Exploring Vulnerabilities of Threat to E-Commerce with Popularity of Search Engine
Shruti Kohli
kohli.shruti@gmail.com

ABSTRACT
Endeavor of this paper is to highlight the threat posed by the emerging online activities. This paper explores the possibility of information leakage indirectly through Search Engine index. It also explores the possible ways through which information over web is being misused and corrupted. Search Engine advertising, a growing online industry, has become a pray of many malpractices that want to monetize at the cost of advertisers and Search Engines. Click fraud has emerged as a big threat for the online advertisers as well as Search Engine. This paper discusses click fraud, cloaking and other threat to the online business industry. An experimental has been conducted to analyze how cloaking can be used as Spammer’s tool to misguide Search Engine and hamper its functioning which ultimately affects e-commerce.

KEYWORDS
Security threat, Search Engine, Online Business Industry

INTRODUCTION
The popularity of Internet has molded the behavior of many information seekers and consumers. Online Business Industry has emerged as one of the most revenue generating industry. This has created a race to expand business online. As any business booms there evolves hackers that want to profitize by ethical/unethical activities. The growing online market has given way to many malpractices that have posed big security threat to online business men, web publishers and Search Engine.

In the last decade, there has been a tremendous surge of interest in Search Engines among Internet users. Fueling this trend is the fact that Search Engines have become popular both as information-seeking vehicles and as an online advertising media. According to a study done by the Internet and Mobile Association of India (IAMAI), there are about 65 million people who use Search Engines in India and the current size of market for Search Engine advertising is $50 million. This market is expected to grow over the next five years [1]. The success of Search Engine lies in its unique features like free access, user friendly interface, and availability of multiple Search Engines [2]. The information on the web is enormous and Search Engines have provided easy way to explore the web. This has posed a threat to the information security as sometimes the information is misused by the hackers to steal some ones information, identity or some times for hacking the website itself.

Search Engine has been found to play an important role in online advertising market. Its popularity has given birth to a new form of online advertising market called Search Engine Advertising which plays an important role in generating huge revenue. According to researchers, online advertising market is playing an important role in online industry. Search Engine advertising programs have emerged as a new model for generating online revenue at null investment. Piper Jaffray [4] stated that the search industry has emerged as the most popular category among the eight broad categories in which the Web data can be divided.

Fig. 1. Search Engine Vs Other Web Categories

The online pay per click model has emerged as a popular revenue generating tool in online industry. The pay per click strategy had been very popular among the web publishers and advertisers but with this popularity the instance of exploitation and revenue loss has increased. Many malpractices have emerged that may result in great revenue loss. Click fraud has emerged as increasing plague to online industry.

In section [2] the security threat posed by Search Engine has been discussed in detail. Section [3] explores the vulnerabilities of Search Engine industry that has emerged as a security threat to the whole online business industry. The cases of revenue loss due to click fraud have been discussed and possibilities to avoid click fraud have been explored. In section [4] and [5] the potential of cloaking as an online tool as well as an emerging threat has been explored.

Search Engine as a Security Threat
Search Engines simplify the process of locating information in cyberspace. They index a huge number of Web pages and
other resources. Statistics in Fig [1] shows Search Engine receives maximum traffic from the net visitors and act as an intermediate between the user and the destination website.

### Fig. 1. Traffic Statistics for various Search Engines

With the emergence of Search Engine as popular information retrieval tool web publishers try their best to get their WebPages indexed in the Search Engine index. Thus it is easy to trace information and misuse it for mal practices. Thus mechanism that provides this great search ability also presents a threat to many organizations. Hackers’ uses these engines to make anonymous attacks, find easy victims, and gain the knowledge necessary to mount a powerful attack against a network.

