Design and Implementation of A Website Record Management System Supporting Automatic Verification in IPv6 Transition Phase

To cite this article: Jingjing Zhou et al 2020 J. Phys.: Conf. Ser. 1575 012033

View the article online for updates and enhancements.
Design and Implementation of A Website Record Management System Supporting Automatic Verification in IPv6 Transition Phase

Jingjing Zhou¹, Gengsheng Deng¹*, Aiying Fu¹ and Zhe Lu¹

Nanchang University Network Center 1, Nanchang 330031,
Email: dgs@ncu.edu.cn

Abstract. Website information registration and filing is the basis of website information security management. In view of the current situation of various websites and increasingly severe network security in Colleges and universities, in order to better manage the websites in Colleges and universities, this paper takes the network environment in the transition stage of IPv6 as an example to design and implement the website filing management system architecture supporting automatic verification. This paper analyzes the methods and shortcomings of the improvement of the traditional website filing management system in the domestic universities, and puts forward the automatic verification mechanism which can actively distinguish the content and status of the website, and actively realize the verification of the filing information by means of scanning detection, page collection, access control, etc., designs and introduces the process of the automatic verification in detail Lists the main code for page collection and reverse proxy services. The actual application shows that the filing management system supporting automatic verification can realize the automatic verification and recording of the information of the completed filing website, detect the website services existing in the managed network, and check out the sites that have not been record, greatly reducing the workload of the management personnel to check the website filing information, and improving the timeliness of the website filing management. Statistics show that IPv6 link access support has become a problem that must be considered by colleges and universities in the transition phase of IPv6. The architecture is based on open source software, and can be used in other universities or similar network environment with only a few configuration and code modification.

1. Introduction
There are many characteristics of university website, such as large quantity, various construction and operation modes. The construction of website provides essential information support for the University. But at the same time, there are a lot of hidden dangers in website security, such as frequent attacks, system information leakage, webpage content tampering, which not only damages the information resources of the University, but also causes bad reputation of the University Therefore, it is very important to strengthen the management of website information security. Website registration and filing is the starting point and basis of website security management. Registration and filing helps colleges and universities to master the number of websites, clarify the operation of each website, determine the main unit of website security responsibility, and sort out the business and management information of the website. [1] In practical work, there are often hundreds of websites managed by colleges and universities. Due to the lack of technical support measures, there are often cases such as incomplete filing websites, non cancellation of offline websites and untimely update of filing information. On the other hand, the current global network is in the process of transition from IPv4 to
IPv6. China has released the action plan for promoting the scale deployment of Internet Protocol version 6 (IPv6) \(^2\). As an important application of the Internet, the website service is also changing from pure IPv4. In the process of accelerating the transition from environment to environment supporting IPv6 \(^3\), however, due to the incompatibility between IPv4 protocol and IPv6 protocol, website filing management also faces more challenges.

In view of the wide variety of current websites and the increasingly severe situation of network security, in order to better manage the campus sites. Domestic colleges and universities also take some measures to improve the traditional website filing management system. Literature \(^4\) introduces that Xiamen University has developed a new filing system based on the original filing system. Based on the original filing function, the annual review mechanism and loophole management mechanism are introduced, and good results have been achieved in the implementation of the filing. Literature \(^5\) introduces the actual needs of the information security office of Beijing Jiaotong University, and develops a campus network security management service platform, which integrates heterogeneous data from different application systems, and analyzes, visualizes, generates reports, generates monthly reports and issues security notices for the integrated data. In addition, literature \(^6\) introduces that Anhui Provincial Communication administration improves the accuracy of website filing information through automatic verification based on communication methods, but only for the verification of contact information and operator registration information. All of the above improvement methods are aimed at the management and improvement of the record website information, failing to automatically discover the unrecorded website system, and the update of the record website information also relies more passively on the registration of the website administrator, and the filing management system lacks the mechanism of actively judging the content and status of the website.

The main content of this paper is to design and implement a website filing management system that can automatically verify the website filing in IPv6 Transition environment. Through scanning detection, page collection, access control and other ways, it can actively verify the filing information, so as to improve the accuracy and effectiveness of the website filing information.

2. Architecture Design of Website Record Management System Supporting Automatic Verification
Website information registration and filing is the basis of website information security management. Through information registration and filing, colleges and universities can understand the number of websites, make clear the operation of each website, determine the safety responsibility unit of the website, manage the business and management information of the website, etc.

Taking the network environment of colleges and universities as an example, this paper designs a website filing management system framework of automatic verification. In the transition stage of IPv6 \(^7\), there are three ways for colleges and universities to provide WEB services to the outside world: pure IPv4 service, IPv4/IPv6 dual stack service and IPv6 service. Therefore, the website filing management system designed needs to support the above three ways.

