International Journal of Computer Applications (0975 – 8887)
International Conference on Futuristic Innovations and Challenges to Diversity Management Emerging Technologies and Sustainability for Inclusive Industrial Growth (ICFIDM 2014)

A Novel Approach for Protecting Web Resources from Web Proxies

L.J. Yazhini Persis
Assistant Professor
JACSI College Of Engineering

Nazareth A. Angeline
Assistant Professor
The Oxford College of Engineering
Bangalore

J. Rajesh Dharmaraj
Assistant Professor
Jyoti Nivas College (Autonomous)
Bangalore

L.J. Arthur Neil
Senior Tech Lead
Happiest Minds Technologies
Bangalore

ABSTRACT
Static HTML was provided as a tool to display pictures and inert information. Consequently, as the internet and web access became more and more ubiquitous so too did the needs of those users who were accessing web applications. A "denial-of-service" attack is characterized by an explicit attempt by attackers to prevent legitimate users of a service from using that service. A novel server-side defense scheme is proposed to resist the Web proxy-based distributed denial of service attack. The approach utilizes the temporal and spatial locality to extract the behavior features of the proxy-to-server traffic, to protect weak signals from the interference of infrequent large values. Then, a new hidden semi-Markov model is proposed to describe the time-varying traffic behavior of Web proxies. The new method reduces the number of parameters to be estimated, and can characterize the dynamic evolution of the proxy-to-server traffic rather than the static statistics. It converts a suspicious traffic into a relatively normal one by behavior reshaping rather than rudely discarding.

Keywords
Traffic analysis, traffic modeling, distributed denial of service attack, attack detection, attack response

1. INTRODUCTION
Web based attacks focus on an application itself and functions on layer 7 of the OSI. John of the Gartner group claims that nearly 70% of all attacks occur at the application layer Application vulnerabilities could provide the means for malicious end users to breach a system's protection mechanisms typically to take advantage or gain access to private information or system resources. A "denial-of-service" attack is characterized by an explicit attempt by attackers to prevent legitimate users of a service from using that service. It includes attempts to "flood" a network, come in a variety of forms and aim at a variety of services. Flack crowd is characterized by the implicit attempt by user can access the web pages if the website respond in a certain time then the user sends the number of request for the page. A Web proxy may be turned into an attacker by two steps: attacker sends attack requests to a Web proxy and forces it to forward the attack requests to the origin server; attacker disconnects connections between itself and the proxy. the attack traffic is mixed with the regular client-to-proxy traffic by each proxy that forwards the traffic, Reference Characteristics[2]. In the final a proxy-to-server traffic, there is no obvious difference between the normal traffic and the attack traffic except their underlying purposes. Thus, the victim server is hard to accurately identify and filter the attack requests .In this paper, three driving mechanisms are defined: Normality, Transition, and Abnormality. The problem of resisting the proxy-based HTTP attack is equivalent to searching the Abnormality state of a Web proxy's access process and filtering those suspicious requests caused by the Abnormality state. Methodologies are the process of analyzing the principles or procedure for that making the resource available to the users by minimizing the workload of the website by recognizing and blocking the hackers Reference Characteristics [3].

2. METHODOLOGY
The deployment of Web application firewalls is the only proven method of protecting critical Web applications. Application firewalls operate at the application layer, not the network (or IP packet) level. They efficiently terminate all application sessions and perform a full, bi-directional parsing of all application data. By inspecting the actual HTML communications and understanding the context of all client requests and application responses in which these are sent to a Web server, an application firewall can enforce correct application behavior and block malicious activity [16].

2.1 Share Trading
The user can view the list of movie which is currently in play and get the details about the movie from the home page. When the user selects the specified movie from the list then they can access the details about the movie by passing query to the application and the details are fetched from the text files

2.2 Web Service
Web service is the technology which is used to put up all the business logic codes to the single application so that whatever the codes which is required by the application can be used by the specific websites by passing parameters to the specific web methods .It contains code for processing the user input.[9]

2.3 Web Proxy
Web proxy module helps to avoid burdens of origin server by receiving the request from client and forwarding it to origin server in order to get aggregated response Reference characteristics [15].

