Design and Implementation of Lightweight Web Asset Identification System

Zhengde Li

lizd@qfnu.edu.cn

Qufu Normal University School of Computer Science, Qufu Normal University, China Rizhao, Shandong, China

Abstract: This paper mainly describes the purpose of Web asset identification system research and development, design ideas, implementation process and so on. Web asset identification system is very popular in the attack and defense confrontation because the client does not need to rely on the environment, and can detect the target site through a variety of methods. With the Web Asset identification system, task scripts can be executed remotely without the need for real-time client execution. Asset detection can be done online. The Web asset identification system supports active and passive scanning. Web asset identification system is mainly divided into information collection module, target site module, file management module, virtual terminal module, user management module, export test module. Each module of the Web asset identification system has its specific meaning and function, but there are connections and interactions between each module. Web asset recognition system is different from the current mainstream asset recognition system, especially the passive scanning module of the Web asset recognition system adopts the form of crawler to crawl the website, which is not available in other Web asset recognition systems.

Keywords: Penetration Test, Web Asset identification System, B/S; Node. Js, Vue.

1 INTRODUCTION

On December 27, 2021, the net letter office issued the "difference" national informatization planning, "planning" digital as the core, puts forward seven development goals, deployed 10 major task, which, in terms of safety, stressing the need to pay equal attention to security and development, in order to realize the network space will markedly enhance its capability of governance and security as the goal, We will deepen the security concept of moving forward and taking precautions in advance, strengthen the mechanism for coordinating cyber security information, develop cyber security technologies and related products, and enhance the capability of independent cyber security defense. As China's network security industry promotion policies continue to increase, the product system is gradually improved, and ecological construction continues to advance, the network security industry continues to grow.

BDEDM 2023, January 06-08, Changsha, People's Republic of China

DOI 10.4108/eai.6-1-2023.2330297

The scale of China's cybersecurity industry reached 156.359 billion yuan in 2019, up 17.1% from 2018, according to the White Paper on China's Cybersecurity Industry (2020) released by the China Academy of Information and Communications Technology. In essence, cyberspace is an environment of interaction and collaboration between people. Network technology is only a tool to connect people together. The core of cyber security is still the issue of people. Under this cognition, it can be regarded as a way to solve new problems in the new era to look back on the wisdom of traditional Chinese culture and use the essence of traditional Chinese culture to grasp the essence of human beings from the perspective of human nature and understand the method of establishing interactive order between people. Cyber security is about you and me. One end of it is connected to Cathay And the other to Min 'an. The Internet has increasingly become a new space for people to study, work and live, a new platform for people to access public services, and an important driving force.

(1) For national development. For example, nowadays, big data, artificial intelligence and other technologies are widely used, and intelligent life has penetrated deeply into people's lives. Most government affairs have been transferred to the Internet, realizing "one-stop operation". Network security risks not only involve countries and enterprises, but also involve everyone.

(2) Research significance: With the rapid development of the Internet, Web applications are becoming more and more extensive. Shopping, banking, airline tickets, forums, tweets, etc. The popularization of Web application has brought great convenience to people's life, but Web security is easy to be ignored. The emergence of new Web technologies greatly speeds up the development of Web applications by enterprises and individuals, but at the same time, the vulnerabilities related to these technologies are also constantly involved in them. The research on Web security has reached the moment of thousands of catch a catch. Based on the above analysis, it is very necessary to study the principle of Web vulnerability and develop an automatic Web asset identification system. On the one hand, it can identify various vulnerabilities accurately, which greatly reduces the threat faced by the Web. On the other hand, it can send a large number of requests quickly and efficiently. Compared with manual, it is not only efficient, but also avoids human factors. On the premise of legality, collect asset fingerprints, analyze the risks of assets, mark the risks and levels of assets, notify asset vulnerabilities through correct channels, and solve problems before threats come.

(3) Research status: Literature 1 puts forward the idea of using crawler to identify assets, but the disadvantage is that web fingerprint library needs to be manually added, which increases the labor cost. Reference 2 proposed the associative component discovery method based on dictionary and the associative component discovery method based on the feature of component source file, but the disadvantage is that the feature code is less. Therefore, Web asset identification system solves the above problems.

(4) Research content: The main content of the research is how to identify assets. Traditional asset identification is to match HTTP response packets through the existing Web application fingerprint database. The author adopts the method of automatic collection to judge and enter the database, so as to quickly collect assets.