The registration of university website is usually initiated by the information management department of the University, which needs to record the domain name, IP, information of the responsible unit of website security (such as the name of the person in charge, contact information, etc.), business type, main data processed, service scope and other information of the website. When the website changes, such as new website, website cancellation, domain name change and contact person change, the website management unit shall report to the school information management department in a timely manner, but in reality, there are often cases such as incomplete website for filing, non cancellation of offline website for filing, and non timely update of filing information. Compared with the traditional website record management system, the system designed in this paper adds the support of automatic verification and IPv4 / IPv6 dual stack.

2.1. Introduction of Automatic Verification Technology
In this paper, automatic verification is mainly realized by three technologies: network scanning, web crawler and embedded script.

Port scanning technology: port scanning is to scan a section of ports or designated ports one by
From the results of the scan, you can know what services are available on a computer. The principle is that when a host requests to establish a connection to a port of a remote server, if the other party has this service, it will respond. If the other party does not install this service, even if you send a request to the corresponding port, the other party still has no response. In this paper, the common ports of IP in the network are scanned to record the ports of Web services.

Web crawler technology: the traditional crawler starts from one or several initial web page URLs and obtains the URLs on the initial web page. In the process of capturing the web page, it constantly extracts new URLs from the current page and puts them into the queue until it meets certain stop conditions of the system. This paper uses crawler technology to crawl the website information according to the URL and compare it with the filing information in the system.

JavaScript embedded script technology: through embedding a small amount of JavaScript code in the web page to realize the tracking and analysis of web page custom events and custom indicators, Google analysis and Baidu statistics all adopt this technology. This paper tracks and records the information of the website in real time through the JS embedded point of the webpage.

2.2. Design of System Architecture

According to the functional requirements of the website filing management system, the project adopts the B/S architecture, and the system architecture is shown in Figure 1. The system administrator and the website administrator access and use the filing management system through the browser. The system management business is completed at the server, and the server completes the interaction with the data layer. The website that completes the filing needs to be added with a buried code. When a user visits the website, the information such as the service status of the website will be automatically updated to the server, and checked with the website filing information in the filing system Proven.

![Figure 1. System architecture diagram](image)

2.3. Design of Energy Module

The system adopts the idea and method of structural design and modular design, and the overall function module diagram is shown in Figure 2, mainly including user management, system information management, basic information management, filing information management, abnormal site list management, filing website access control, access statistics, filing information query, reverse agent service, filing website information collection and other modules.
2.4. Design of Network Architecture

The network of the system is mainly divided into two parts: filing management and collection agent service. The collection agent service is deployed in the IPv4 / IPv6 dual stack network environment [11], and the filing management service and collection agent service need to be accessible to each other. The collection agent service needs to be able to access the record sites and potential networks. Ordinary users need to be able to access the reverse agent service and website service. In order to ensure that ordinary users cannot directly access the filing management service, only administrators are allowed to access. In case of resource shortage, the filing management and collection agent service can be deployed on the same machine, and the access of ordinary users to the filing management can be restricted by access policies.

Figure 3 is the schematic diagram of the system deployment network architecture designed in this paper. Among them, $S_{BA}$ is the filing management service, $S_{SR}$ is the reverse agent service, $S_{DB}$ is the database service, $S_{web1}$ is a record IPv4 site, $S_{web2}$ is a record IPv4 / IPv6 dual stack service site provided through the collection agent, $S_{web3}$ is a not record IPv4 / IPv6 dual stack website, and $S_{web4}$ is a not record IPv4 website.

2.5. Automatic Verification Process

Through the network scanner to actively detect the web services within the scope of the managed network; the web crawler crawls the web content to check with the currently record information; the embedded script is added to the web page, on the one hand, when a user visits the web site, the current information of the web page is automatically recorded in the web site filing system, on the other hand, to obtain the current filing status of the web site To show or control access. The specific process is as
follows:

Step 1: the administrator enters the managed network segment and domain name information;
Step 2: the scanner scans all the commonly used web service ports for entering network segments and domain names regularly, such as 80, 443, 8080, 8443, etc., and saves the open service IP and port information and other scanning results to the database;
Step 3: the crawler regularly crawls web service information according to the IP and port information saved in the database, including response time, response status, header information, title information, web page content, etc., and saves the crawled information to the database;
Step 4: the system regularly compares the web service information obtained by crawling with the information in the filing system. If it is found that the web service information does not exist in the list of websites that have been recorded, it will prompt the discovery of websites that have not been recorded. If it is found that the web service information is consistent with the filing, but the title information is inconsistent, and the status is abnormal, it will prompt the abnormality of the filing website;
Step 5: add the embedded script to the record website. When a user visits the record website, obtain the website information through the script, and update the website status information and statistical information in the database;
Step 6: the system regularly compares the website information obtained by the script with the information in the filing system. If the website information is found to be inconsistent with the filing title information, the status is abnormal, etc., it will prompt the filing website to be abnormal. If the website is found to have no visits in recent period or the visits are very low, it will prompt the website to be abnormal and zombie websites.