2.4 HTTP Attacker
Attacker will make suspicious request by initiating bulk of same URL request to the web proxy and he will force the web proxy to forward those requests to origin server [15].

2.5 Attack Detection

We can separate the normal user request and suspicious request that are generated by both client and attacker. Then we will discard those suspicious http requests[14].

3. SYSTEM ARCHITECTURE

To resist the Web proxy-based distributed denial of attack. The approach utilizes the temporal and spatial locality to extract the behavior features of the proxy-to-server traffic, which makes the scheme independent of the traffic intensity and frequently varying Web contents. A nonlinear mapping function is introduced to protect weak signals from the interference of infrequent large values. Then, a new hidden semi-Markov model parameterized by Gaussian mixture and Gamma distributions is proposed to describe the time-varying traffic behavior of Web proxies, Reference characteristics[5].The new method reduces the number of parameters to be estimated, and can characterize the dynamic evolution of the proxy-to -server traffic rather than the static statistics. Two diagnosis approaches at different scales are introduced to meet the requirement of both fine-grained and coarse grained detection. Soft control is a novel attack response method proposed in this work. It converts a suspicious traffic into a relatively normal one by behavior reshaping rather than rudely discarding Reference characteristics [3].

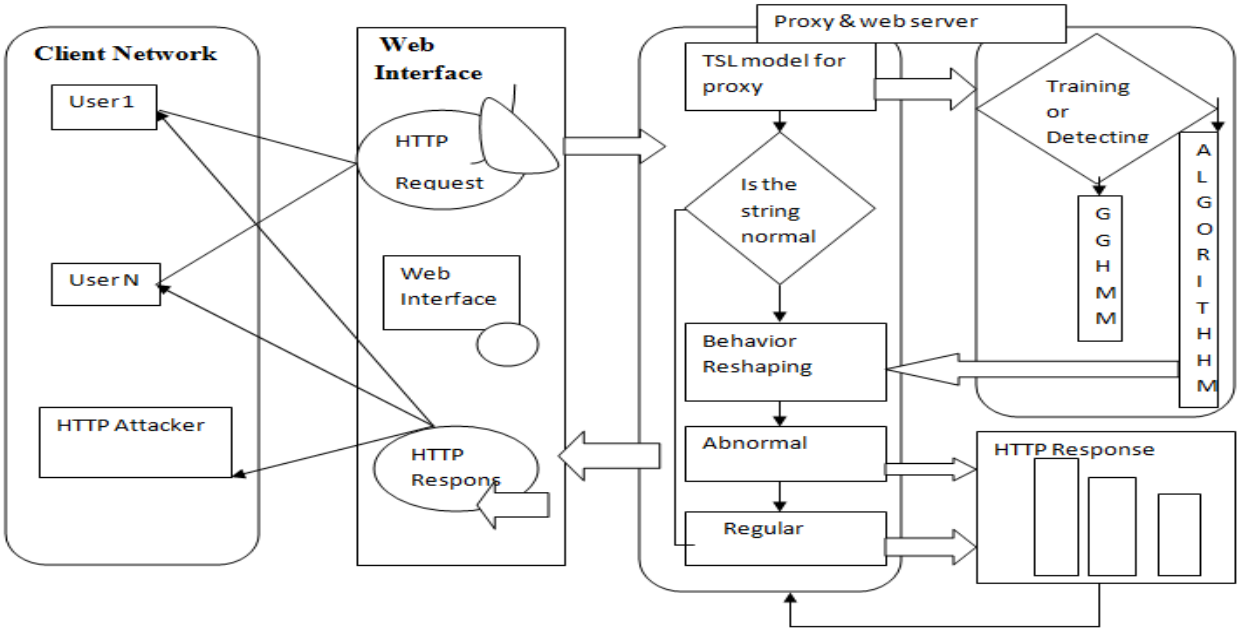


Fig 1: Proxy to Server traffic

4. TYPES OF ATTACKS

4.1 Web Based Attacks

4.1.1 Spoofing

Spoofing is the act of mimicking another user or process to perform a task or retrieve information that would normally not be allowed. An attacker could use a crafted HTTP request containing the session id information from another user and retrieve the targeted users account information [14].