2 PREPARATION

Vue: Vue is a set of progressive frameworks for building user interfaces. Vue is designed to be applied layer by layer from the bottom up. Construct a data-driven Web framework through vue.js, which contains a variety of view components to realize data response and composition for developers to use, including a wide variety of component libraries for developers to choose from [3].

Node.js: Node.js is a JavaScript runtime environment based on Chrome V8 engine. Node.js uses an event-driven, non-blocking I/O model, which is not restricted by the client (browser), so that JS has the same operation permissions on files, network and operating system processes as the back end, and has little difference from the functions of Java, Python and other programming languages [3].

Python crawlers: The Requests library is a concise and simple third-party library for handling HTTP requests. Written in Python, the Library is based on URLlib and is Licensed under the E2 open source protocol. It is more convenient than URLlib, can save us a lot of work, fully meet the REQUIREMENTS of HTTP testing the biggest advantage is that the program writing process is closer to the normal URL access process. There are many ways to

open a web page, the most common of which are GET and POST. GET is used to access the page by directly entering the URL in the address bar of the browser. GET () is a GET request corresponding to HTTP for obtaining

HTML web pages, and data is obtained by making a request to the specified URL [4].

Database related technology: SQLite is a lightweight, open source, embedded relational database. It is an ACID-compliant relational database management system. It is a zero-configuration database that requires no configuration on the system. As an open source database, SQLite is widely used by major software companies, such as Firefox, iPhone, iPad and Android, etc [5].

With the rapid development and continuous update of the Internet, database has been more and more widely used. Database development up to now, has produced a variety of functions of different types, among them, the more widely used SQL Server database and MySQL database. MySQL database has the characteristics of small memory occupation, relatively low development cost, relatively fast running speed and can support a variety of computer programming languages, and its corresponding source code is free. Therefore, MySQL database is deeply trusted by the majority of small and medium-sized websites and corresponding enterprises [6].

This chapter explains the relevant technical basis involved in the paper. Only by understanding the basic knowledge of relevant technologies, can we continue to understand the design and principle of corresponding modules in the paper works. Web asset identification system adopts B/S architecture and MTV mode, so loose coupling relationship between components is maintained. M is primarily used to take care of business objects and database objects, T is responsible for how pages are presented to users, and V is responsible for business logic and calls M and T when appropriate.

3 SYSTEM ANALYSIS

System process analysis: The modules of the Web asset identification system can be divided into the following types: information collection module, target site module, file management module, virtual terminal module, user management module and export test module, as shown in Figure 1.

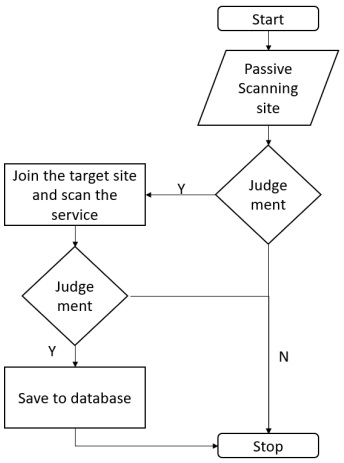

Figure 1 Overall system structure

Figure 1 Overall structure of the system System process analysis mainly analyzes information collection. Other module processes are short and are only analyzed in corresponding modules as shown in Figure 2.

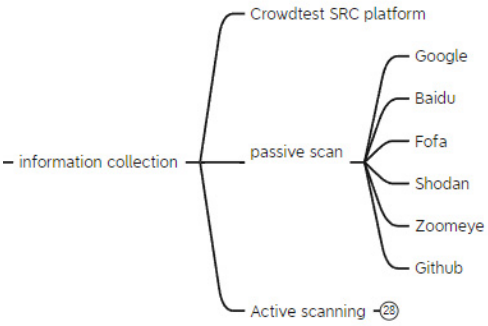

Figure 2 Flow diagram of Web asset identification system

(2) Passive scan module analysis: Passive scan module Analysis book submodule includes three parts, namely the SRC platform, passive scan and active scan, among which passive scan includes. See Figure 3