3. Implementation of Website Record Management System Based on Automatic Verification

Based on Linux platform, open-source software nginx, mysql, redis combine PHP and Lua language to realize the whole web service architecture. Among them, nginx is used to realize web service and reverse proxy service, MySQL is used as database, redis is used as cache database, PHP is used to develop record management service, Lua is used to realize collection statistics and reverse proxy service. Nginx is an open source web page and reverse proxy server, which can run in Linux, windows and other operating systems. It has the characteristics of high reliability, simple configuration and low resource consumption. Among them, the implementation of collection statistics and reverse proxy adopts NGX Lua module embedded in nginx. By writing Lua script, the data flow can be processed, and the function can be realized flexibly under the condition of high concurrent service ability. The following describes some key codes to realize automatic verification of website filing.

3.1. Key Configuration Of Collect Web Server Information
Collect the page information of the corresponding URL through the curl component on the collection agent server, and the obtained information includes content_type, http_code, header_size, request_size, FILETIME, redirect_count, total_time, namelookup_time, connect_time, download_content_length, http code and other information.
$obj_curl=new CurlComponent();
$obj_curl->set_value(FALSE);
$return = $obj_curl->multigetheader($data['url']);
for($i=0;$i<count($return);$i++){
    $arrInfo = $return[$i]['info'];
    $header  = $return[$i]['html'];
    ...
}  

3.2. Reference and Generation of JavaScript Embedded Script
At the bottom of the registered website source code, the JavaScript embedded script is referenced, and the modified script is quoted from the collection proxy server SSR. The nginx configuration on the collection proxy server SSR references the JavaScript embedded script. The specific script content is judged and output by Lua script. The JavaScript script can obtain the domain name, URL, title,
resolution, browser type, language, etc. of the page.

Reference code of embedded script of filing site:
<script src='http://beian.example.edu.cn/ba.js' type='text/javascript'></script>

The main configuration codes of nginx are as follows:

```nginx
location /ba.js {
    default_type 'text/javascript';
    add_header Expires "Fri, 01 Jan 1980 00:00:00 GMT";
    add_header Pragma "no-cache";
    add_header Cache-Control "no-cache, max-age=0, must-revalidate";
    content_by_lua_file /usr/local/openresty/nginx/conf/lua/content_ba_js.lua;
}
```

The main Lua scripts for real-time access to website information in the background are as follows:

```lua
local insert_sql = "insert into t_log_visit (msec,remote_addr,u_domain,u_url,u_title,u_referrer,u_sh,u_sw,u_cd,u_lang,http_user_agent,u_utrace,tjsj)
values(ngx.now(),ngx.var.remote_addr,ngx.var.arg_domain,ngx.var.arg_url,ngx.var.arg_title,ngx.var.arg_referrer,ngx.var.arg_sh,ngx.var.arg_sw, ngx.var.arg_cd, ngx.var.arg_lang, ngx.var.http_user_agent,uid,now())"
res, err, errno, sqlstate = db:query(insert_sql)
```

3.3. **Key Configuration of IPv6 Reverse Proxy Service Port Listening**

Configure the IPv6 port of RSR, and IPv6 users can visit www.example.edu.cn to access the real web service sweb2 through proxy server:

```nginx
server {
    listen [::]:80 default_server ipv6only=on;
    server_name www.example.edu.cn;
    location ~ / { 
        resolver8.8.8.8 ipv6=off;
        proxy_pass http://www.example.edu.cn;
        body_filter_by_lua_file conf/6to4.lua;
        ...
    }
}
```

4. **Implementation Cases and Effects**

The design of this paper is based on open source software. It can be used in other universities or similar network environment with only a few configuration and code modification. At present, hundreds of initial filing information have been recorded in the filing management of school websites in the author's unit. At present, more than 50 websites are monitored through the filing management system, and nearly 20000 visiting logs are recorded for each filing website every day. The websites that have not been record and the filing information has changed can be checked automatically, indicating the possible "zombie" websites. When a new network or domain name is added, only the corresponding address or domain name needs to be added in the configuration, the system will detect the possible website services in the network, and remind the management personnel to carry out record management or access control. In order to improve the security of the system, the management user needs to log in before using the system. The user who is not logged in can only query the filing status of the site individually.