4.1.2 Denial of service

Denial of service attacks are likely the most well-known of all application attacks, often generated by malicious users, competitors or script kiddies. Motivations for this type of an attack range from personal to political reasons in hopes of stifling an organization's ability to field online business. Famous examples include attacks upon SCO a couple of years ago by individuals upset about lawsuits aimed at LINUX Reference Characteristics [9].

4.1.3 Insecure Direct Object Reference

A direct object reference occurs when a developer exposes a reference to an internal implementation object, such as a file, directory, database record, or key, as a URL or form parameter. Attackers can manipulate those references to access other objects without authorization Reference Characteristics [10].

4.2 General Attacks

A Denial of Service attack is an attempt by a person or a group of persons to cripple an online service. Distributed denial-of-service attacks (DDOS) pose an immense threat to the Internet users to keep endorsed users of a website or web service from accessing it, or limiting their ability to do so. A flash event (FE) is a large surge in traffic to a particular Web site causing a dramatic increase in server load and putting severe strain on the network links leading to the server, which results in considerable increase in network traffic. A denial of service attack (DOS) is an explicit attempt by attackers to prevent legitimate users of a service from using that service. We interpret this definition broadly---we consider any attempt to undermine a Web site to be a denial of service attack. Network traffic anomaly detection can be done through the self-similar analysis of network traffic. In this case, the abnormal condition of network can be indicated by investigating if the performance parameters of real time data locate at the acceptable ranges Reference Characteristics [10].

4.2.1 Flash Crowd Attack

A Denial of Service attack is an attempt by a person or a group of persons to cripple an online service. Distributed

denial-of-service attacks (DDOS) pose an immense threat to the Internet users to keep endorsed users of a website or web service from accessing it, or limiting their ability to do so. If a DDOS attack or flash crowd attack occurs during a flash event, a Web server should aim to ignore DDOS requests and handle the legitimate requests. This requires the Web site to be able to distinguish between the two sets of requests and block both types of attacks Reference Characteristics [10].

4.2.2 Traffic Pattern Attack

Traffic patterns as seen by the Web site are important for several reasons. Overall traffic volume determines how much a server should provision resources to keep the site operational up to a certain level. If server load exceeds its maximum tolerance level which is pre-defined by its capacity, the server begins to slow down and can be driven to a shutdown. Thus, watching traffic patterns allows us to articulate the period when an unusually large number of clients can overwhelm a site and how much time the server has from the start of an FE or DDOS to take defensive measures [9].

5. TEMPORAL AND SPATIAL LOCALITY BEHAVIOR[6]

To resist the Web proxy-based distributed denial of service attack. The approach utilizes the temporal and spatial locality to extract the behavior features of the proxy-to-server traffic, which makes the scheme independent of the traffic intensity and frequently varying Web contents, Reference characteristics [15, 16].

5.1 Temporal Locality

Temporal locality refers to the property that a referencing behavior in the recent past is a good predictor of the referencing behavior to be seen in the near future, whereas the resource popularity metric only represents the frequency of the requests without indicating the correlation between a reference to a document and the time since it was last accessed.

5.2 Spatial Locality

Spatial locality refers to the property that objects neighboring an object frequently accessed in the past are likely to be accessed in the future. Spatial locality indicates correlation among a cluster of HTTP requests, capturing spatial locality can help mine the access behavior of Web proxies.

