A Real-time Comprehensive Architecture and Solution for Web Bulletin Board Security from Spamming Attack

Nune Sreenivas#[1], Dr.Gadiparthi Manjunath#[2]

[1]*Assistant Professor, School of Electrical & Computer Engineering, AAIT, Addis Ababa University Addis Ababa, Ethiopia*
[2]*Assistant Professor, Department of IT & SC, AAIT, Addis Ababa University Addis Ababa, Ethiopia*

Abstract *- Spamming is one of the most prominent challenges over the web. It is an event of flooding various website with sort of data that people would not have otherwise received. Generally spamming is performed in two major ways: Using automated programs and using Human Spammers. A forum spamming is an issue where either automated programs or human belonging to a spamming society registers itself with various Forums (Web Bulletin Board) and post same message. This is primarily done for two main purposes: To create back link of a website from numerous websites to increase the search engine ranking of the website and to advertise certain products whose advertisements are refused to be accepted by web audience. In this paper we analyze various security threats and their preventive measures with implication on server and the Bulletin Board itself with the help of real time data from grasshoppernetwork.com which is a forum running on mybb Bulletin Board. We analyze various threat models and propose a comprehensive model to protect the forum from Spammers. A Php based Modification is developed as anti spamming measure and integrated with the existing website. Result data shows significant improvement in preventing spam with our algorithm.*

Keywords: *Web Security, Bulletin Board Security, Spam.*

I. INTRODUCTION

Forums or digital Bulletin Boards are the web based message boards for sharing web audience opinion, knowledge. Over last few years, the popularity of such forums has increased due to consistent user participation and many online communities. These forums are basically software engines which are powered by various software like Php/MySQl (VBulletin, mybb, phpbb, SMF), .Net (YAF). The general architecture of the forums is presented below. It is clear from Fig. 1 that main components of the forums are: Users which are further divided into groups, Forums which are the categories on which audience participation is expected, Threads: main discussion components, posts: the messages presented on the threads by user and user groups. Anonymous users or the users who are not registered with the web sites are allowed to view the posts but are not permitted to post. In order to prevent automated software (Called Bots) from registering to the forum various protocols are used which are called challenges. One of the most common forms of the challenge is CAPTCHA. There are several other security proposals towards this direction like asking simple questions for registration, enabling java script support and so on. Though they minimize the attack from weaker spam bots, human and strong spamming applications can still break in.

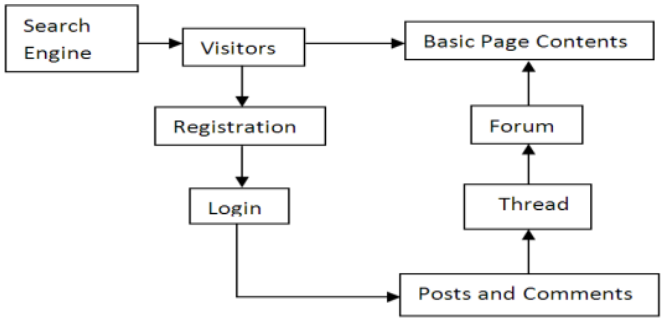

Fig. 1. Basic Forum Engine

Hence developing anti-spamming standard is not restricted to making their registration difficult but at the same time to monitor the Bulletin board constantly for spamming activity.
Types of threats are categorized in Fig. 2.

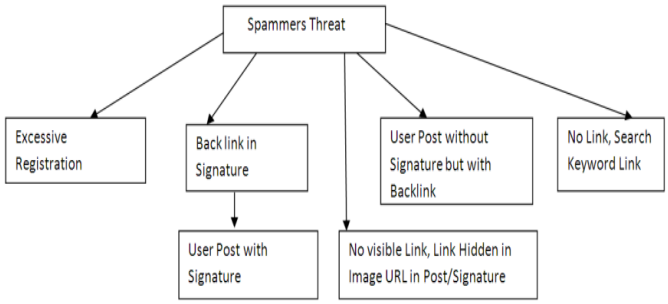

Fig. 2. Type of Attacks of Bulletin Boards

One of the most common approaches that is widely adopted by the Forum administrators to protect the forum from such spammers is by blacklisting the IP addresses and by ensuring that no spammers are revisited from the same IP address. Also the user

account and the email addresses are banned which makes it impossible for the spammers to reuse.

The main problem with this theory is that normally spammer's uses proxy IP addresses. Hence banning the IP address never prevents the spammers from attacking. Many administrator uses anti Proxy strategy for registration which disallows users from registering if they are trying to access the forum from proxy address. Many user these days uses wireless and Wi-Fi internet connection which allocates a temporary IP address to the user. Therefore using this strategy minimizes the user participation and many valid users are also denied.

