**Imperial Journal of Interdisciplinary Research (IJIR)**
**Vol-2, Issue-5, 2016**
**ISSN: 2454-1362,** http://www.onlinejournal.in

# Advanced Traffic Management with Bot Detection and Geo Location Monitoring

A. Rama Rao[1], P. Surya Narayana Raju[2], P. Madhuri[3], S. Kalyani[4] & P. Darma Teja[5]

[1] Professor & Head of Department, Computer Science Engineering, Lendi Institute of Engineering & Technology, Jonnada, Vizianagaram

[2,3,4,5] Student, Department of Computer Science Engineering, Lendi Institute of Engineering & Technology, Jonnada, Vizianagaram

***Abstract*****:** *Internet is growing all over the world and become trend even in small villages. But providing fast access is becoming an issue. Controlling the traffic is the effective way to overcome this problem. Different types of traffic like organic traffic, bot traffic, paid traffic and direct traffic visit these websites every day. According to recent survey 56% of total traffic over internet is the bot traffic. Security from botnet has become a major issue for these websites and web services. Here we propose a bot detection machine for a single host which analyze traffic and detect spam. An enhanced user traffic profile is generated and used to filter out the normal traffic. The detection system is tested using real world bot. the proposed system achieves a high detection rate and a low false positive rate other system.*

*KEY WORDS: internet, websites, traffic characterization, botnet, traffic profile and security.*

## 1. Introduction

The media industry does not have any shortage of what to worry about. Audiences elsewhere are going, whether it's to internet sites like Facebook or digital-only opponents such as BuzzFeed, therefore advertising revenues continue steadily to fall--not for print just simply, but also for digital and video and just about everything. But there's been a bigger trouble for ad-based media which doesn't get discussed much: Namely, the known reality that a large chunk of the advertising market is dependent on fraud.

Department store magnate John Wanamaker famously stated: "I understand half the amount of money I devote to advertising is wasted, I simply have no idea which half." He was talking about traditional advertising in print newspapers and magazines and other formats, which were complicated to measure notoriously. As a total result, media companies could actually charge huge sums predicated on the assumption that lots of men and women saw an advertiser's message.

That was all likely to change when marketing and mass media went digital, since among the advantages of the web is that you could track almost every facet of someone's behavior, whether it is the period of time they spend on a full page, where they originated from, what browser they work with and what they clicked on.

It doesn't mean measuring the potency of marketing has gotten any much easier, however. Actually, it's arguably gotten actually harder, for several reasons. One is that no-one can appear to acknowledge what specifically media companies ought to be measuring: Clicks? Page-views? Different monthly visitors? Time spent on a page? Many advertisers remain attached to the thought of page-views or visitors as representing eyeballs, although analytics corporations like Chartbeat want to wean them from these metrics.

And that's only the start of the issues with the $14-billion online ad organization, as Sam Scott described in a recently available post at Moz. Another pressing issue is that, according to some estimates, over fifty percent of the marketing on the web is never basically seen by a individual. A study completed by ComScore that viewed ad campaigns in 2012 and 2013 deducted that 54% of the advertising in those promotions were never proven to a human being visitor, yet they were all without doubt counted as "impressions" in someone's ad budget.

As Scott explains, there are numerous of explanations why an ad wouldn't normally be demonstrated to a genuine reader: The ad could possibly be broken, and so it generally does not load or display effectively; it could be so sluggish to load that the individual browsing the website clicks away before it really is seen (but it's nonetheless counted as the feeling); and in some full cases it can be displayed below the edge of the screen, but nonetheless counted as having been experienced. In

part, that's for the reason that description of what an "impression" is, is still vague at best.

Those are a number of the less nefarious explanations why an ad wouldn't be observed, but would be counted. There are other darker explanations why that might be the case, however, including the use of tricks to inflate advertising counts--tricks such as for example what's called "pixel stuffing," where an ad made to appear at 1,024 by 480 pixels is crammed right into a one-by-one pixel square, but counted as the feeling still. Additionally, there is "ad stacking," where multiple advertisings are programmed for an individual slot.

