

Design and Implementation of an Intelligent Enterprise Email Filter Based on Service Plugins

Jing Zhang^{1,*}, Jianmin Yao², Ling Zhang¹ and Qihua Zhang¹

¹Communication and Computer Network Lab of Guangdong Province, School of Computer Science and Engineering, South China University of Technology, China

²School of Computer Science and Engineering, Suzhou University, China

*Corresponding author

Abstract—An enterprise email filtering system is introduced integrating present business systems with service plugins, which connects the email and related business parts as office automation, cloud storage etc.. The user can fulfill his business-related process in the business system with personalized rules. The system also has adopted a special white list rule for a special business, which, on basis of the traditional email filtering functions, can recognize and process important emails for users. The design has promoted user friendliness of the mailbox and the business efficiency. The system is deployed in Guangdong Province Key Laboratory of Computer Network for email service and shows excellent performance in service.

Keywords-Business plugin, email filtering, business system

I. INTRODUCTION

Email service is often needed in enterprises, which are mainly satisfied by private email system or by outsourcing enterprise mailbox. Private email system service is suitable for large-scale enterprises by private security and steady communication service. Outsourcing enterprise mailbox can provide professional service and satisfy the needs of customers, which suits the small or medium-sized enterprises better.

But with the fast development of internet, enterprises are very active in building information systems. According to the 37th Report of Internet Development in China by the CNNIC [1], 40.7% enterprises have deployed information systems by December 2015. Enterprises are in great demand of office and mailing software. While present mail service mainly focuses on mail receiving-and-sending and social communications, little attention is paid on enterprise business routine connections. An efficient, personalized email system will be welcomed by enterprises when connected with enterprise business process.

A mail filtering architecture based on enterprise business plugin technology is introduced in the second section, which incorporates the enterprise business into the mail filter with a plugin system. The mailbox is more closely connected with the business and the staff can configure the mailbox interface to process related mail contents in corresponding business systems. Taking the Guangdong Key Laboratory of Computer Network as an example, the incidence relation between the business and mail service is analyzed in the third section by introducing the deployment of the mail filter with plugins for merging the mailing system and other systems like the Web-

lib cloud storage system. The performance of mail filtering is introduced by the actual operation conditions.

II. CHARACTERISTICS OF BUSINESS MAILBOX AND DESIGN OF MAIL FILTER

To design a mail filtering system for enterprise customers, the characteristics of enterprise mail contents and demand of customers are considered, which decide the function and architecture of the enterprise mail filtering system.

A. Reliable Senders and Classified Mail Filtering Strategy

Enterprise staff are taken as reliable senders, whose mails are not spams except under virus attack or other special conditions. For this reason, the following actions are taken in mail filter design to improve the mail service efficiency.

A mailbox white list rule is set. The administrator is responsible for addition and deletion of mailbox users, and chooses to set up the enterprise white list, which is pushed to all users. Individual users needn't maintain his own white list and thus the mail service is less complex for daily use.

Classified mail processing mechanism is realized. Traditional mail filtering classifies the mails as spam or normal by machine learning mechanisms. Because of the influence of the training corpus, the recall and precision of normal mail recognition is often contradictory to each other. To lower the probability of normal mail classified into the spam mailbox, spams are often allowed into the normal mailbox.

By setting up a white list in the enterprise mailbox, mails are classified into three levels: important, ordinary and spam. Enterprise and private white lists include important contacts among enterprise colleagues or social circles, mails from whom is of the highest priority and must be processed and classified as important. All mails from the white list needn't go through the mail filtering process with a special configuration of white list directory.

Except the white list contacts, the user has to process large amounts of mails from various contacts whose mails are related to the work, e.g., students' assignments for the teacher, while the students' mailbox needn't be recorded in the white list one by one. The mail filtering system of the enterprise mailbox will process this part of mails specially.

Other mails are from contacts outside of the white list, which are judged by the mail filter, and goes to ordinary mail directory or spam directory according to the recognition result.

