Security Analysis on Craigslist (December 2009)

Bryan Lengle, Mathew Sam, Jiyan Lam, Adrian Lee, *University of British Columbia*

***Abstract*— Craigslist is a popular website that provides local classifieds and forums for jobs, housing, sale and services. Since the website was started, it has been targeted for phishing scams and email harvesting. This paper will analyze the security of Craigslist, and provide countermeasures to resolve these security issues.**

***Index Terms*—Web Security, Web Crawlers, Craigslist, SQL injection**

I. Introduction

Craigslist was started in 1995 originally as a simple email distribution network for advertising local events. The following year it was deployed as a web service. Since then the services offered have expanded and follow a generic paradigm of advertising the sale or willing purchase of goods and services. It is worth noting that in 2004 eBay purchased a 25% stake in the company. These two services share the mechanism of connecting buyers with sellers but employ completely different policies. Craigslist still makes the majority of its revenue from posting paid job advertisements.

The Craigslist site is first divided by region; both country and city. This encourages local exchanges which increases safety from fraud. The service is further categorized into sections such as: for sale, job postings, personals, housing, services, etc. Within these categories, anyone can read and post new postings. Posting are valid for 30 days before being removed from the site. A post consists of information on the item or service to be sold or wanted as well as some form of contact information for readers to reach the poster.

Since the web service was started, it has been plagued by a series of security issues including mass phishing scams and email harvesting. The amount of help literature on scams throughout the site is evidence that there are clearly security issues with the system, however there has been very little change in the system in the past few years.

Our objective is to present some of the security issues associated with Craigslist. To this aim, we will implement, 1) a crawler to harvest email addresses and phone numbers from Craigslist, 2) a spammer that will send anonymous emails to the addresses that were harvested with the crawler and 3) a flagger that will help boosting a post, and flag posts as spam even though they are not.

In this paper, we will discuss in details of some of the potentials flaws with Craigslist, and explain how the crawler, spammer and flagger work. Finally, we will present the results of the crawler, spammer and flagger that we have evaluated. We will also document some of the current and suggested countermeasures that could avoid the security issues that we have outlined.

Manuscript received December 07, 2009
B. Lengle (b.lengle0@gmail.com)
M. Sam(madmanmathew@gmail.com)
J. Lam(jiyanlam@interchange.ubc.ca)
A. Lee(a.ysl@msn.com)

II. Possible risks

A. Assets at Risk

The assets at risk can be separated into different categories, depending on the situations associated with the risk

- *Low Risk*
- *Medium Risk*
- *Medium High Risk*
- *High Risk*

A low risk situation would be a Craigslist account being hacked, or the Craigslist user leaves the webpage on their computer and someone else accesses it. Assets at risk include email addresses, the Craigslist account, items that the Craigslist user is interested or trying to sell and personal information.

Medium risk situations would consist of a Craigslist user being scammed or being defrauded. Assets at risk include money, home addresses, credit card information and banking information.

Next, a medium high situation is one where a malicious user replies to a post via email and provides t he recipient with a link to a virus. In this case, assets at risk could range from personal information on a computer, or the whole computer itself, as hackers are able to take total control of a computer.

Finally, consider a situation where a Craigslist user is meeting in person with someone he/she met online. In this high risk situation, the reason for meeting with a stranger could include finalizing a transaction or personal interests. Assets at risk could potentially be a life.

B. Classes of Threats

- Disclosure

  Snooping- emails and phone numbers can be accessed by a crawler

- Deception
  Spoofing- using an email spammer, one can send their own ad to everyone
- Disruption/Usurpation
  Denial of Service- improper flagging can cause disruption by either making a post high rated, or one removed

C. CIA

Confidentiality - is reduced when a crawler is used to find personal data.
Integrity - is reduced if someone uses spamming to send Craigslist emails across the internet, making Craigslist less credible
Availability - is reduced if one uses an auto flagger to remove posts on Craigslist.

III. Security Issues

A. Login- Craigslist's login system will automatically need a user to verify the login via a recaptcha after five to six failed attempts. However, to bypass the recaptcha, one can use another login ID and begin to brute force the password for another five to six attempts. Also, to use the forums for Craigslist, one must create a handle for that account. The handle can be used as a login ID and is visible to all other users on Craigslist. Therefore, setting up a web crawler to harvest the handles and implementing them with the login brute force strategy, is idealistic. Finally, one can take these handles and abuse the "reset password" function. Emails are sent to the user's personal email account and with respect to how many times one sends that email, it can be a nuisance.