However the biggest fraud of most is the utilization of "bots," software packages designed to mimic the actions of individual browsers. Such applications can drive large sums of visitors to websites, and may scroll through a niche site and select links even, as a human web browser would just. According for some estimates, around 60% of the traffic on the web is non-human.

As you participant in the online-advertising black industry described, there are visitors brokers who buy particular amounts of traffic made by bot farms and sell that visitors to marketers and publishers. And regarding to this broker, a lot of the ad and publishers systems mixed up in market plead ignorance, but know full very well what they are undertaking.

## 2. Internet Bot

An Internet bot, otherwise called web robot or just bot, is a product application that runs automated scripts over the Internet. Commonly, bots perform tasks that are both straightforward and basically redundant, at a much higher rate than would be feasible for a human alone. The biggest utilization of bots is in web scraping, in which a computerized script gets, breaks down and documents data from web servers at higher than the rate of a human.

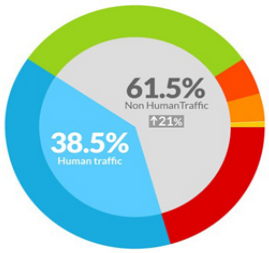
Given the outstanding pace with which bots can perform their generally basic schedules, bots might likewise be actualized where a reaction speed speedier than that of people is required. Normal cases including gaming bots, whereby a player accomplishes a critical point of preference by executing some dull routine with the utilization of a bot instead of physically, or auction-site robots, where a minute ago bid putting rate might figure out who puts the triumphant bid – utilizing a bot to put counterbids manages a huge favorable position over bids put physically. A bot is a malicious code or a software which is utilized as a platform for attacks such as distributed denial-of service attacks, fraudulent activities such as spam, phishing, identity theft and information eliminate and such other fraudulent activities. In our proposed paper we are going to detect bots and segregate traffic into bot traffic and normal traffic. Our scope of bot detection includes visiting count, referral link and mouse movement.

## 3. Traffic Analysis

There are also some good bots which are mastered by Google and Yahoo. The study states, “The more popular a website got, the harder it was for the good bots to keep up with the influx of human and bad bot visits.”

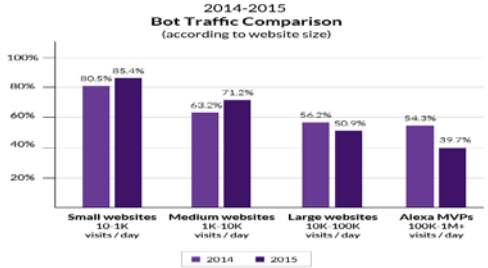
As the reports points out, there are two main reasons why good bots might visit a site: indiscriminate crawls (like search engine bots or marketing research tools) and targeted scans to learn more about your website (like uptime or SEO). “In both cases,” the report explains, “high website popularity is unlikely to translate into increased good bot traffic.”

## 4. Bot Traffic




Bot Traffic is the piece of online activity and exercises misleadingly created via computerized bots and creepy crawlies. Bot movement is apparently difficult to assess, yet as indicated by a few sources it can be evaluated to run from 10 to 20 % of activity. Every site naturally gets a set level of Bot Traffic. In this way, the lower the authentic activity is, the higher is the bot movement extent. Bot Traffic extents additionally shift as indicated by the site nature or action.

## 5. Bot Traffic Comparison




Most prominently, we saw that bot traffic to the most popular websites (those with 100,000 daily visits or more) went down from 54.3 percent to 39.7 percent.

Most prominently, we noticed that on websites with 100,000 or more daily human visits, good bot activity went down from 21.9 percent to 9.3 percent. This, interestingly enough, was in contrast to activity on low-tier websites, where good bot traffic actually went up.

