Internet Programming I

Extra- Credit

Zainab Ebrahimi

Protecting your website from Spam

We live in a Internet driven society. We use Internet services everyday to share information, to schedule a meeting, to watch movies, etc. With the great amount of advantages also comes the inevitable threat of spam. A revolutionary security technique called a honeypot has now been implemented to trap website spammers. How do they work and how beneficial would it be to implement it on my website?

Spam is basically an over flood of the same unsolicited “questionable ”commercial message. The problem with spam is not only that it’s annoying but it’s also a security threat. You can’t be sure what’s behind that link or what will happen if you fill out that form. And most importantly, if a user of your website get spammed by using your website, you have a big problem. The worst part is that spam is incredibly easy to send; it’s cheap and lucrative. But where and how does a spammer get millions of valid e-mail addresses to spam? The basic answer is websites. There are tens of millions of Web sites, and spammers create search engines that spider the Web specifically looking for the telltale "@" sign that indicates an e-mail address. These programs are called **spambots** (Brain). Other spammers create sites specifically to attract e-mail addresses. We are more than familiar with the "Win $1 million!!!” ad or website. Just type your e-mail address and you are now part of a large spammer database. Some sites create "opt-in" e-mail lists by asking, "Would you like to receive e-mail newsletters from our partners?". Probably the most common source of e-mail addresses, is a "dictionary" search of the email servers of large e-mail hosting companies like Hotmail, Gmail or Yahoo!. A dictionary attack utilizes software that opens a connection to the targetail server and then rapidly submits millions of random e-mail addresses. Many of these addresses have slight variations, such as "jdoe1abc@hotmail.com" and "jdoe2def@hotmail.com." The software then records which addresses are "live," and adds those addresses to the spammer's list. These lists are typically resold to many other spammers (Delio).

There are a couple of ways to protect your email address. You can modify the address on your site in such a way that they are invalid but easily fixed by human users. Or just simply hide it on your pages so that spambots can’t find them. You can modify part of your address by inserting invalid random text. For example,

carol@**REMOVETHIS**example.com  
carol**DELETEBEFORESENDING**@example.com  
**Z**arol@example.**Z**om (replace **Z**s with **C**s)

Spambots will still harvest your address, but if spammers send something it will bounce. The only problem is that legitimate users may not be able to send messages to you because they don’t get your security technique. On the other hand, in order to hide your address and avoid your email to be harvest you eliminate the @ from the address[[1]](#footnote-1), since spambots find addresses by searching for text patterns that look like emails, specifically @ signs. Example:

carol**-at-**example.com  
carol**(at)**example.com  
carol **AT** example **DOT** com

This approach, although successful, still may cause difficulties for valid users (Livingston). So, let’s to identify visitors that are spambots and either not show email addresses to them or completely deny access to your site. We can do so by checking the referrer string or cookie acceptance. A referrer string is where a visitor linked from. Usually, spambots don’t set the referrer string or they link it to a third party’s URL. Also, they don’t deal with cookies. These might look like fairly advanced thing to do. The other option is to use a honeypot. A honeypot is a security technique that works very much like a trap for intruders, hackers and in this case for spammers. Usually, when we talk about honeypots we refer to network honeypot, which are targeted to identify hackers and system vulnerabilities. Project Honey Pot is the first distributed system for identifying spammers and the spambots they use to scrape addresses from your website (Unspam Technologies, Inc.). This project built in PHP5, some Perl and C is free of cost and provides a great service, supporting websites running PHP, Perl, ASP, JSP and Cold Fusion. It helps you identify spammers and lead them to incorrect data. Once identified Project Honey Pot gathers a list of the IP addresses of known spambots to assist webmaster (Unspam Technologies, Inc.).

Your website is your property and you are allowed to restrict who is allowed to visit it. You should consider including a no-email-collection metatag at the <head> are at the top of the pages of your site.

<meta name="no-email-collection" content="[link to your terms]" />

You should also consider participating in Project Honey Pot. Eventually, as the project gets bigger spambots will attempt to avoid participating websites.

1. A more sophisticated version of hiding your address, which still allows human users to see the addresses without any apparent munging, involves using ASCII character codes. ASCII character codes are like machine language for representing characters on a web page. For example, if you want to represent an @ you can either use the character itself, or you can use it's ASCII character code: **&#64;** (ampersand number-sign six four semi-colon). [↑](#footnote-ref-1)