This 3-level classification mechanism helps enhance the mail filtering efficiency, and is convenient for the user to read and process mails.

B. An Intelligent Mail-Work Flow Strategy Merging Mails And Business Process

Important mails in the enterprise mailbox are often business-related, and need further processing after reading, e.g., to copy the attachment to other business systems or to extract partial information for special business systems. For administration staff or business managers, mails need to be archived, and the users have to download the mail and upload or store it to other application systems for limited capacity of the enterprise mailbox.

Because the mail contents are often related to enterprise business, an intelligent work-flow strategy can be set up in the mailbox for automatic processing of mails. This will reduce the repetitive mechanical work of downloading and uploading.

For example, the university students are required to submit their assignments to the online learning system, but sometimes they still send their assignments to the teacher's mailbox for various reasons. The teacher have to download the attached files and upload them to the online system, which is a complex and trivial process. The intelligent workflow strategy can directly upload the assignment mails to the cloud storage system for the teacher to process, or upload to corresponding directories of the online system, which reduces the complexity of the teacher's operations.

III. MATRIX INTELLIGENT ENTERPRISE MAIL FILTER BASED ON BUSINESS PLUGIN

After analyzing the characteristics of the enterprise mailbox, the MATRIX intelligent enterprise mail filter based on business plugin is proposed.

A. The Operation Logic of MATRIX Intelligent Enterprise Mail Filter

The location and operation logic of the mail filter in the process of mailing service is shown in Figure I. The MX records of the DNS are utilized to filter the mails: With the MX records of the DNS, all mails sent to a domain can be processed by the mail processing system, after which the mails are sent to the real mail server. The advantage of this processing is that the mail processing system is not related to the software of the mail server, and thus compatible with common mail servers, and can support multiple mail servers at the same time.

Because the user can forward mails from other mailboxes to the MATRIX processing system, these mails can also use all functions of the system. On the other hand, all processing functions of the MATRIX system can be fulfilled at the server end but not at the browser or mobile client end, thus all clients can use the same mailing rule configuration at various

terminals. The same configuration at the server end applies to all clients, and the client can manage multiple mailboxes through different terminals.

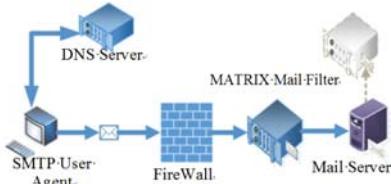


FIGURE I. THE OPERATION FLOWCHART OF MATRIX INTELLIGENT ENTERPRISE EMAIL FILTER

B. Functions and Implementation of MATRIX Intelligent Enterprise Mail Filter

Since the DNS MX records point to the MATRIX mail filter, and all mails to the server pass the filter first, the filter must have the function of an SMTP server for receiving and intelligent processing the mails. If the same process serves to receive and to process the mails, it may fail to respond in time to SMTP requests to receive the mails. Thus two independent process groups are designed for independent receiving and intelligent processing the mails. After the intelligent processing, mails are sent to the mail server and distributed to office sub-systems according to the user configuration. The system architecture is shown in Figure II.

The reception node is responsible for receiving mails, operating concurrency control, frequency control and mail feature control when processing the SMTP request. The processing node will scan and process the mails in turn. The modular design enables the filter to be customized, which calls different modules according to the filter's functions. Modules in this paper include the SPF detection module, the black list detection module, the virus detection module, the white list detection module, the mail filtering module and tagging distribution module etc..

The contact processing module is used to recognize the white list users, whose mails are directly channeled to the important mail directory without the mail filtering processing, which make it convenient for users to recognize and process. As mentioned above, the white list includes the enterprise public list and the user private list. The enterprise staff name and mail address list can be sent to all staff members via the MATRIX enterprise administrator configuration interface, and save the enterprise users from repetitive labor.