B. Least Privilege- Access to Craigslist is open to the public. The only time one needs to sign in, is to create a posting. All the other features of Craigslist can be used anonymously. This becomes an issue because the more important functions, for example, flagging and replying posts, can be used maliciously.

C. Email- To reply to a post, one is given a reference email address. This Craigslist reference email address will automatically forward any emails to the actual user's personal email address. By using relay emails, Craigslist helps ensures privacy for posters. These reference emails have the following composition:

[category]-[string]-[posting ID]@Craigslist.org

This generic format applies to all reference emails. Category is usually a four letter representation of the actual section the post is located, for example: PERS, SALE, SERV and JOB. Next, string is a random sting consisting of five lower/uppercase letters and numbers. The posting ID is simply a string of 10 numbers, easily found in the post itself. This system can be exploited by randomly generating emails for spam.

IV. Crawling

A web crawler is an application that browses the Internet automatically. We have built a web crawler, and are using it to crawl Craigslist and harvest sensitive information. How it works is it makes URL requests and receives back HTML. The HTML is then parsed for useful data and links such as specific anchor tags to crawl deeper into another page. We developed a Craigslist web crawler tool that will take three inputs: location, category and number of pages to crawl. Craigslist is laid out in a hierarchy: location -> category -> post, given the location and category we can crawl through a desired number of posts. The posts are ordered by date so we crawl from the earliest to the latest date. When the tool is run, it returns a list of email addresses and phone numbers; it also returns some stats on the frequency of finding emails and numbers.

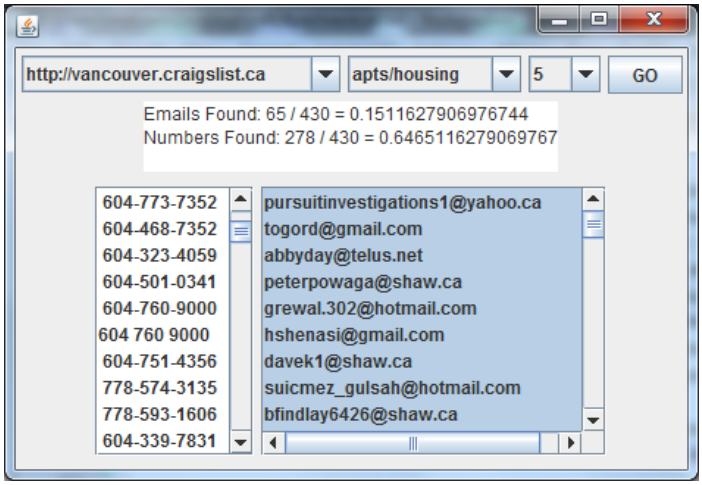

Figure 1: Craigslist Crawler

The assets that are available in using our crawler are emails and phone numbers as well as Craigslist emails. A Craigslist email is a relay email address that expires after 30 days, you can still email these address but we suspect they have a heavy spam filter. We managed to successfully email these emails and receive responses, which will be explained later in the report. One may say that these assets are not very valuable, but we disagree. The interesting part of our tool is that it categorizes the phone numbers and email addresses into location and category. This can be extremely useful for advertisement and marketing. People pay huge amounts of money on advertisement to get a small portion of customer's responses. With this simple tool you could retrieve hundreds of emails and numbers in your area for people interested in your product. Say you were a realtor, you could crawl Vancouver, housing wanted and five pages and you could get up to two hundred phone numbers and seventy five emails. Not to mention, there are around five hundred Craigslist emails for people looking for housing.

Using our tool we were able to get some concrete statistics on people posting their emails and phone numbers based on the category they chose.

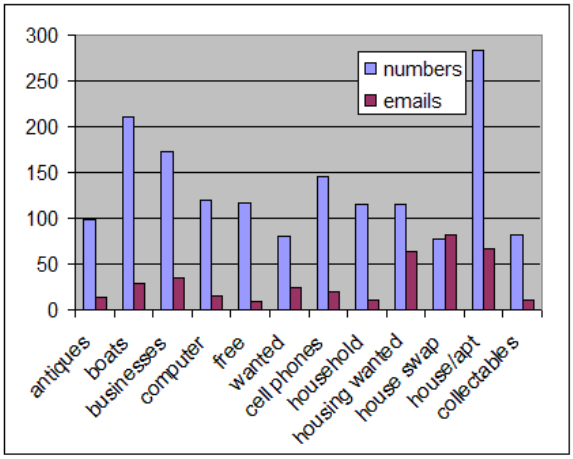

Figure 2: Data Found per Category

We collected data on twelve categories from the Vancouver area, such categorized as antiques, boats, free, housing etc. We found that on average 25% of people post their telephone number and 5.9% post their email addresses. The highest percentage for numbers was the House/Apt section at 56%. We believe this is because most people are more desperate to find or sell houses or apartments. The highest percentage for emails was house swap, at 16%. It is actually more than the numbers found with house wanted, at 15.4%.

