Web Vulnerability Scanner (SPYDER)

Siddharth Jain, Siddharth Choudhary, Shubham Malviya, Prashant Gupta, Prof. Amit Khare
Computer Science Engineering, Acropolis Institute of Technology and Research, Indore

Abstract

In Today's world, website security is the most important feature of securing an organization and should be given higher priority. As we are seeing, every new day hackers are targeting on the informative websites and web-based applications like forms, sensitive areas like login pages, shopping carts, dynamic pages, etc. Insecure web applications cause uploading backdoors on the server which allows access to databases, website-hosted servers and also allows hackers to perform illegal activities using the host server like email spamming, proxies. A website that has been hacked can be used for criminal activities, while illegally using the website's bandwidth and making its owner liable for these unlawful acts. While developing the websites, so many developers/site owners forget to remove sensitive data from a website that is not supposed to be exposed to public users. These data consists of untested vulnerable forms, database backup, and site backup in compressed format. A hacker tries to search for such kind of data and tries to collect important information from it like login detail from that data.

Introduction

In the conditions of the Internet, SPYDER is a specialized software designed to crawl and browse the World Wide Web usually for the purpose of indexing Web pages in order to give them as search results for user search queries. The most popular SPYDER is Google bot, Google's crawl crawler, which helps ensure that relevant search results are returned.

SPYDERs are also known as Web crawlers, search, or simply bots.
A SPYDER is a program which is used to harvest information from the World Wide Web. It crawls with the web pages of the websites extracting information and indexing it for later use, usually for search engine results. The SPYDER visits its websites and web pages with the various links to and from the pages, so a page without a single link going to it will be difficult to index and they may be ranked really low on the search results page. And if there are a lot of links pointing to a certain page, this would mean that the page is popular and it would appear higher up on the search results.
Steps involved in Web crawling:

- The SPYDER finds the site and begins crawling its pages.
- The SPYDER used to index the words and contents of the site.
- SPYDER visits the links found on the website.

SPYDERs or WebCrawler are just programs and, as a result, they follow systematic rules set by the programmers. Websites owners can also get in on this by telling the SPYDER which parts of the site to index and which should not. This is done by creating a "robots.txt" file containing the SPYDER's instructions regarding on which parts to index and links to follow and which ones it should ignore. The most important SPYDERs out there are those owned by major search engines such as Google, Bing, and Yahoo, and those designed for data mining and research, but there are also some harmful SPYDERs labeled to find and collect emails for the user to sell to advertising companies or to find vulnerabilities in Web securities.

A vulnerability assessment is defined as the structure for defining, displaying, classifying, and prioritize injection attacks vulnerabilities in computer systems, applications, and the network infrastructures and providing the organization with the necessary knowledge, awareness, and risk background to understand the threats to its environment and react accordingly.

A vulnerability assessment is a process that aims to identify potential threats and the risks they pose typically involves the use of automated tools, such as network security scanners, the results are then listed in a vulnerability assessment report.

A penetration test, which is also known as a pen test, is a simulated cyber attack against your computer system to check for all the exploitable vulnerabilities. In the context of web application security, penetration testing is commonly used to add a web application firewall (WAF). Pen testing may include attempts to violate any number of application programs, (e.g., application protocol interface (APIs), frontend / backend servers) to detect vulnerabilities, such as unauthorized inputs that may be attacked by sqlinjection attack .

System

The architecture of the system is shown below how the system works we will see it in detail.

I. URL Crawling :

URL crawler is mostly used to crawl the URL's from the search engine. If we search any keyword using a search engine the crawler will find the number of web pages for that particular search. When we search any One keyword it can contain a number of web pages and each web page has its URL but when we click on that URL it will again a number of URLs for that one web page so we can say that it is been called recursively for one single search. So basically the URL Crawler is used to crawl the web pages for the particular search it will automatically crawl the URLs and will show limited URLs or web pages. when we search for any keyword the limit is given to 6-7 web pages that will appear after we search for a specific keyword.

A URL crawler, SPYDER, or search engine downloads and indexes content from all over the Internet. The purpose of such a bot is to learn how (almost) all the webpage on the web is about so that the information can be retrieved when it's needed. They are called "web crawlers" because crawling is a technical term for automatically accessing a website and obtaining data with the help of a software program.

These bots are almost always driven by search engines. By using a search algorithm for the data collected by web crawlers. Search engines can provide relevant links in accoradance to user search queries, generating the list of webpages visible after the user has typed a search on Google or Bing (or another search engine).

