Classification of Spamming Attacks to Blogging Websites and Their Security Techniques

Rizwan Ur Rahman
Maulana Azad National Institute of Technology, Bhopal, India

Rishu Verma
Jaypee University of Information Technology, India

Himani Bansal
Jaypee University, Solan, India

Deepak Singh Tomar
Maulana Azad National Institute of Technology, Bhopal, India

INTRODUCTION

With the explosive expansion of information on the World Wide Web, search engines are becoming more and more significant in day to day lives of humans. In China only, there are more than one twenty million Internet users, among which just about eighty percent use search engines regularly and eighty eight percent using search engines as a key means to acquire newly appeared information (Liu et al., 2008).

Even though a search engine generally gives huge number of results for certain query, the majority of the search engine users simply view the first few web pages in result lists. Consequently, the ranking position has become a most important concern of internet service providers.

Spamming is the method of intentionally manipulating HTML pages to achieve high ranking of search engines. Spamming is exploited to deceive search engines indexing program and to gain ranking position.

The spammers exploit vulnerabilities in web application especially Blogging sites. Security issues in Blogging Websites are still exploratory and in spite of an increase in Blogging Websites research and development, lots of security challenges remain unanswered. Spamming are the most malicious threats to the web application, especially Blogging Websites (Rahman et al., 2008).

The main objective of this chapter is to examine to what level spammers could be threat to Blogging Websites. In first section the terms related to spamming are defined, and a sufficient overview of spamming will be presented in order to give the reader with an understanding of the background for the remaining Sections.

The subsequent section will present the indications and signs of spamming attacks. This section will provide the information of different categories of spam such as Content Spamming and Form Spam. A section devoted to Security Techniques including detective and preventive techniques will be presented.

The first section introduces the overview of web spamming including Content Spamming and Link Spamming, Click Spam and Form Spam. Further, this section elaborates the attacks form spamming on blogging Sites.

DOI: 10.4018/978-1-5225-9715-5.ch058

The next section presents the vulnerabilities in blogging sites their types and vulnerability scanners.

The last section presents the taxonomy of detection and prevention methods such as various forms of CAPTCHA, and Honeypot. It also explores the Data mining techniques including Support Vector Machine. At last conclusion of the chapter is presented.

TYPES OF WEB SPAMMING

Content Spamming

Content Spamming is the first and most widespread type of web spam because it exploits search engines based information retrieval models. These models are build from page contents which further ranks the pages on the bases of the page rank algorithms. As the result the spammers analyse the weakness of these models being implemented and exploit them. The different types of content spamming are Title spamming, Body spamming, Meta tag spamming, Anchor Text spamming, URL spamming (Henzinger, 2007).

As the title field is very important in the information retrieval, spammers try to overfill it in order to increase the page rank and this type of spamming where one overfills the title is called title spamming. In body spamming the body of the page is modified and injected with certain content or queries that are frequently searched. The meta-tags play a special role in document description as when we use search algorithm n search engines the results are fetched on the bases of the meta-tags on the webpage. So placing the spam in this content will be the most efficient way to spam the document (called meta-tag spamming). Anchors tag are the tags used to include links on the website so in anchor spamming the spammers create the links with the desired anchor text to get the right term for the target page. In URL spamming the content to be searched is itself injected to the URL.

Link Spamming

There are two main categories of link spamming namely

1. Outgoing Link spamming, in this the spammers have direct access to the page and can therefore add any content to the web page. They can easily copy the entire web catalogue.
2. Incoming link spamming, where the spammers try to raise the Page Rank and boost the number of incoming links to the page (Gyongyi et al., 2008).

Cloaking and Redirection

Cloaking is a way to provide different versions of a page to the user/crawlers based on the information contained in the request made or query being searched. The other way is to redirect users to the malicious pages by executing JavaScript. The JavaScript redirection spam is the most widespread and is the one of the most difficult detection spam by the crawlers.

Click Spam

In this type of spamming, attacker executes clicks for end users who have not made them.