The possible threat posed by Search Engine can be categorized as follows [5]:

1. Information Leakage
2. Revenue loss

#### Information Leakage through Search Engine

Information leakage is the major problem that Search Engines can cause for an organization. Organizations are continually making resources web accessible. In many cases, these organizations only intend for internal users to access these resources. Many organizations do not password protect these resources. Instead, they rely on security through obscurity to protect them. Organizations will create web pages that they think have no links pointing to them or that are in some obscure directory that they think no one would look in. What they forget is that Search Engines will catalog and index log files and software just as readily as they catalog and index widget pages. A report like the one produced by WebTrends contains a lot of useful information. A typical report will contain [9]:

- Top 10 requested pages.
- Top 10 countries requesting pages.

#### Revenue Loss through Search Engine

The information loss can ultimately contribute to revenue loss for any organization or web master. In the case of a business, sensitive data will include any data that may give a competitive advantage to a competitor [6]. Business and marketing people will agree that dissemination of this kind of information to a competitor could help them to have a competitive advantage. Research shows that Search Engine has emerged as a popular shopping tool. According to a survey conducted by Jupiter Research 62 percent of Search Engine users click on a search result on the first search results page and 90 percent click on a result one of the first three pages. This race of getting good ranking has given birth to many malpractices to get the site ranked without concerning its quality or relevancy to the search query. The possible threat posed by Search Engine can be categorized as follows [5]:

1. Information Leakage
2. Revenue loss

### Factors contributing to the security threat using Search Engine

#### Hacker’s identity cannot be easily traced

By using Search Engine hackers can easily hide their identification. One reason so few hacking attempts get reported is that there are so many of them. Trace routing any hacker’s IP address to its source often ends at a hop completely unrelated to the hacker’s actual ISP or local network. This makes reporting the hacker to the upstream provider difficult. Search Engines are dangerous largely because users are careless. In the age of DSL and broadband cable accounts, users often keep their machines turned on and connected to the Internet for days. Most of them would be shocked to find that they consider hackers target their machines up to several times a minute. Most home-
machine hack attempts seek to make their targets zombies in a distributed denial-of-service attack. Search Engines make discovering candidate machines almost effortless. It is not possible to secure all channels against hackers trying to penetrate a vulnerable system. But Search Engines need not be wide-open channels that continue to help hackers find and penetrate weak systems.

b. Search Engine is more prone to hacking as compared to directories.

The term "Search Engine" is often used to refer to many different searching technologies. Search Engine can be broadly divided into two main categories depending on the way they work to index pages and display search results. In case of Search Engine robot also called a spider or crawler, reads the page then follows every link contained within that page. The read data is handed off to the indexer that builds a searchable list of terms that describe the page. Directories are maintained by humans. The robot's systematic following of every link it encounters is where the problem lies. A human writing a description of a site for a directory probably would not consider log files and CGI programs as being important to the description. A robot, on the other hand, will happily traverse into the log file directory, if it is accessible from another page. It will follow a link that points to some CGI program or database interface program. These URLs and the "contents" of the URL, even if they are just error messages generated by a CGI program that was called with bad inputs, will be sent back to the indexer. The indexer builds the database of keywords that a user searches against. Many Search Engines provide a powerful language for constructing search queries. Google allows a user to search not only by the common "X and Y but not Z", but also to specify if a search term must be contained in the URL. It also allows searches to be narrowed by domain names; anything from abc.xyz.com to just .com is a valid parameter [7].

c. Unawareness among webmasters/organizations

Many organizations do not realize that Search Engine robots and indexers behave in the above manner. More importantly, organizations do not realize the kinds of information they are making available and just how easy it is to find that information. If you use Google, for instance, and are not blocking cookies, the Search Engine likely has placed a cookie on your system that won't expire until 2038. That cookie lets Google track what you searched for, when you conducted the search, and which results you clicked. The cookie does not identify you by name, but it does identify you by your system information and IP address. This is what the U.S. government was after when it subpoenaed Google for search records of millions of random users to establish the need for a federal online pornography law. The company was fighting the subpoena as this article went to press, but AOL, MSN, and Yahoo have already given the government at least some of the kinds of data it wants. The case highlights the sensitivity of search records in general, and Google's in particular. The company's position at the top of the Search Engine means that its archives could contain years of detailed logs on what millions of users search for and where they surf. Fortunately, there are well-established ways to rid your PC of tracking cookies, either using your browser or one of many third-party anti-spyware and system cleanup utilities [7].