II. Search Engine :

A search engine is a website where users can search online content.To do this, users enter the particular keyword into the search field. The search engine then try to look through its index for relevant websites and displays them in the form of a list. The internal search engine algorithm determines that on which position a website will get in the search results. Google, Microsoft edge and Yahoo, Mozilla firefox are example of popular search engines. Unlike web guides, which are updated only by human editors, search engines also maintain real-time data by running an algorithm on a web crawler. The Online content that cannot be searched by a web search engine is generally described as the deep web.

III. Remote Sites :

A remote site is a website that is physically located a distance away from the user's site..When we apply search on search engine it will show us plenty of sites as our result. We get results from multiple sites and when we click on a particular URL from that URL we again get multiple URL's means the sites are recursively called for this the directory traversal is used in the remote site so that the number of sites on a particular search will reappear.

IV. Domain Reputation :

Domain reputation is mainly used to check the black-listed sites so that reputation for that site will check. There are various domains which are available for checking the reputation of sites different domains are listed in the system architecture. This will help you to validate the vulnerability of sites using different domains mainly the RBL's are used to check mail server's IP and it checks whether the server's IP is black-listed or not.

V. CMS Scan :

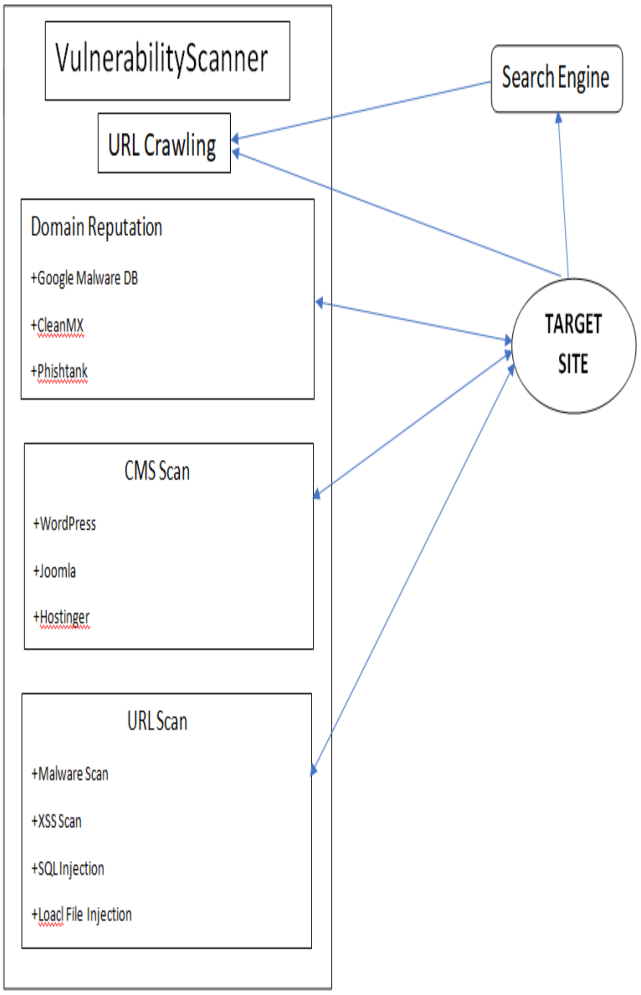
There are many tools which are available to design the website and due to this easily available tool, more changes are to cause Vulnerability. Tools like Joomla, WordPress are some examples of CMS. In this, the hackers used to check for the loopholes to detect Vulnerability. This tool is open source so it is easy for hackers to detect code also this kind of tool used a specific type of format and these formats are easily available for hackers and they detect Vulnerability much rapidly.

VI. URL Scan :

UrlScan is basically a security tool that is used to restrict the types of HTTP requests that Microsoft Internet Information Services (IIS) will process. By blocking specific HTTP requests, the UrlScan security tool helps to prevent the harmful requests from reaching the server.

We can do URL Scan with the help of two types of methods GET and POST. Using these two types of methods we check the Vulnerability for URL. For GET method there is a particular pattern is being define and for every URL these specific pattern is applied to check Vulnerability while in POST method no specific type of pattern is applied in this the whole URL is applied to check the result.