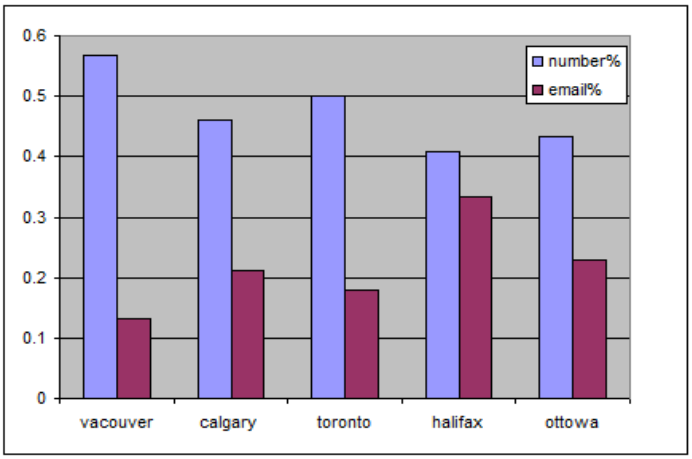

Figure 3: Data Found per Region

We also collected data on five cities with the housing/apt category. Each city data was pretty similar across Canada 40%-56% numbers and 13%-33% emails

Craigslist lets you mask you r email with a Craigslist email relay that will expires in 30days. So people have the option to not disclose their email. The feature is actually mandatory but you can still openly post your email in your post even though a much safer option is provided.

Craigslist does track IP addresses, if you crawl too much you will be blocked, but your IP is easily changed thought proxies and virtual private networks. You can crawl about 2500 posts before your IP becomes blocked

V. Email Spammer

Craigslist implied a "post a friend" feature; basically the idea is if you see a post you think a friend may like, you can forward it to them. All you have to do is fill in two emails, the receiver and the sender. Note that, these emails are not verified. What we did was create a small tool repeatedly filling out an html form to exploit this feature. Given a sender emails, a location, unique posting id and a list of receiver emails, you can forward that post to all the users in the provided list. This can be very useful to spam your own post to other people. It's essentially free advertisement for your own Craigslist posts. The email is forwarded from the Craigslist mail server, so it's as if you're spamming on Craigslist's behalf. You can also use our first tool, the crawler, to get a huge list of emails to spam. If you got a list from the houses wanted category, you could forward your own house for sale post to all those emails of people who posted in the housing wanted category.

As far as our testing, Craigslist does little to stop this. You can also put yourself on the non forward list, but by default no one's email is on it. As a counter measure they could require login to do this, and have a max of 5 forwards per day. They also have a number of forwards per IP address per hour, but you can easily get around it again with proxies and virtual private networks.

VI. Flagging

One of the vulnerabilities in Craigslist is lack of security around the flagging mechanism used to determine whether or not a post is useful or spam. Because the system is designed to be open to the public as well as moderated by the public, everyone has the ability to flag posts. Post can be flagged as: miscategorized, prohibited, over post/spam or best of Craigslist. Craigslist bases the system on the assumption that you get a rough approximation of the true status of a post (good or spam) given an average over a set of votes. This assumption; however, depends on the assumption that you can trust the voters.

The flagging system can be misused to do one of two things. One can flag a post as best of Craigslist even though it is not. This option can help boost a post, resulting in postings which are not necessarily helpful, residing in the best of' listings section. This removes a spot in the listings for a post which is actually helpful, and it generates undeserved advertising for the poster. This vulnerability could be exploited by an advertising company. The company could post an add and flag themselves as best of Craigslist enough times that the system puts them in the 'best of' listings. The other misuse of the system is slightly more malicious. One can flag posts as spam even though they are not. This results in posts being removed from the system which are legitimate. This vulnerability could be exploited by any user which wishes to remove competition from the system. Simply flagging any competitive posts enough times will have them removed; only leaving the user's post visible to visitors to the site. Keep in mind that advertising is one of the major sources of revenue for major companies like Google and even Craigslist itself. The asset value of exploiting Craigslist for free advertising is

higher than you might expect and the flagging mechanism lends itself to these kinds of exploits.