d. Vulnerable security policies implemented by Search Engines

Search Engine security policies are excellent tools for helping hackers to attack machines anonymously, search for easy targets or gather confidential data. Securing all channels against hackers trying to penetrate a vulnerable system isn't possible, but there is no reason for Search Engines to be wide-open channels that continue to help hackers find and penetrate weak systems. Because it is so easy to use a Search Engine to cloak an attack, search-engine-based hacker abuse has become a real threat that poses serious risks.

Click Fraud looms as Major Threat to Online Advertising Industry

Online advertisers in developed markets agree to pay hosting website each time an ad is clicked. With performance-based deals becoming dominant on the Internet, pay per click has emerged as one of the popular online revenue generating strategy. Many Internets advertising companies like Click2freemoney.com shares profits with their members. Such companies provide various options for clicking on ads like clicking on website links via e-mails that they send, by clicking on banners and text ads in their paid-to-click section, and by referring others to the website [10]. Major Search Engines like Google have their paid-placement strategies where they publish advertiser’s advertisement on the side of Search Engine as sponsored link and charge advertiser for each click made on the advertisement. They also place advertisements on web publisher’s website and pay website owners for each click. Because of its easy working and almost null investment it is gaining popularity among web publishers and online advertisers. Google pays Ad publishers 78.5 cents on the dollar as revenue share [6]. The revenue generated by some website using pay per click strategy like Google’s Adsense has been summarized in Table [1].

Table 1. Revenue generated by Web Publishers through Search Engine advertising strategies.

<table>
<thead>
<tr>
<th>Website</th>
<th>Revenue generated via Adsense (per month)</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="http://www.Freeweblayouts.net">www.Freeweblayouts.net</a></td>
<td>$100,000</td>
</tr>
<tr>
<td><a href="http://www.SeatGuru.com">www.SeatGuru.com</a></td>
<td>$10,000 to $20,000</td>
</tr>
<tr>
<td><a href="http://www.podcastdirectory.com">www.podcastdirectory.com</a></td>
<td>$30,000 to $40,000</td>
</tr>
</tbody>
</table>

‘Internet Cloaking’: A new Web security threat

Researchers have been struggling for combating web spam using Cloaking Technique. Cloaking can be described as a technique to deliver different contents to different visitors
based on user agent or IP address of the visitor. It is popular for manipulating search results by delivering different pages to Search Engine and browsers. It can be also addressed as a systematic theft through negative Search Engine optimization. Now, Cloaking has not evolved not only as a quality spam but also a security threat for the web users. Study shows that Cloaking is being used by terrorists to provide misleading information to intelligence agencies. According to Lance Cotrell, founder and chief scientist at Anonymizer of San Diego, terrorist organizations and other national enemies have launched bogus Web sites that mask their covert information or provide misleading information to users they identify as federal employees or agents. Among the risks of the terrorist cloaking practice are that the organizations can provide bogus passwords to covert meetings, populate wrong misleading information. Through such practices they can pinpoint federal intelligence agents who attend the meetings, making them vulnerable to being kidnapped or can become the unwitting carriers of false information.

Experimental Analysis of Cloaking
To analyze Cloaking author performed experimental analysis of Cloaking. Author made a simple script that displayed a webpage on basis of the user agent. If the user agent is Mozilla/5.0 (Google Bot) it gets the message “Googlebot will see this” else if the user agent is any browser or any other Search Engine it will receive a page with message “This is for the browser”. The script was uploaded on a website and the different versions for the user agent were checked using the User Agent Switcher feature available with Fire Fox browser. This feature helps to view the webpage as it will be visible to different user agents. The script is as follows:

```php
<?php
Suser_agent = $_SERVER['HTTP_USER_AGENT'];
//echo($user_agent);
//Sipadress = REMOTE_ADDR;
//Echo "Your IP is Sipadress!"
if ($user_agent=="Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html")
echo "Google Bot will see this";
else
echo "This is for the browser";
?>
```

Inference
By using the Cloaking technique spammers display a fully optimized webpage to Search Engine and display a page to user which will attract them. This helps them to get high ranks among the Search Engine results even if the quality of the content of webpage is lower then the other websites. The spammer gains as good ranking attracts visitors to his website. This hampers the accuracy and relevancy of Search Engine and may lead to customer dissatisfaction with Search Engine results.