Using the Vulnerability Assessment and Penetration Testing technique gives an organization a more proper view of the threat facing its applications, making the the business to better protect its systems and data from malicious attacks. Vulnerabilities are found in applications from third-party vendors and internally made software, but most of these flaws are easily fixed once found. Using a Vulnerabilty assessment and peneteration testing provider enables IT security teams to focus on reducing critical vulnerability while the VAPT provider follows to discover and classify vulnerabilities. Rather, than only validating for Vulnerability the system is a product which can be applied to detect Vulnerability and to solve that Vulnerability the advance part in the system is that the system will be fully automated, we don't have to perform any type of manual work is required to do the system fully develop to be automated and anyone can understand it easily and use it. Due to automated patterns of system it will work fast and results will generated quickly.



ISSUES

The program has many features some of the things the system includes are:

1. DDos:

Distributed Network Attacks are known as the Distributed Denial of Service (DDoS). Such type of attacks takes the benefit of the specific capacity limits that apply to a network resource – like the infrastructure enables a company's website. This attack send's multiple requests to an attacked web resource – with the aim of exceeding the website's capacity to handle multiple requests… & it prevents the website from working properly.

2. Domain reputation in Google, Phishtank.

Check whether domain is black listed with the databases. Databases and organizations used to store IP addressess of machine and domains which are involved in malware, spamming, phishing activities.

3. Phishing Attacks

Phishing is a type of social engineering attack that is used to steal user data, including login credentials and credit card numbers. It is often discovered when the attacker, masquerading as a trusted entity, dupes a victim into opening the email, instant message, or text message. The recipient is then told to open malicious link with some false tricks, which can lead to the installation of malware, the freezing of the system as part of a ransomware attack or revealing of sensitive information.

4. Scan SQL Injections

SQL injection is a web-based security vulnerability. With the help of SQL an attacker is allowed to interfere with the queries that an application makes to its database. It basically allows an attacker/hacker to view data that they are not usually able to retrieve. This data belongs to other users, or any type of other data that the application itself is able to access. In many cases, an attacker can modify or delete this data which can cause persistent changes to the application's content or behavior. It can use poorly defined or poorly escaped SQL queries in classifying dynamic user input data.

5. Scan XSS - Cross Site Scripting

A Cross Site Scripting is a kind of vulnerability found in Web applications. It allows attackers to inject a client-side script into the Web pages viewed by other users and also detects on the WebPages and also scan for a GET and POST requests.

6. Scan Malware

A website disfigurement is an kind of attack on a website that changes visual appearance of a website or a webpage. It scans JavaScript code snippets against all the generic signatures and checks for JavaScript dangerous functions like eval, base64_decode, etc.

7. Detect and Scan CMS.

CMS scan is used for detecting WordPress, Joomla, and Bulletine. It Scan Themes, Plug-ins, unprotected admin area. User enumeration. Brut forcing for the simple password detection.

8. Scan for Directory Indexing

When a user types the request for a page on a website, a web server processes the request and then it searches the web document root directory for a default file name, and then sends this page to the other source. If the server can't find a page then it issues the directory listing and then sends output in HTML format to the user.

Definitions, Acronyms and Abbreviations

- **Vulnerability assessment**:

It is the process of identifying, identifying, classifying and prioritizing some of the vulnerabilities in computer systems, applications and network infrastructure and providing the organization along with the necessary knowledge, awareness and it also risk background to understand the possible threats to its environment and react appropriately.

- **Penetration testing**:

A penetration test(pen test), is a simulated cyber attack against your computer system to check for exploitable vulnerabilities. In this web application securities, penetration testing is mainly used to augment the web application firewall (WAF).

- **BurpSuite (PT)L**:

Burp Suite is a graphical tool for testing the security of the web applications. The tool was developed by Port Swigger web security and is written in Java programming language. The company has created the mobile application containing a similar tool which is compatible with iOS 8 & above versions.

- **Metasploit(PT):**

Metasploit is called an exploit development framework which facilitates penetration testing of the IT systems. It initially started off as the game and was taken over by Rapid 7 for the maintenance and further for the development.

- **Sublist3r** :

Sublist3r is a python specific tool. It is designed to identify the sub-domains of the websites using OSINT. It also helps the testers and bug hunters to collect and gather sub-domains for all the domains they are targeting.

- **Nmap** :

Network Mapper or Nmap is a open-source network scanner designed by the Gordon Lyon**.** Through this we can discover the hosts and the services on the computer network by sending packets and also analyzing the responses.

- **Dirsearch** :

Dirsearch is known by an other name as the simple command line tool designed to brute force directories and files in websites. The tool is supported on several operating systems including the Windows XP/7/8/10, GNU/Linux and MacOSX.

- **Acunetix :**

  Acunetix an automated web application security testing tool that audits your web applications by checking for the vulnerabilities like SQL Injection, Cross site scripting and other exploitable vulnerability too.