In order to appreciate the magnitude of the vulnerability in the system, we designed a tool for use in auto flagging posts on Craigslist. We chose to exploit the system in order to negatively flag posts in order to get them automatically removed. As mentioned before, the flagging mechanism does not employ any external anti spam techniques, so anyone including a machine can simply perform an HTTP post operation to the form dedicated for flagging. This form requires two inputs: the post id which is the unique numeric identifier attached to each post in the system and a flag code. Each category of flag has its own numeric code value i.e. flagging as spam has a flag code of 15 and flagging as best of Craigslist has a code of 9. Once the form is submitted to the server, the response is a webpage which contains the words "Thanks for flagging". This pattern in the response can be checked by a program in order to ensure that the flag was submitted. The final tool created was based on the Craigslist crawler previously described. The program takes as input a region and category of interest as well as a user entered string to search for. The program begins crawling the site and when it reaches a post with text matching the input string it submits a flag form using the posting id and the flag code for a spam post. The output of the application is a list of posts which were flagged as well as statistics of post flagged versus posts crawled.

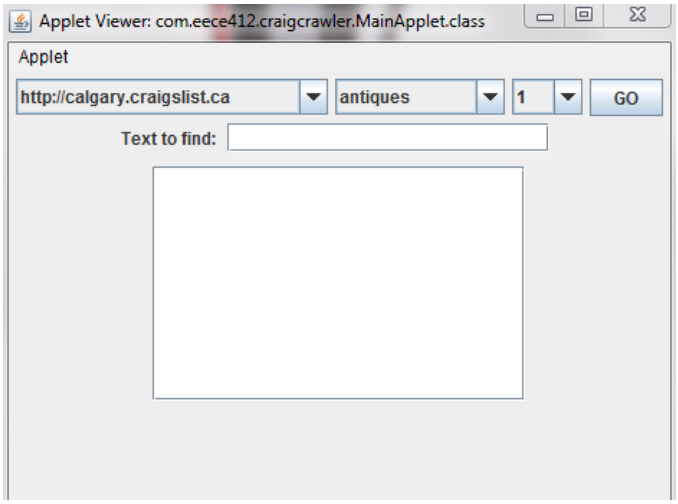

Figure 4: Craigslist Flagger

Craigslist does employ preventative measures in the form of IP tracking to ensure that a computer from a single IP cannot mass vote. However, this mechanism does not prevent a mass group of computers from skewing the integrity of the votes. In order to circumvent this mechanism, our program can route itself through proxy servers to hide its IP address. This method works well, but it proved difficult to find a large list of relatively reliable proxy servers in time for the submission of this paper. The program was tested using up to 15 proxy servers to flag a single post which we had created. The post was flagged as spam once using each proxy. Despite the use of proxies, Craigslist proved resistant to this method. We hypothesize that the system relies on an algorithm on the back-end which only removes posts under certain conditions. These conditions may depend on age of the post as well as the type of user flagging; logged in versus anonymous user. In either case, a post flagged 15 times, each from a different IP is much more likely to be automatically removed than a post which is not. As a result, the integrity of the system is compromised using our method although not to the magnitude which a malicious user would have hoped. This form of preventing malicious activity by using an unknown algorithm breaks one of the principles of designing secure systems. A system should not rely on the secrecy of design in order to maintain security. With the opportunity to investigate and test further in order to determine the exact conditions required to automatically remove a post, our auto flagging program would become a very malicious tool able to remove any desired post.

VII. Craigslist Experiment

To analyze the trust levels of Craigslist users, we created a fake email address and expressed interest to a post in a very plain and general way. The email contents are as follows:

Subject: Craigslist Posting
From: Gerald Dickson (g.dickson@live.ca)
Sent: November 26, 2009 4:44:03 PM
To: sale-ncaqu-1482832317@Craigslist.org
Hi,
I am writing in regards to your posting on Craigslist.
I'm interested in seeing if the price can still be negotiated!!
Hope to hear from you soon,

Sincerely,
Gerald

It seems that as long as anyone expresses interest in any way to a product, the owner will respond no matter how general and weird the contents of the email could be. However, if we were to ask for credit card information in the preliminary email, people are more aware of the consequences. Statistics concluded to be twenty emails sent, with fourteen respondents.

VIII. Cross Site Java Scripting/SQL injection

We tried to use cross site java scripting and SQL injection but they were not possible. Craigslist is a well established site and we are sure they have gone to all the measures to remove these sorts of common attacks. We tried to input a standard "<script>alert('test')</script>" into a lot of the input fields, but there was no success.

Craigslist either removes the brackets or any tag period, so the initial script would be reduced to "alert('test'). Also, we considered if Craigslist did the validation client side but that is not the case. In addition, we also tried some standard SQL injections but again no results were found. PHP provides methods to remove SQL and Script tags.