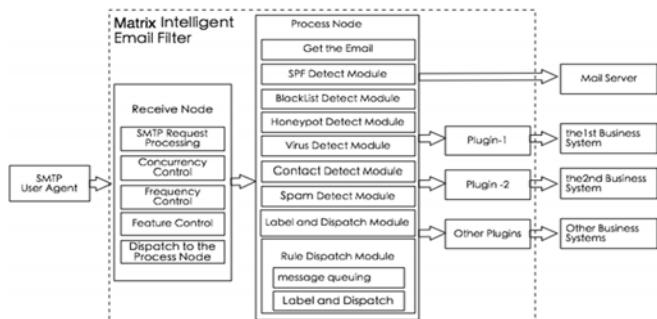


FIGURE II. ARCHITECTURE OF MATRIX INTELLIGENT ENTERPRISE EMAIL FILTER

Besides, the processing node also handles SPF detection, black list detection, virus detection and spam filtering etc., and tags the processed mails for further distribution. The mail distribution function includes two aspects. 1) All mails are sent to the mail server system; 2) Abiding by the user configuration, the engine module transfers the mails to various business systems via plugins in other business systems for further processing work-flow.

The rule-based mail distribution engine module monitors the message list notification from the processing module, and detects the mails. When some rules of the user configuration

are triggered, the mail will be transferred to corresponding sub application systems according to the rule content.

The concurrency control, frequency control, feature control at the reception node, and all modules at the processing node can support hot swap function. The configuration interface of the dynamic control reception node is shown in Figure III(a), the dynamic swap processing node in Figure III(b). When the system administrator changes the configuration at the management terminal through MATRIX system administrator's configuration interface, the change takes effect without restarting the service.



FIGURE III. (A) SETTING OPTIONS IN THE RECEIVE NODE (B) SETTING OPTIONS TO ENABLE/DISABLE THE MODULES IN THE PROCESS NODE

IV. APPLICATION OF THE MATRIX INTELLIGENT ENTERPRISE MAIL FILTER IN BUSINESS

The MATRIX Intelligent Enterprise Spam Filter is now deployed in practical operation environment in the mailing service at the Guangdong Key Laboratory of Computer Network.

A. Business Demand Analysis

The Guangdong Key Laboratory of Computer Network is a research organization with network technology and applications as its key research topic. There are around 80 members in the lab who are university teachers, administration members and students. There are various application systems in the lab, such as the enterprise cloud disk system Weblib for data cloud storage, the online learning system for instruction work, the equipment management system for lab device man-

agement, etc.. The connection of the Weblib enterprise cloud disk and the mail filter is taken as an example below to introduce the deployment and application of the MATRIX Intelligent Enterprise Mail Filter system.

The Weblib enterprise cloud disk is commonly used in the laboratory, which provides storage space for the members. For the teachers, the mailbox is full of students' assignments, which, in former mailing service, the teachers will download and process one by one. With the MATRIX system, the scattered student assignments will be selected and archived under the "Student Assignment" directory, which will enhance the efficiency and avoid omission of some assignments.

For the administration members, a weekly report must be submitted to the executive staff, which will automatically be archived by the MATRIX system to a specific directory and the cloud storage.

B. Deployment of the Intelligent Enterprise Email Filter

With the data interface and corresponding plugins provided by the Weplib enterprise cloud disk system, Weplib was connected to the MATRIX system, as shown in Figure IV.

All mails are forwarded to the mail server system after the mails are tagged in the processing node. The mails that satisfy the user configuration rules are forwarded to the Weplib cloud disk with the engine node. The processing node adopts the SpamAssassin as a spam filtering module.