6. HIDDEN SEMI MARKOV MODEL

HMMs and HSMMs are, first of all, that both are built from two stochastic processes: an observed process and an underlying 'hidden' (unobserved) process Reference Characteristics [6]

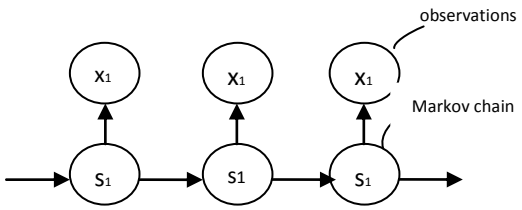
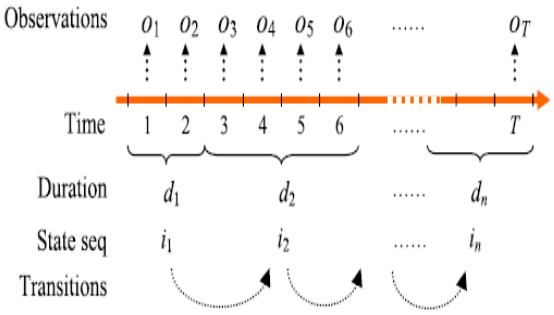


Fig: Structure of HsMM

- St1:t2= i – state i that the system stays in during the period from t1 to t2. In other words, it means St1= i, St1+1 = i. . . and St2= i. Note that the previous state St1−1 and the next state St2+1 may or may not be i.



- Hidden states : the (TRUE) states of a system that may be described by a Markov process (e.g., the weather).[6]
- Observable states: the states of the process that are `visible' (e.g., seaweed dampness).
- S[t1:t2] = i – state i which starts at time t1 and ends at t2 with duration d =t2 −t1 + 1. This implies that the previous state St1−1 and the next state St2+1 must not be i.
- S[t1:t2= i – state i that starts at time t1 and lasts till t2, with S[t1= i, St1+1 = i, . . . , St2= i, where S[t1= i means that at t1 the system switched from some other state to i, i.e., the previous state st1−1 must not be i. The next state St2+1 may or may not be i.
- St1:t2] = i – state i that lasts from t1 to t2 and ends at t2 with St1= i, St1+1 = i, . . . , St2] = i, where St2] = i means that at time t2 the state will end and transit to some other state, i.e., the next state St2+1 must not be i. The previous state St1−1 may or may not be i.

Web proxy's access behavior by an HSMM, each hidden semi-Markov state represents a driving mechanism of a type of proxy-to-server traffic. Two driving mechanisms are defined: Normality, and Abnormality. The problem of resisting the proxy-based HTTP attack is equivalent to searching the Abnormality state of a Web proxy's access process and filtering those suspicious requests [16] .

$$P(S_{t+1} = S_{t+1}) \,|\, (S_t = S_t, S_{t-1} = S_{t-1}, \ldots, S_0) = S_o) = P(S_{t+1} = S_{t+1} | S_t = S_t)$$

7. ALGORITHM USED

Gaussian distributions and Gamma distributions in Hidden semi-Markov Model (GGHsMM)

A hidden semi-Markov model (HSMM) is an extension of HMM by allowing the underlying process to be a semi-Markov chain with a variable duration or sojourn time for each state. Therefore, in addition to the notation defined for the HMM, the duration d of a given state is explicitly defined for the HSMM. The basic HsMM consists of a pair of stochastic processes:

The Observed process {ot} and the hidden semi-Markovstate process {Xt}, where t $\in$ {1; 2; . . .} is the

number of observation (also called event). {ot} is associated with {Xt} by the conditional distribution depending on the state process that is a finite-state semi-Markov chain. To model a Web proxy's access behavior by an HsMM, each hidden semi-Markov state represents a driving mechanism of a type of proxy-to-server traffic. In this paper, three driving mechanisms are defined: Normality, Transition, and Abnormality [6]. The problem of resisting the proxy-based HTTP attack is equivalent to searching the Abnormality state of a Web proxy's access process and filtering those suspicious requests caused by the Abnormality state. Given a behavior model (i.e., HsMM), this objective can be achieved by seeking the optimal underlying semi-Markov chain for an observed proxy-to-server traffic. The reasons for using the GGHsMM are as follows:

It has been proved that a finite mixture of Gaussian components can model/approximate any continuous distribution with arbitrary precision if a sufficient number of components is provided and the parameters of the model are chosen correctly.