## 6. Proposed System

Our proposed system includes 2 phases. Firstly the traffic generated by the user should belong to particular pattern or category. This fact is used to detect whether the user is a bot or a normal user. This detection will be done based on the visiting count and referral link from which he came into our website. Based on these factors, we can detect most percentage of bot traffic. Visitor count is the parameter which is considered by the number of visits of the user for particular website [4]. As visitor count for a normal user will always be up to a limit as a single user wont visit a website again and again in the same day. Hence we can segregate user into normal or malicious in this phase. The visitor having valid referral link and moderate visitor count will be considered as normal traffic and directed to valid page. There is a possibility that some normal users can be treated as malicious hence we developed phase 2 validation where user will be analyzed in detail. A visitor if found malicious in phase 1 will be sent for detailed analysis in phase 2.

## 7. Detailed Analysis

Detailed analysis of detection includes which includes captcha validation and analysis of user heatmap of mouse. CAPTCHA stands for Completely Automated Public Turing test to tell Computer and Humans Apart. It is a type of challenge response test to ensure that the response is only generated by humans and not by a bot [8]. CAPTCHA is the word verification test that the users come across the end of the verification for the website.


It is more easily possible for humans to look at an image and pick out the patterns than a bot. This is because bots lack the real intelligence that humans have by default. CAPTCHAs are implemented by presenting users with an image which contains distorted or randomly stretched characters which only humans should be able to identify. Sometimes, characters are stroked out or presented with a noisy background to make it much harder for a bot to figure out the patterns as it relies on visual test.

Now a days super bots are developed which scrape through jscript & html and find the captcha value and enter them. Hence we included heatmap of mouse movement into validation page. We analyze the heatmap and then find out whether visitor is normal or bot. Bots can't move their mouse pointer as a human do. Hence we can perfectly segregate whether visitor is bot or normal. Then the normal user is sent to valid site and malicious user is sent to invalid site.

## 8. Flow Diagram of Proposed System

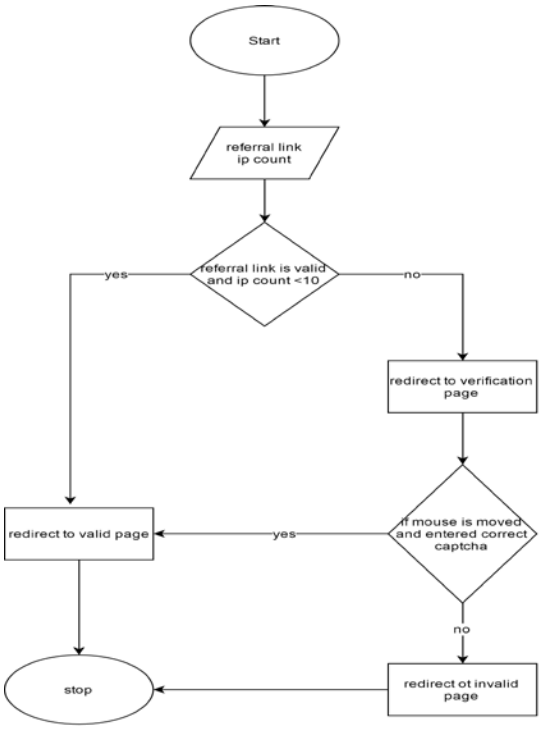



**Fig: flow diagram of proposed system**

Flow Diagram explains the complete flow of data in our proposed system which is from entering into our service through url to entering into valid website or invalid website.

## 9. Testing

We test our system to check its reliability, effectiveness and efficiency. We also checked this system to perform on any platform which has a browser in it. So there is no need for high performance systems to run it. The following test cases check every possible mistake our system can make.