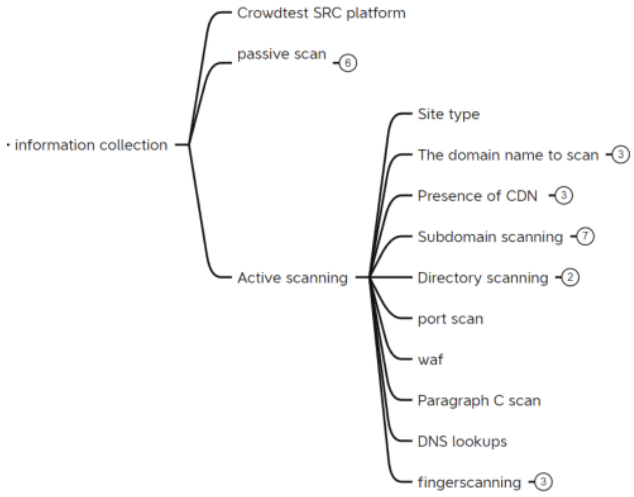

Figure 3 Mind map of passive scanning module

(3) Active scan module analysis: This sub-module includes three parts, namely, the SRC platform, passive scan and active scan, among which the active scan includes. See Figure 4

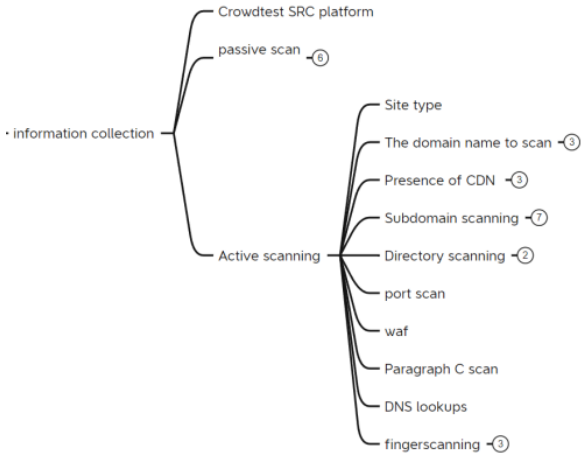

Figure 4 Mind map of active scan module

(4) Target site module analysis: This module includes four parts, which are adding sites, scanning sites, deleting sites, modifying sites. See Figure 5

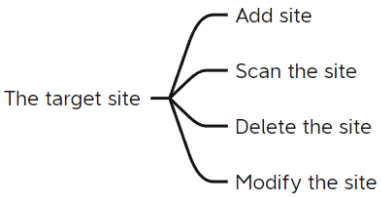

Figure 5 Flow chart of target site module

(5) Document management module analysis: File management module analysis This module includes six parts, respectively upload, download, copy, compression, delete, rename. See Figure 6

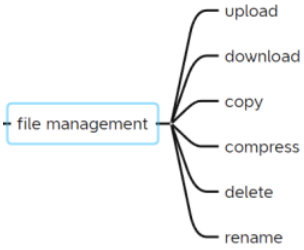

Figure 6 File management architecture diagram

(6) Analysis of virtual terminal module: Analysis of the virtual terminal module The virtual terminal module can remotely connect to the server on the Web and quickly enter the server to troubleshoot faults.

(7) Quick start module analysis: This module includes three parts, which are adding software path, generating registry download, running registry. See Figure 7

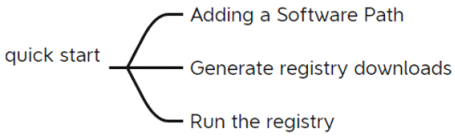

Figure 7 Quick start architecture diagram

4 DESIGN AND IMPLEMENTION OF THE SYSTEM

(1)The system display module using UI is Vue. Vue has a rich library of components that can quickly complete the layout of the front end.

The development environment needs to meet the requirements of the Windows system, and also need to install Vue 2.6.11, Node.js v14.17.6, element-UI 2.15.6 and so on.

The modules in the menu include "Information Collection", "Target site", "File Management", "Virtual Terminal", "Quick start", "User Management", "Export Test" and "Panel Setting".

Click the information collection module, a new menu will pop up. The menu of information collection module includes "SRC platform for mass testing", "Active Detection scan", "Passive detection scan", and "Proxy Settings" and "environment variables" are set on the panel.

(2) Design and implementation of active scanning module:

First, the user enters the domain name of the target site; Then, click the Search button; Then, the scanning script requests information from the target site according to the domain name. The request method can be either get or post. After getting the returned information from the target site, it stores it. If necessary, the next step is comparative analysis. At the same time, the returned information is displayed on the page.

(3) Design and implementation of passive scanning module: Design and Implementation of passive scanning module The implementation process of passive scanning module is as follows: The user enters the advanced syntax of the corresponding search engine and clicks the search button. The front end sends the value entered by the user to the back end through get request. The back end calls the Python crawler script to obtain the information

(4) Design and implementation of SRC platform for mass testing: Design and Implementation of SRC Platform for mass testing The implementation process of SRC platform module for mass testing is as follows:

Collect the mainstream emergency response center on the market on the Internet and store it in the database. When the user clicks the crawl button, the front end transmits the value to the back end through GET request to crawl the list of manufacturers and save it in the SQLlite database. When the user clicks the query, the back end calls the database to query the corresponding manufacturer information.