II Related Work

One challenge in web security education is its interdisciplinary and practical nature which is explained by [1]. [2] Explains some web security measures that [3] introduces the commonly used Web security comprehensive evaluation tools. A component-based framework as well as an open source solution is given in [4].[5] propose a formal model of web security based on an abstraction of the web platform and use this model to analyze the security of several sample web mechanisms and applications. [6,9] On the heels of the widespread adoption of web services such as social networks and URL shorteners, scams, phishing, and malware have become regular threats. Despite extensive research, email-based spam filtering techniques generally fall short for protecting other web services. In [7], a mobile service, CAIS, is proposed. CAIS stands for Context Awareness Information Services. It is composed of several mobile services to achieve: 1. context-awareness communication, 2. location-based communication, 3. mobile information sharing, 4. autonomous communication, 5. user privacy protection, 6. negligible cost.

III Proposed System

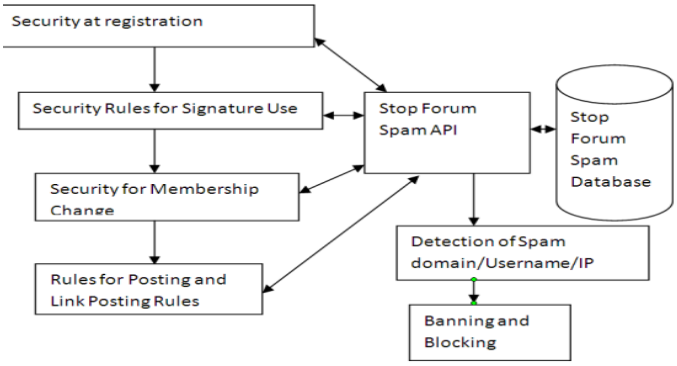

Fig. 3. Proposed Architecture of the System

In this paper we present a comprehensive architecture for spam detection and forum protection against spam based on combined data mining, manual and automated process. In this paper a detailed algorithm is presented for anti-spamming realization which is implemented in php. The proposed architecture diagram explains the security extension proposed in the paper. At various levels, Rules are added. These rules are collaborated with Stop Forum Spam database with an API bridge. At each level of Forum activity detection of spammers is performed. If new domains and IP addresses are identified, then the StopForumSpam database is updated so that the next attempt of spamming from the database can be prevented.

IV Methodology

Algorithm:

1) Registration:

a) Add Bio data Field which must be minimum of 240 characters. The Bio data must include purpose of registration of the user, his opinion about the site and other description.

b) Upon submission compare the content of Bio data field with tag library of the website to compare the content similarity.

c) If the Bio data content is not similar to 10% of the site content or type of visitors, than block the registration process.

d) Use a Hidden JavaScript for Spammers and bots cannot evaluate and parse JavaScript.

e) Use a Simple security Question that adds a question at the time of registration. The questions are randomly picked and only answerable by humans.

f) Use registration confirmation

Many spamming domains do not have the facility of incoming mails. They generally are used for flooding mails and newsletters to email addresses and forums. Enabling registration confirmation sends an authentication script to the email of the registering member. Once clicked on that script, the member's registration process is completed.

Although the mod comes with the core of mybb, this needs to be conFig.d.

2) Prevention against Human Spammers Posting

a) No permission to use signature before at least 10 posts are made.

b) No Permission to use any html or hyperlink before 10 posts are made.

This prevents the Human spammers to just join the Forum and bombard forum with irrelevant posts containing link to Spam Sites.

3) User Grouping

a) Initiate two separate Groups: Newly registered User and Registered Users.

b) Newly Registered Users are not allowed to visit any other forum other than the introduction forum which is not indexed by the search engine.

c) Only after making one post in the introduction section, user is migrated to Registered Group. Therefore preventing the spammers from creating deep link with the site.

4) StopForumSpam mod with Bulletin Board

a) StopForumSpam is an internet wide group that maintains and manages a database of the spammers IP address and domains. We use them in two different

ways: For protecting against allowing registration from blacklisted domain and for updating the database once a human spam was detected for users failing to migrate from registered group from Newly Registered Group. Our Script detects and filters those IP addresses whose users have successfully registered but have failed to be qualified as valid members.