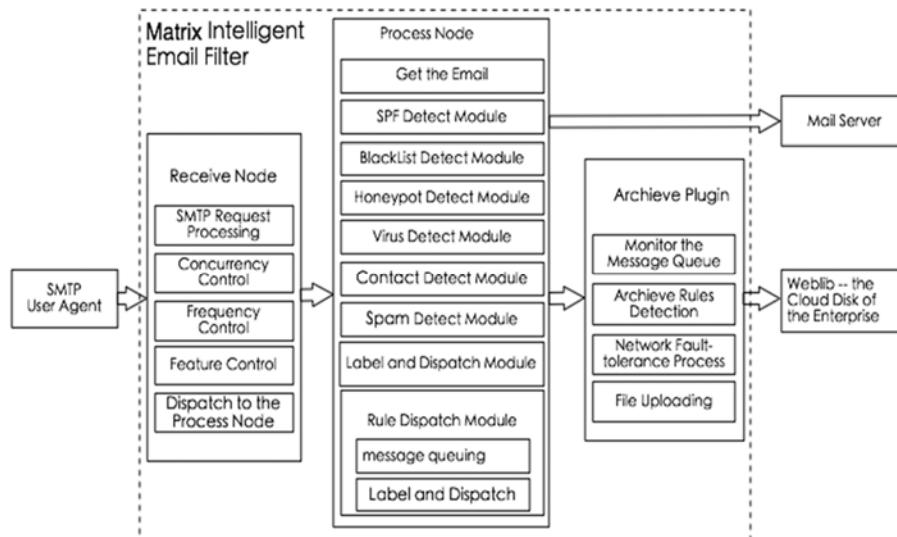


FIGURE IV. ARCHITECTURE OF MATRIX EMAIL FILTER WHEN LINKING WITH THE CLOUD DISK OF THE ENTERPRISE

The system will forward specific mails or part of the mail contents to the cloud disk according to the archiving rules, in which the symbol “*” stands for arbitrary contents, while the symbol “|” stands for or relationship. The system supports drop-down prompt to complete the rule content automatically. And the user can change the order of the rules by pressing a certain line with the mouse and dragging the gray block. The user can also start or stop the rules easily.

Hot swap is supported in association between the MATRIX mail filter and other business systems. Taking the Weplib enterprise cloud disk as an example, when it is configured in the MATRIX system, the configuration interface will change to the Weplib archiving rule configuration interface as shown in Figure V. If the enterprise user no longer

C. Operation After Associating the MATRIX And the Weplib Cloud Disk System

The operation condition after associating the MATRIX mail filter and the Weplib in Guangdong Province Key Laboratory of Computer Network is shown in Figure 4. The user defined mail archiving rules through the MATRIX management interface shown in Figure V.

needs to forward the mail to the Weplib cloud storage, the enterprise administrator may remove the binding between the MATRIX and the Weplib systems at the configuration terminal, which automatically deletes the Weplib system configuration interface and mail archiving function from the MATRIX configuration interface.

When the archiving rules are configured by the user for the Weplib, all mails abiding by the rules will be archived to the corresponding directories in the Weplib cloud disk. A directory structure that corresponds to the user rules in the Weplib cloud disk is shown in Figure VI. The mail attachment storage example is shown in Figure VII. The user may read related mails from Weplib even if he doesn't log in to the mailbox.

规则	停启	删除	更改
主题 匹配 周报、日报 则将 邮件正文 发送至 报告类邮件	启用 禁用	删除	更改
发件人 匹配 联系人 则将 邮件正文 发送至 联系人邮件	启用 禁用	删除	更改
附件 匹配 * 则将 当前附件 发送至 邮件附件	启用 禁用	删除	更改
附件 匹配 png、jpg、jpeg 则将 当前附件 发送至 邮件图片附件	启用 禁用	删除	更改
主题 匹配 * 则将 正文和附件 发送至 我的邮件	启用 禁用	删除	更改