Emergence of Cloaking as a Security Tool
The properties of cloaking that can be used to provide misleading information or hack websites can be used to counter attack intrusions to once private space. For years companies have tracked nosey competitors' through reverse IP address lookup sites like whois.com or by retrieving information from Search Engine results or through website’s code. With the development in Internet technologies it’s easy to collect information about IP address's domain, physical location and sometimes even contact information and clipboard contents of site traffic. Such data can then analyze for patterns such as traffic on websites, activities of the company promotional schemes etc. Masking such activity is gaining in popularity, particularly by using tools and services that run traffic through a different network[11]. IP based cloaking can be used efficiently for counterstriking competitor’s activities. IP-based cloaking, which configures a legitimate Web site to display inaccurate or incomplete information only when it is accessed from certain IP addresses. Similar technique called IP Spoofing can be used to direct certain IP addresses to fake Web sites containing false or misleading information. The tactic can be used to throw off rivals and provide them with misleading data. During the IT boom of the late 1990s, a Fortune 500 company set up its site so that anyone coming from a competitor's IP address was sent to a different home page -- one opening with a job offer. Among the more interesting tricks is dynamic spoofing, which culls select criteria from known IP addresses to alter or hide pages, individual links, ads or banners and even price and availability information. For instance, if one airline knows a competitor checks its site for fares daily, it can jack up the price only when a rival's IP address tries to access the site. This can cause the other airline to advertise higher, unattractive tickets. Thus this technique can be easily deployed to counter security threat through competitor’s intrusion[12].

Conclusion
With the emergence of Internet Online Industry has emerged as a highly revenue generating industry. Every business needs to be secured by thieves and hackers and its activities need to be monitored. Online Business Industry is expanding at a great speed specially the Online Advertising Industry. In this paper major security threats to the Online Industry have been analyzed and the impact of these threats on the industry has been summarized. There is a need to scrutinize and update the vulnerable policies of Search Engine. However, not all blame should fall on those who operate the Search Engines. Search Engines aren't responsible for the huge numbers of poorly configured and insecure machines all over the Internet even if the Search Engines do aid in identifying them. The web users and web publishers need to be aware of the possible security threats and should take proper measures to monitor and counter such attacks.
To counter click fraud the online advertising industry has established Click Measurement Working Group as part of an Interactive Advertising Bureau (IAB) with the participation of major search providers such as Google and Yahoo. The purpose of this group is to “provide the detailed definition of a “click” and the standard against which clicks are measured and counted including the identification of invalid clicks and/or fraudulent clicks.” However, research needs to be done to eliminate such threat completely from chapter of Online Business industry.

FUTURE SCOPE
Threat due to Search Engine popularity is great matter of concern to the Search Engine Advertising Industry. Authors are examining the dependency of Search Engine users and Web Publishers on Search Engine and extracting factors that are threat to the working of Search Engine Business Model. Attempt with this analysis is to find out the various kind of threats Search Engine / Web publishes are prone when they use Search Engine for their requirements. After determining the major threats authors will taken up spamming as the major threat and work on monitoring spamming activities specifically Cloaking. Combating such acts is still an area of much needed research.

REFERENCES
[1] Internet and Mobile Association of India (IMAI), http://www.iamai.in/section.php?id=1288&secid=16
[12] Anne Saita,”IP Cloaking becomes a business necessity”, http://searchsecurity.techtarget.com/originalContent/0,289142,sid14_gci1151253,00.html