(5) The design and implementation of the target site: the design and realization of the target site target site principle mainly through the interaction of other modules or manually add, the user can choose the site for vulnerability scanning, add site by entering a name and a domain name to achieve the effect of added manually, click ok after sent via a get request to the backend interface call for automation.

(6) Design and implementation of document management: Design and implementation of file management File management has upload, download, new, copy, compression, delete, rename, search functions

(7) Design and implementation of virtual terminal module: The benefits of using a virtual terminal are as follows: In the presence of a jumping machine environment, if you have any open web services jump machine itself, that could be deployed on the springboard for virtual terminal, then don't via SSH or RDP access to jump machine, open the browser can directly in the form of web SSH remote access network device, by detecting the in some network firewalls don't allow SSH, But it is useful in environments that allow HTTP and HTTPS. In addition, SSH client software such as putty and Secure CRT is not required.

(8) Design and implementation of fast start: the design and implementation of a fast start principle of environment variable is submitted by the user in the absolute path to automatically

generate the corresponding registry files in the background, when users click the registry to download again to download the registry file, double click on the run in the computer system registry, finally can be run in the quick launch module local procedures.

The following table shows that the Web asset identification system is superior to current asset identification system tools in terms of user customization.

TABLE I. COMPARISON OF ADVANTAGES AND DISADVANTAGES OF WEB ASSET IDENTIFICATION SYSTEMS

Assets system	Architecture	CDN identification	SRC vendor list can be crawled
Goby	C/S	N	N
lighthouse	B/S	Y	Y
Web asset identification system	B/S	Y	Y

5 CONCLUSIONS

Web asset identification, Web vulnerability scanning, Web vulnerability utilization and other tools. They are all indispensable automatic and intelligent tools in red blue confrontation. Each tool has different advantages. For example, WebShell management tool is good at modifying the files of the target host and has high stability. Web asset identification system is a comprehensive tool, which can not only scan vulnerabilities but also collect assets. If WebShell management tools and EXP are integrated into the browser, it will be a powerful tool for individual combat. Web asset recognition system is the first successful project that the author summarizes and develops through practical experience. There are still many deficiencies in function points, such as WebShell management, EXP utilization, POC detection, etc. These function points are an indispensable part of red-blue confrontation. In the long run, WebShell management function can use the national secret to confuse the transmission data, so as to bypass the packet detection of security vendors, greatly improve the efficiency of security testing, and avoid the product interception of security vendors. EXP uses the collected Nday, 1day, 0day to conduct one-click GetShell, so as to quickly collect assets.

REFERENCES

[1] Zhang Chengyu. Based on Web crawler's design and implementation of automated testing system of Web components [D]. Beijing university of posts and telecommunications, 2021. The DOI: 10.26969 /, dc nki. Gbydu. 2021.001072.

[2] Zhou Shuangfei. Web fingerprint analysis [D]. Chongqing university of posts and telecommunications, 2020. The DOI: 10.27675 /, dc nki. Gcydx. 2020.000173.

[3] Hao Zeng, Haowen Huang, Liqiang Zhang & Sihui He.(2020). Design and implementation of learning communication Platform based on B/S. Journal of Enterprise Science and Technology (10),49-51.
[4] Feng YANru.(2021). Design and implementation of Web crawler system based on Python. Computer and Information Technology (06),47-50. doi:10.19414/j.cnki.1005-1228.2021.06.014.
[5] Zhu Chunyang.(2015).Python operation of SQLite database. Computer programming skills and maintenance (15), 65-66. Doi:10.16184/j.cnki.com.prg. 2015.15.027.
[6] Zhu Baoshan, Chen Guangpu, Li Pengcheng, Wang Shen. Modern Electronic Technique, 201,44(14):65-69.
[7] Luo Qiang, DUAN Mengjun, Wu Zhilin. Communications Technology, 201,54(09):2235-2241.
[8] Zhang Wei. Analysis of Web vulnerability risk scanning technology [J]. Network Security Technology and Application,2022(06):14-15.
[9] Rui Kunkun, RUan Jinjun. Development of Django WebSSH application integrated with edX platform [J]. Journal of yili normal university (natural science edition),2020,14(04):53-59.
[10] Wei guangxing. Design and implementation of file warehouse management system based on MVC [J]. Journal of changchun university,2018,28(12):34-38.