FIGURE V. (A) UI FOR THE USER TO CONFIGURE ARCHIVE RULES IN MATRIX TO ARCHIVE THE EMAILS TO WEBLIB

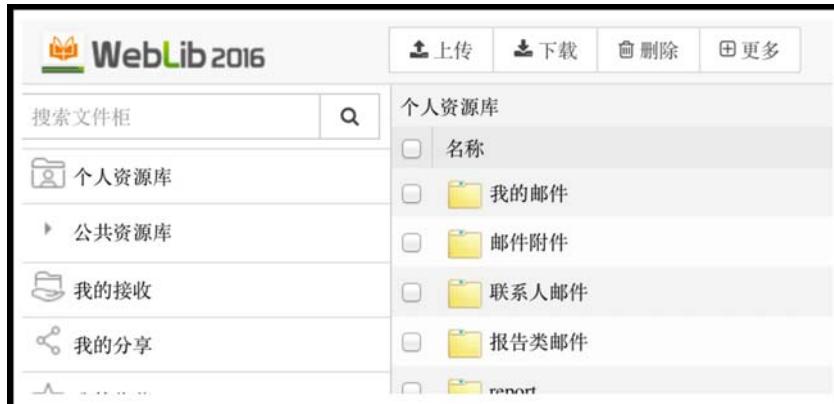


FIGURE VI. ILLUSTRATION OF FOLDERS IN WEBLIB



FIGURE VII. ILLUSTRATION OF EMAILS ARCHIVED IN WEBLIB

MATRIX configured multiple user accounts with various permissions except the staff accounts. The enterprise administrator account has the admissions to set enterprise white list, to choose other business systems associating with the system, to check various reports from the system etc.. The system administrator can specify in the MATRIX system operation which hot-swap modules are to be applied, the training period of the spam filtering, viewing performance data of the system operation, viewing alert information of the system operation etc.. The observer account has only the admission to read and to search, to view the performance data, the alert information of the system operation.

D. Performance Test of the MATRIX System

The MATRIX intelligent mail filter has operated in the laboratory for over one year with good performance as shown in the function test and performance test. The performance test involves average response time, delay time and throughout capacity^[2,3], which justifies the performance of the system.

The SMTP pressure test tool SMTP-Source is used to send large amounts of mails to the system. A record is taken at the receiving and processing nodes of the mail processing system to record mail receiving time, process starting time, process ending time and other logics, after which the respond time, delay time and throughout capacity are automatically

measured. Considering that the archiving node performance is mainly decided by the Weblib performance (including uploading time of the attachment etc.), and the Weblib is an enterprise business system which is not controlled by the MATRIX system, the archiving performance is not tested. In real practice, the archiving delay is within acceptable amount.

Take the mail size and concurrency as variables, the test is taken on mail sizes of 8k, 16k and concurrency of 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900 and 2000, each concurrence operates 5 times. Delete the maximum and minimum, we get the response time, delay time and throughout capacity.

In the test, we set up 300 enterprise IP black list, 300 enterprise mail black list, 300 honeypot mailboxes, 300 enterprise contacts and 1 million contact records. And so we got 10,000 users and 100 contact records for each user. Because the contact records are indexed in the MATRIX system, the search expense of contact detection is not heavy.

The system evaluation is made in an operation environment of a CPU of 2.1GHz*12 cores. For the heavy expense of rule detection of SpamAssassin, the SpamAssassin and its node processes are bound to the CPU0 through to CPU7, receiving nodes to the CPU8 through to CPU11, to avoid its influence on

the receiving node. This method avoids the influence of filtering processing on receiving mails.

1) Response Time Evaluation

Response time refers to the time consumption between the user sending a mail and receiving the mail, which is a key indicator of user friendliness to the mail sender. The smaller this value is, the better. The average response time in various concurrency conditions are shown in Figure VIII.

As shown in the figure, when the concurrency is less than 800, the response time is less than 1 second, which is acceptable to users.

2) Delay Time Evaluation

Delay time refers to the time consumed after the MATRIX system receives the mail and before finishes processing the mail, which is a key indicator of user friendliness to the mail receiver. The smaller this value is, the better. The Delay time is composed of two parts, one of which is the waiting time after the system receives the mail at the receiving node and before processes the mail at the processing node, and the other part is the time consumption of processing the mail at the processing node. In the concurrency test, the average waiting time, average processing time and average delay time are shown in the evaluation curves in Figure IX, X and XI.