Since the online trial operation in May 2019, more than 480 external logs have been collected and recorded, and nearly 100 websites that have not been recorded have been automatically verified, and nearly 20 "zombie websites" with extremely low daily average visits have been found.

In addition, according to statistics, in the current IPv6 transition stage, the proportion of IPv6 access to website services is between 14.11% and 19.5%, and the details are shown in Table 1.
Table 1. IPv4 / IPv6 network access of the website registered at the current stage

| Date     | Number of visits IPv4 | Number of visits IPv6 | Number of visits All | Percentage of IPv6 |
|----------|-----------------------|-----------------------|----------------------|--------------------|
| 2019-06  | 5326                  | 1230                  | 655699               | 18.76%             |
| 2019-07  | 4499                  | 8221                  | 532211               | 15.45%             |
| 2019-08  | 4107                  | 7534                  | 486136               | 15.50%             |
| 2019-09  | 5976                  | 1379                  | 735558               | 18.75%             |
| 2019-10  | 4164                  | 9098                  | 507400               | 17.93%             |
| 2019-11  | 3843                  | 8742                  | 471817               | 18.53%             |
| 2019-12  | 4703                  | 1139                  | 584244               | 19.50%             |
| 2020-01  | 2895                  | 5984                  | 349391               | 17.13%             |
| 2020-02  | 4666                  | 7672                  | 543402               | 14.12%             |

From Table 1, the proportion of users from IPv6 access is more than 15%, so IPv6 access has become a network link that must be considered by colleges and universities.

5. Concluding Remarks
In view of the requirement of the website record management in the campus network in the transition stage of IPv6, this paper designs and implements the website record management framework and system which supports automatic verification. Through scanning detection, page collection, access control and other ways, it actively realizes the verification of the record information, introduces the process of automatic verification in detail, and lists the main generation of page collection and reverse proxy service Code. The practical application shows that the system designed in this paper can realize the automatic verification and recording of the information of the website that has been put on record, detect the website services existing in the managed network, and check out the website that has not been put on record, which greatly reduces the workload of the administrator to check the website's record information, and improves the timeliness of the website's record management. In addition, according to the statistics of record management, in the transition stage of IPv6, IPv6 link access has become a problem that colleges and universities must consider. In the future, we will continue to study and design the system architecture supporting distributed deployment, so as to improve the performance of the filing management system.

6. Acknowledgment
Supported by CERNET Innovation Project(NGII20170405)

7. References
[1] Haiyan Wu. Automatic identification technology "forced" website filing, China Education
[2] Circular of the general office of the Ministry of education on the implementation of the action plan for promoting the scale deployment of Internet Protocol version 6 (IPv6), 2018. http://www.moe.gov.cn/srcsite/A16/s3342/201809/t20180903_347090.html

[3] Yongting Chen, In September, cngi-cernet2 backbone network traffic increased by 317% year on year, China Education Network, 2019; 9

[4] Xiaolian Jiang Haishan, Zhen. The design and implementation of the filing system for security loopholes management in Colleges and Universities. Network Security Technology & Application, 2017(04): 150-152.

[5] Yanyan Yin. The Design and Implementation of Campus Network Security Management. Beijing Jiaotong University. 2016.

[6] Yueling Wang, min Li. The method of improving the accuracy of the website's record information based on the automatic verification system of communication mode. TELECOMMUNICATIONS NETWORK TECHNOLOGY, 2017, 000(001): 44-46.

[7] Jingguo Ge, Wei Mi, Yulei Wu. IPv6 Transition Mechanisms: Survey, Evaluation Criteria and Deployment Considerations Journal of Software, 2014, 25(4): 896-912.

[8] Rodney R Rohrmann, Vincent J Ercolani, Mark W Patton. Large scale port scanning through tor using parallel Nmap scans to scan large portions of the IPv4 range[C]/ 2017 IEEE International Conference on Intelligence and Security Informatics (ISI). IEEE, 2017.

[9] Yuliang Shi, Ti Zhang. Design and implementation of a scalable distributed web crawler based on Hadoop[C]/ 2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA). IEEE, 2017.

[10] Dadzie P , Walt T V D . Website Statistics 2.0: Using Google Analytics to Measure Library Website Effectiveness[J]. Libri, 2015, 27(3): 261-78.

[11] Gengsheng Deng, Zhihui Yan, Weiping Zou. IPv6 Transition Based on dual stack reverse proxy technology application research . China Education Network, 2018; 1