Gamma distribution is a flexible distribution to express different distributions of practical signals by adjusting its two parameters. Moreover, our preliminary experiment has showed that the Gamma distribution is more flexible than other single-form distributions to fit the various duration distributions of the hidden states. The GGHsMM has fewer parameters to be estimated than the discrete HsMM, which greatly reduces the computational complexity[9].

8. CONCLUSION AND FUTURE ENHANCEMENT

In this proposal, we tried to filter the attack traffic from the aggregated proxy-to-server traffic, which is a new problem for the DDoS detection. A novel resisting scheme was proposed based on temporal and spatial locality. GGHsMM, multi precision diagnostic method and soft-control were proposed to improve the detection performance. This IP trace back scheme helps to determine some restrictions to attacker in order to prevent websites from n – number of time attacks refused by attacker can be included. Spatial locality refers to the property that objects neighboring an object frequently accessed in the past are likely to be accessed in the future.

9. REFERENCES

[1] Yi Xie, S. Tang,Y. Xiang,and J. Hu,"Resisting Web Proxy-Based HTTP Attacks by Temporal and Spatial Locality Behavior",IEEE Transcations on parallel and Distirbuted systems, VOL. 24, NO. 7, JULY 2013

[2] T. Peng, C. Leckie, and K. Ramamohanarao, "Survey of Network- Based Defense mechanisms Countering the Dos and DDos Problems," ACM Computing Surveys,vol.39, no. 1, hgp. 3, 2007.

[3] S. Lee, G. Kim, and S. Kim, "Sequence-Order-Independent Network Profiling for Detecting Application Layer DDos Attacks," EURASIP J. Wireless Comm. and Networking, vol. 2011, no. 1,p. 50, 2011.

[4] 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.

[5] S.-Z. Yu and H. Kobayashi, "An Efficient Forward-Backward Algorithm for an Explicit-Duration Hidden Markov Model," IEEE Signal Processing Letters, vol. 10, no. 1, pp. 11-14, Jan. 2003.

[6] http://sist.sysu.edu.cn/~syu/Publications/HSMMs_AIJ.pdf.

[7] A. Mahanti, D. Eager, and C. Williamson, "Temporal Locality and Its Impact on Web Proxy Cache Performance," Performance Evaluation, vol. 42, nos. 2/3, pp. 187-203, 2000.

[8] 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.

[9] http://www.ijert.org/view-pdf/9043/packets-flow-based-intrusion-detection-technique-for-websites.

[10] http://www.ipnetworksinc.com/pdfs/citrix/Citrix%20Appl.%20FW%20 White%20Paper.pdf.

[11] S. Choi and R. Wette, "Maximum Likelihood Estimation of the Parameters of the Gamma Distribution and Their Bias," Technometrics,vol. 11, no. 4, pp. 683-690, 1969.

[12] J.Devore ,Probability and Statistics for Engineering and the Sciences. Cengage Learning, 2008.

[13] Y. Zhong, X. Shen, and C. Ding, "Program Locality Analysis Using Reuse Distance," ACM Trans. Programming Languages and Systems, vol. 31, no. 6, p. 20, 2009.

[14] https://www.sans.org/reading-room/whitepapers/ application/web-based- attacks-2053

[15] http://www.computer.org/csdl/trans/td/2013/07/ttd20130 7140 1-abs.html

[16] http://www.researchgate.net/publication/260358229_Resisting_Web_ProxyBased_HTTP_Attacks_by_Temporal_and_Spatial_Locality_Behavior