The figures showed that the waiting time and processing time increase with the concurrency. It is mainly because of the heavy expenses of SpamAssassin while the limited capacity of the CPU. The final result is an overall delay time of less than 10 seconds with a concurrency under 800. For laboratory or general enterprise applications, this performance is acceptable. If higher processing speed is needed, part of the SpamAssassin rules can be forbidden or take other optimization methods.

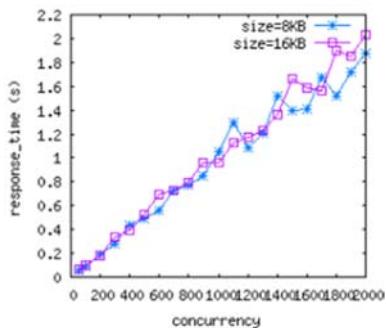


FIGURE VIII. CURVE OF AVERAGE RESPONSE TIME PERFORMANCE OF MATRIX

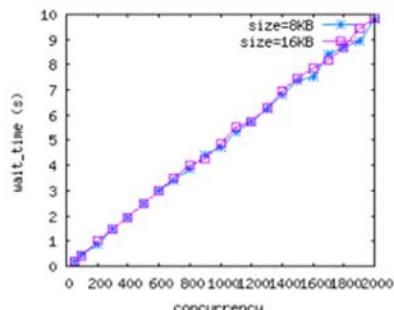


FIGURE IX. CURVE OF AVERAGE WAITING TIME PERFORMANCE

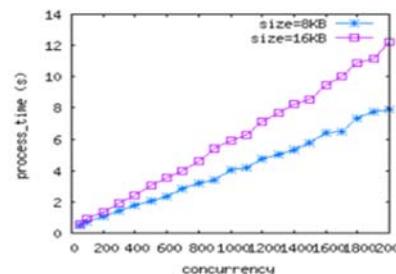


FIGURE X. CURVE OF AVERAGE PROCESS TIME PERFORMANCE

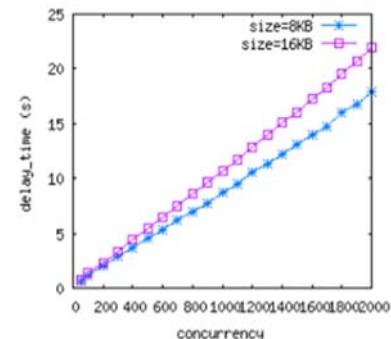


FIGURE XI. CURVE OF AVERAGE DELAY TIME PERFORMANCE

V. CONCLUSION

A MATRIX mail filtering system based on hot swap is introduced, which supports regular mail filtering functions and serves as a bridge connecting the user mailing service and the present business system, and thus embeds the mail or part of the mail into the business system.

As it becomes cheaper and cheaper to deploy new products, the small and medium-sized enterprises start to pay attention to cloud computing, big data and other new technologies. The plugin based mail filtering architecture can satisfy the enterprise demands in mail archiving in cloud storage, in combining the mail service and enterprise business, and in data mining from business data.

ACKNOWLEDGEMENTS

Thanks are due to former students at the South China University of Technology for their help in calculations, in particular, Huang Kangquan. This work was supported by the National Natural Science Foundation of China under Grant No. 61672367 and the Teaching and Research Project of South China University of Technology under Grant No. Y915001w. This work was also supported by the CERNET Innovation Project under the Grant of “The Next-Generation Intelligent Enterprise Mail Processing System”.

REFERENCE

- [1] Statistical Report on Internet Development China, Published by China Internet Network In China https://www.cnnic.cn/gjwm/xwzx/rdxw/2015/201601/t20160122_53283.htm, January, 2016
- [2] Fang Weidong etc. How to Evaluate the System Performance[J]. China Education Network, 2008(06):p20-20
- [3] Huang Kangquan, Design and Implementation of the Enterprise Email Filtering System[D]. South China University of Technology, 2015