### Case-1

Phase-1:

Given Input: Entered through redirection less than 10 times through a ip

Expected Output: Valid page should be opened

*Phase 2 won't be activated

**Case-2**

Phase-1:

Given Input: Entered url directly
Expected Output: detailed validation page should be activated

Phase-2:

Given Input: Entered correct captcha without mouse movement

Expected Output: Invalid page should be opened

**Case-3**

Phase-1:

Given Input: Entered url directly
Expected Output: detailed validation page should be activated

Phase-2:

Given Input: Entered correct captcha with mouse movement

Expected Output: valid page should be opened

**Case-4**

Phase-1:

Given Input: Entered through redirection more than 10 times through an ip

Expected Output: detailed validation page should be activated

Phase-2:

Given Input: Entered correct captcha without mouse movement

Expected Output: Invalid page should be opened

**Case-5**

Phase-1:

Given Input: Entered through redirection more than 10 times through an ip
Expected Output: Captcha should be activated

Phase-2:

Given Input: Entered correct captcha with mouse movement

Expected Output: valid page should be opened

## 10. Conclusion

We conclude that our proposed system is valid for detecting the bot traffic. We have achieved a detection rate of 100% with a false positive rate of 0%. Our approach is intended for ad networks which need to pay for visits and views of the normal users but not for bots. This system can also be used by other web services which are vulnerable to denial of service attacks which happen stealthily. Our Proposed system gives best results in stopping these attacks. In future we are going to provide this system as a service to different web sites. We are going to analyze, manage traffic and provide a best report about the traffic.

## 11. References

[1]G. Gu, R. Perdisci, J. Zhang, and W. Lee, "BotMiner: Clustering analysis of network traffic for protocol-and structure-independent botnet detection," in Proceedings of the 17th USENIX Security Symposium, pp. 139-154, 2008.

[2] J. R. Binkley and S. Singh, "An algorithm for anomaly-based botnet detection," in Proceedings of USENIX Steps to Reducing Unwanted Traffic on the Internet Workshop (SRUTI), pages 43-48, July 2006.

[3] W. Strayer, D. Lapsley, B. Walsh, and C. Livadas, "Botnet detection based on network behaviour,"Advances in Information Security, vol. 36, pp. 1-24 Springer, 2008.

[4]K. Takemori, M. Nishigaki, T. Takami, and Y. Miyake, "Detection of bot infected PCs using destination-based IP and domain whitelists during a non-operating term,"IEEE Global Telecommunications Conference, pp. 1-6, 2008.

[5]T. Wang and S. Z. Yu, "Centralized botnet detection by traffic aggregation," in IEEE International Symposium on Parallel and Distributed Processing with Applications, pp. 86-93, 2009.

[6]H. Xiong, P. Malhotra, D. Stefan, C. Wu and D. Yao, "User assisted host-based detection of outbound malware traffic," in Proceedings of the 11th International Conference on Information and Communications Security, pp. 293-307, 2009.

[7]W. Yan, Z. Zhang, and N. Ansari, "Revealing Packed Malware," IEEE Security and Privacy, vol. 6, no. 5, pp. 65-69, 2008.

[8]L. Zhuang, J. Dunagan, D. R. Simon, H. J. Wang, and J. D. Tygar, "Characterizing botnets from email spam records," in 1st Usenix Workshop on Large-Scale Exploits and Emergent Threats, pp. 1-9, 2008.

[9] H. Binsalleeh , T. Ormerod , A. Boukhtouta , P. Sinha , A. Youssef , M. Debbabi , and L. Wang, "On

the analysis of the Zeus botnet crime ware toolkit," in Eighth Annual International Conference on Privacy, Security and Trust , pp. 31-38, 2010.
[10] R. Borgaonkar, "An analysis of the asprox botnet," in 4th International Conference on Emerging Security Information, Systems and Technologies, pp. 148-153, 2010.
[11]B. Soniya and M. Wiscy, "Detection of TCP SYNscanning using packet counts and neural network," in IEEE International Conference on Signal Image Technology and Internet Based Systems, pp. 646-649, 2008.
[12]L. Shuai, G. Xie, and J. Yang, "Characterization of HTTP behavior on access networks in Web 2.0," in IEEE International Conference on Telecommunications, pp. 1-6, 2008.