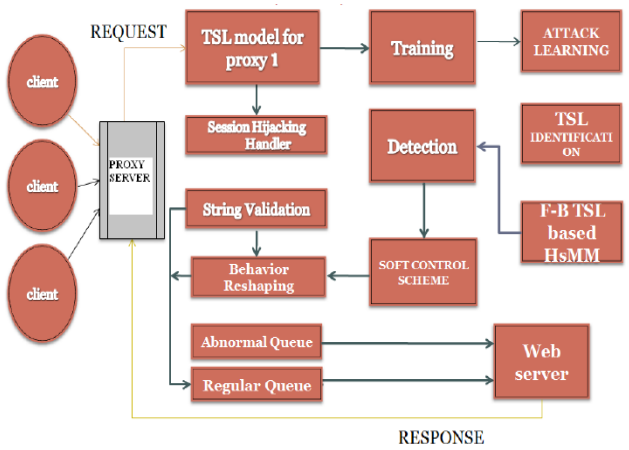


**Figure 1:** System Architecture

## 5. Performance Diagram

In the existing system, both false positive and false negative ratio increases as the difference between number of attacking and non attacking packets increases. But from Fig.2, it is clear that it have no effect on proposed system.

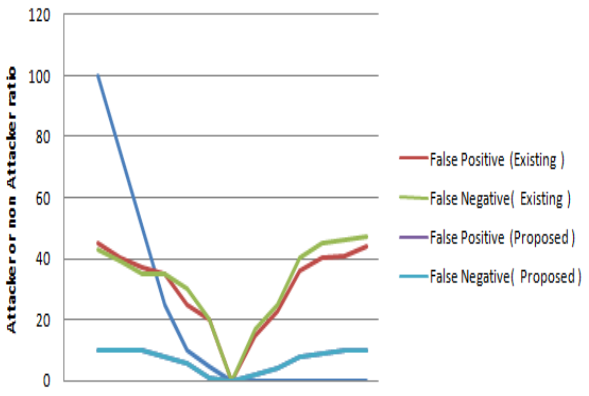


**Figure 2:** Performance Graph

## 6. Conclusion

A novel resisting scheme was proposed based on TSL behavior. This project made an attempt to filter attack traffic from the aggregated proxy-to-server traffic, which is considered as a new problem for the DDoS detection. Forward-Backward TSL based HsMM and soft-control scheme were proposed to improve the detection performance. Here detection performance is better than the pure statistical methods and it is independent of the traffic intensity. The major advantage of this scheme is that it can detect the particular attacker client, so the problem of blocking the innocent proxy never arises. The experiments can be further extended and in future the technology can be made improved so that it can be capable of handling other serious HTTP attacks such as IP spoofing.

## References


[1] P. Garcia-Teodoro, J. Diaz-Verdejo, G. Macia-Fernandez and E. Vazquez, "Anomaly-Based Network Intrusion-Detection: Techniques, Systems and Challenges," Computers and Security ,vol. 28,nos. 1/2, pp. 18-28, 2009.

[2] Yi Xie ,S.Tang and J.Hu,(2013), "Resisting web proxy-based Http attacks by temporal and spatial locality behavior," IEEE Transactions on Parallel and Distributed System.

[3] S. Yu, "Hidden Semi-Markov Models," Artificial Intelligence, vol.174, no.2, pp.215-243, 2010.

[4] J. Ferguson, "Variable Duration- Models for Speech," Proc. Symp. Application of Hidden Markov Models to Text and Speech, pp. 143-179, 1980.

[5] Y. Xiang, K. Li, and W. Zhou, "Low-Rate DDos Attacks Detection and Traceback by Using New Information Metrics,"IEEE Trans. Information Forensics and Security, vol. 6, no. 2,pp. 426-437, June 2011.

[6] S. Yu, W. Zhou, W. Jia, S. Guo, Y. Xiang, and F. Tang, "Discriminating DDos Attacks from Flash Crowds Using Flow Correlation Coefficient," IEEE Trans. Parallel and Distributed Systems,vol. 23, no. 6, pp. 1073-1080, June 2012.

[7] Y. Xie and S. Yu, "A Large-Scale Hidden Semi-Markov Model forAnomaly Detection on User Browsing Behaviors,"IEEE/ACM Trans. Networking,vol. 17, no. 1, pp. 54-65, Feb. 2009.

[8] Tao Peng and Christopher Leckie and Kotagiri Ramamohanarao (2006), "Survey of Network-based Defense Mechanisms Countering the DoS and DDoS Problems," ACM Transactions on Computational Logic.

[9] XIE Yi and YU Shunzheng, "A Detection Approach of User Behaviors Based on HsMM", ITC19/ Performance Challenges for Efficient Next Generation Networks.

[10] Y. Xie and S. Yu, "Measuring the Normality of Web Proxies Behavior Based on Locality Principles," Network and Parallel Computing, vol. 5245, pp. 61-73, 2008.


## Author Profile


**Jabin A** received B.Tech in Computer Science and Engineering from Kerala University, India in 2008 and currently doing M.tech in Computer Science and Systems Engineering at Government Engineering College , Idukki , Kerala. Research interests includes network security, data mining and protocol design.