IX. Countermeasures

Although the security issues associated with Craigslist are not extremely major, there are still methods to improve the overall functionality of the website. To solve the matter of the principle of least privilege, mandatory login for all functions associated with Craigslist should be enforced. It can be argued that one can still create fake accounts and be able to access these functions. However, a counterargument is that Craigslist can monitor account usage, and determine if that specific account is being used maliciously. For example, a certain number of accounts can be deliberately flagging other posts. Craigslist can observe those accounts and shut them down at their discretion.

To resolve the problem of programs crawling pages, saving information, and creating email spam, recaptcha can be implemented as a security feature. Currently, recaptcha is only needed for the personal postings; however, every post should adapt this feature to maximize the security. Also, recaptcha is currently used after five to six failed login attempts into an account. As stated above in the security issues section, recaptcha can be bypassed by using another login ID. Therefore, exponential back off, or locking an account and notifying the owner could be possible ideas to improve the login system.

Moreover, if Craigslist separates the login ID and the forum ID, then the handle for the forum will be unable to be used as a login ID. Resulting in the current handles safe from being harvested and used in a brute force system. Next, Craigslist can change the reference email formats to a randomly generated string consisting of upper/lower case letters, numbers and/or symbols. This would prevent hackers and scammers from specifically targeting a certain category.

As a countermeasure to minimize the number of people openly posting their phone number they could implement the system as stated below.

The poster would add his/her phone number and it would remain anonymous and not openly visible on a post, if someone would like to get the phone number of the poster, they would have to request it. To request it, they would have to input their phone number and Craigslist would send them a text message back with the posters number. This completely removes potential crawlers and also can maintain a list of people who requested phone numbers. Phone numbers as authentication is a very good mechanism implemented these days. It is starting to be used much more often, where a phone number is essentially identification and the name and number address is stored with it.

Finally, Craigslist has developed a new system for added security called Phone Verified Accounts, PVA. Phone verified accounts were initially not a requirement for posting ads in Craigslist. With rapidly increasing traffic to the site, junk and spam advertisements became common, so the need for PVA was introduced. PVA is a good form of security because the user needs to verify themselves and is only given forty eight hours to post an ad after proper account verification. Also, if the user fails to post within this time frame, the account becomes blacklisted. Currently, these accounts are only being introduced in North America; however, it is only matter of time before this security feature will be introduced to all of Craigslist.

X. Conclusion

After the initial analysis of the Craigslist, it became clear that, although simple, the system does provide opportunities for malicious users to abuse certain features. The asset value in these cases was determined to be high enough to warrant sufficient protective measures. At this point Craigslist's position on the matter of security seems to be to provide as many warning notices and much internet safety literature as possible in an effort to push the responsibility onto the user. This can be effective especially with a system as simple in design as Craigslist; however, it assumes that the users read and understand these warnings. As the popularity of Craigslist increases worldwide, experiments like the ones we performed for this project become increasingly pertinent. These types of experiments are important to prove not only that the vulnerabilities exist and can be exploited fairly easily, but that the users who use the system are not taking the issue of computer security particularly seriously or simply do not understand the risk. Just as companies sometimes enforce a certain level of password security used by their employees, sometimes it is best to remove some of the responsibility from the user and place it on the policies of the system itself. We believe that the countermeasures mentioned earlier are a good start towards this goal.

We believe that we have done a more formal security analysis of Craigslist. Most of the existing information deals with scams and fake postings. Our crawling tool is unique such that it intelligently harvests information by location and category. In combination with our email spammer, it can be considered different than any of the existing software. We also found some useful statistics on email and phone number post rates and the reply rate from anonymous people with ambiguous email. In the process, we also found a lot of security features which Craigslist did well to prevent users from exploiting their site further.

References

[1] M. Stamp, *Information Security Principles and Practices*. John Wiley and Sons, 2006.

[2] M. Young, *The Techincal Writers Handbook.* Ross J. Anderson, 2008.

[3] Wikipedia. Craigslist. Internet. http://en.wikipedia.org/wiki/craigslist, [Dec. 05, 2009]

[4] EECE 412, "Introduction to Computer Security," http://courses.ece.ubc.ca/412/sessions/EECE_412-02-introduction-printable.pdf , [Dec. 05, 2009]

[5] EECE 412, "Principles of Designing Secure Systems, " http://courses.ece.ubc.ca/412/sessions/EECE_412-08-design_principles-printable.pdf , [Dec. 05, 2009]