V RESULT AND ANALYSIS

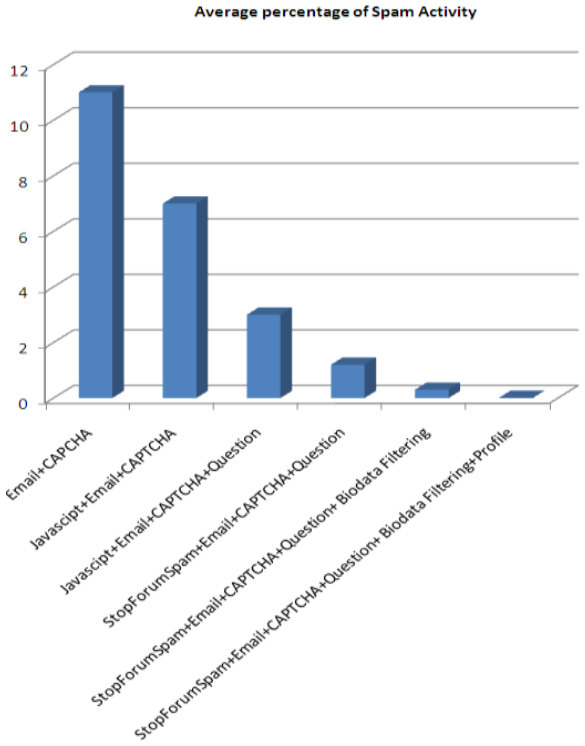

Fig. 4. Result of Usage of Modules and average Spam activity over one week and

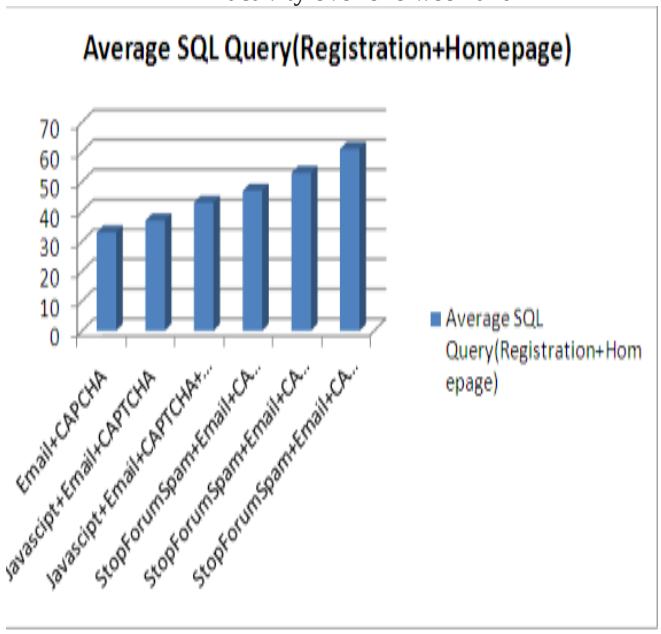

Fig. 5. Adopted Security extension v/s Average Number of SQL Query

We define average spam activity as the average ratio of total spam activity to the total number of transactions (registration, posting, new thread creation, Giving Reputation to member, Profile updation, posting) over a week. The spamming activity is marked by the admin and is logged in the record for the test.

Fig. 4 and 5 demonstrates that by adding the suggested module, amount of query being executed at the server increases marginally but the performance gain interms of avoiding the spammer increases many a fold.

VI CONCLUSION

Discussion forums and bulletin boards are widely used software over the web. But most of the bulletin boards suffer from Spamming attack. Such attack not only disturbs the integrity of the forums by adding noisy data, but also increases the server overhead. Back links from blacklisted sites degrades the page rank of the website. Beside there are chances of intellectual property theft and hacking of the sites. Though there are several modules available to provide protection against such attacks, there are no open standards for the same. Hence the problems of spammers are increasing by each passing day. In this work we developed the architecture and the solution against spamming attack and implemented the same in www.grasshoppernetwork.com website. The results show astonishing improvement interms of driving the spammers away, both Human as well as Bots. Though spamming activity could not be brought down to zero, it was minimized nevertheless.

The work can be further improved with data mining based security enhancement and rule integration such that the Spammers and the Bots can be prevented at the server level itself. i.e. spammers can be disallowed to view the site itself.

REFERENCES

[1] Tao, L., Chiou, C. L., Lin, C.: Work in progress — Improving web security education with virtual labs and shared course modules. Frontiers in Education Conference (FIE), IEEE (2010)

[2] Curphey, M., Arawo, R.: Web application security assessment tools. Security & Privacy, IEEE, Volume 4, Issue 4 (2006)

[3] Hui-zhong S., Chen, B., Yu, L.: Analysis of Web Security Comprehensive Evaluation Tools. Networks Security Wireless Communications and Trusted Computing (NSWCTC), 2010 Second International Conference Volume 1 (2010)

[4] Sheng-Kang, L.: From Web server security to Web components security. IEEE 37th Annual International Carnahan Conference on Security Technology Proceedings Digital Object Identifier: 10.1109/CCST.2003.1297556 Page(s): 176 – 182 (2010)

[5] Akhawe, D., Barth, A., Lam, P. E., Mitchell, J., Song, D.: Towards a Formal Foundation of Web Security. Digital Object Identifier: 10.1109/CSF.2010.27 Page(s): 290 – 304(2010)

[6] Thomas, K., Grier, C., Ma J., Paxson, V., Song, D.: Design and Evaluation of a Real-Time URL Spam Filtering Service.

Security and Privacy (SP), IEEE Symposium on Digital Object Identifier: 10.1109/SP.2011.25, Page(s): 447 – 462 (2011)

[7] Yih-Jiun, L., Kai-Wen, L.: Location Based Enabled Context Awareness Information Service, New Trends in Information and Service Science. NISS '09. International Conference on Digital Object Identifier: 10.1109/NISS.2009.160 Page(s): 944 – 947 (2009)