SPYDER TOOL

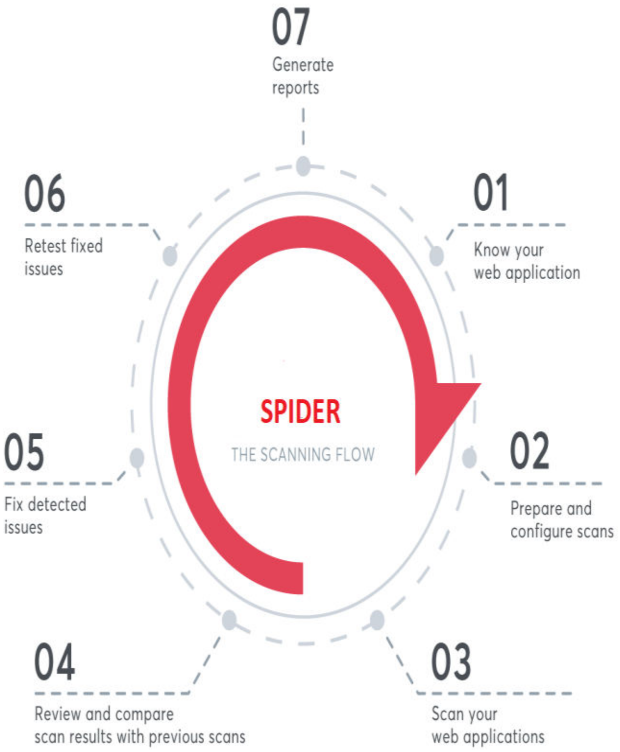
Web application penetration testing is the process of using penetration testing techniques on a web application to detect all of its vulnerabilities. This Project is an automated tool of various vulnerability tools like Burpsuite, Dirsearch, Nmap, Sublistr3r, Dirbuster, Acunetix, etc. The main purpose of this project is to minimize the time required to perform Vulnerability test. Performing vulnerability test using SPYDER is very convenient. It is identical to a penetration testing and aims to break into the web application using any penetration attacks or threats like SQL injection, HTTPs certification, Payment gateway bypass, XSS attacks, DOS attack, etc. To over the threads and make it make it more secure from this threads.

SCOPE OF THE PROJECT

The scopes of this project are:

- Protect from cyber attacks.
- Identify the possible vulnerability for a website.
- To ensure the security of website.
- SPYDER will consume less time and storage.
- Reveal lots of major vulnerabilities your security or development team never considered.

PROCESS FLOW DIAGRAM


07
Generate reports
01
Know your web application
02
Prepare and configure scans
03
Scan your web applications
04
Review and compare scan results with previous scans
05
Fix detected issues
06
Retest fixed issues
SPIDER
THE SCANNING FLOW

CONCLUSION

This is to conclude that the project and research paper that we undertook was worked upon with a sincere effort. Most of the requirements have been fulfilled up to the mark and the requirements which have been remaining, will be completed with a short extension. This project and research paper would definitely satisfy all the requirements of the users and would be beneficial for organizations to ensure security of their system.

ACKNOWLEDGEMENT

I am thankful to Prof. Amit Khare, Prof.Vandana Kate and Prof. KavitaNamdev for their valuable guidance and encouragement. I would also like to thank the Acropolis Institute of Technology and Research, Indore for providing the required facilities, Knowledge, content. At last, we must express our gratitude to all the Teaching and Non-teaching Staff members of Computer Science and Engineering Department who helped us for their valuable time, support, comments, suggestions and persuasion.

REFERENCES

[1] WebApplicationSecurity,ScannerEvaluation Criteria.WebApplicationSecurityConsortium.

[2] Acunetix Security. [Online]. Available:https://www.acunetix.com

[3] Common Vulnerabilities, and Exposures. [Online]. Available:http://cve.mitre.org

[4] WebApplicationSecurityStatistics.WebApplicationSecurityConsortium.[Online].Available:http://projects.Webappsec.org/Web-Application-Security-Statistics

[5] F. Maggi, W. K. Robertson, C. Krugel, and G. Vigna, "Protectionof a dynamic target: Addressing web application concept drift," in RAID, 2009, pp. 2140.

[6]http://portswigger.net/burp/documentation/scanner

[7]https://www.canada.ca/en/revenue-agency/services/about-canada-revenue-agency-cra/protection-your-privacy/xenon-web-crawling-initiative-privacy-impact-assessment-summary.html