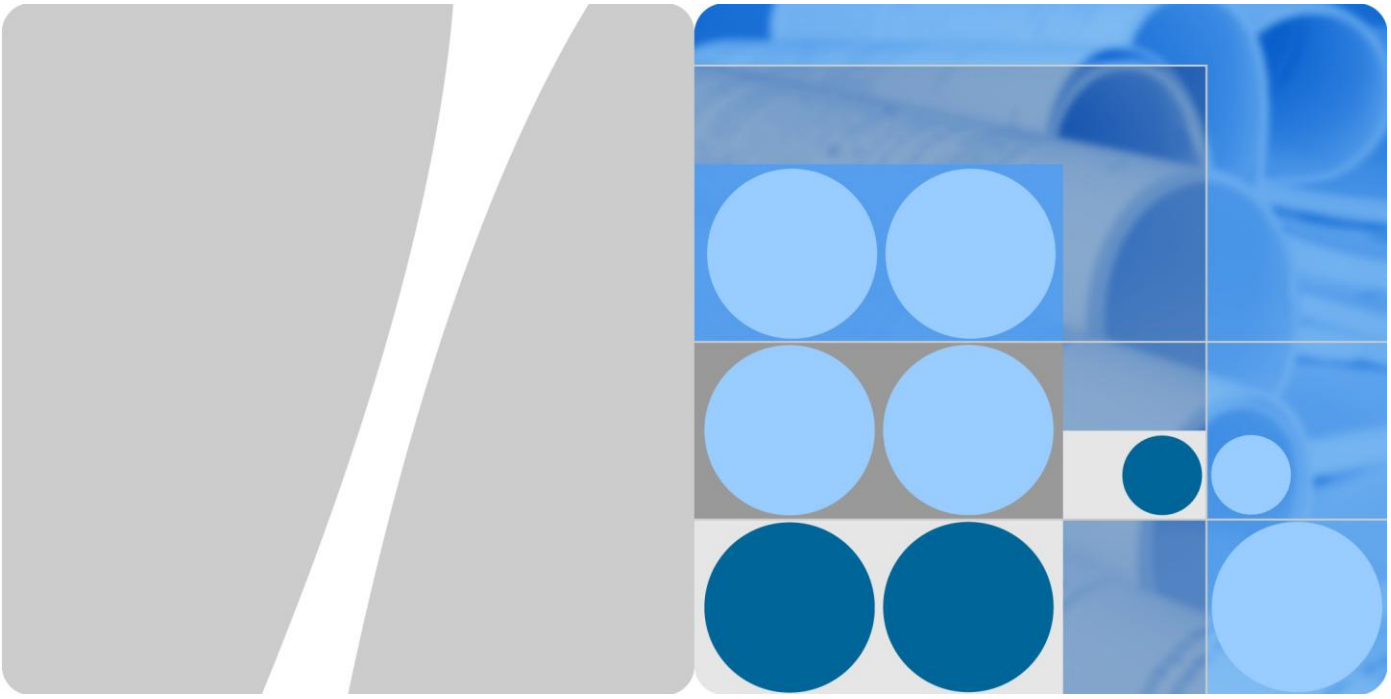
Doc. code



HUAWEI Secospace USG Series Content Filtering White Paper

Issue 1.0

Date 2014-03-27

HUAWEI TECHNOLOGIES CO., LTD.

Trademarks and Permissions


and other Huawei trademarks are trademarks of Huawei Technologies Co., Ltd. All other trademarks and trade names mentioned in this document are the property of their respective holders.

Notice

The purchased products, services and features are stipulated by the commercial contract made between Huawei and the customer. All or partial products, services and features described in this document may not be within the purchased scope or the usage scope. Unless otherwise agreed by the contract, all statements, information, and recommendations in this document are provided “AS IS” without warranties, guarantees or representations of any kind, either express or implied.

The information in this document is subject to change without notice. Every effort has been made in the preparation of this document to ensure accuracy of the contents, but all statements, information, and recommendations in this document do not constitute the warranty of any kind, express or implied.

Huawei Technologies Co., Ltd.

Address: Huawei Industrial Base
Bantian, Longgang
Shenzhen 518129
People's Republic of China

Website: http://www.huawei.com

Email: support@huawei.com

Contents

1 Background

The rapid development and popularization of the computer network in various fields of the social life bring convenience to obtaining, sharing, and transmitting information. Various organizations such as the government and enterprises become increasingly dependent on networks which can provide related information in a timely manner facilitate the organization development. Meanwhile, information security becomes more and more significant.

On the one hand, organization networks are confronted with too many risks such as phishing networks, viruses, information interception, and various malicious codes on the Internet.

On the other hand, mass Internet applications and information attract employees to spend more and more time on the Internet, resulting in the inefficient use of the innately vulnerable organization network. For example, employees may access to illegitimate websites, spread illegal speeches, transmit sensitive information, and abuse network resources by using applications such as P2P downloading, IM, online games, online video, and stock transactions. These behaviors affect the work efficiency of the organization, brings threats to information security, and may cause irreparable loss.

How to ensure the network security, regulate employees' online behaviors, and ensure the proper operation of the network system has become a major concern for organizations.

2 Purpose of Content Filtering

As reported in Status Quo and Developing Trend of IT issued by IDC, content security problems have become the primary threat to enterprises. There is a dramatic increase in network security events such as security risks, decreased work efficiency, sensitive information leaks caused by services such as normal network access, email receiving and sending, P2P downloading, IM, forum, and online video. Besides, with the in-depth development of Internet applications, more and more network security events occur on the application layer.

Traditional firewall is a basic infrastructure and plays an important role in ensuring network security. However, they are vulnerable to new network security problems.

In most cases, they function as access control devices to permit traffic matching security policies. Advanced firewalls detect the robustness of protocols and allow legitimate traffic .However, new attack methods spring up and a large amount of attacks penetrate to the application layer. Traditional firewalls are impotent because these attacks seem to be secure on the network layer.

IPS and antivirus devices focus on obviously insecure internet access, protect servers in the organization from malicious attacks, and prevent employee PCs from being infected by viruses due to inappropriate Internet access. However, they cannot prevent employees from accessing secure but illegitimate content, such as pornographic websites, unauthorized transmission of intellectual property right works. They can neither prevent employees from intentional betrayal of organization confidential information, which may result in huge economic losses.

2.1 Harms of Network Security Problems

For an organization, internal employees' incorrect access to network brings loss in productivity and bandwidth, and severely compromises the network security construction and information system of the enterprise. Improper or illegal content on the network may even greatly harm the employees' physical and mental health and bring legal issues to the enterprise. The following describes the harms in details:

Employees are prone to the threats of insecure links or malicious downloads when accessing websites. Their PCs, if planted with malicious code programs, will make the organization network a botnet or infected with viruses.

Employees are prone to spoofing attacks such as network phishing, resulting in bank account and password compromise that may cause economic losses.

Employees may access, download, or transmit copyright protected contents such as unauthorized thesis, audios, and videos illegally. The penalty for such behaviors is increasingly severe, and the organization is the ultimate victim that pays for the behaviors.

Employees may access radical websites, participate in illegal online activities, deliver sensitive information, spread illegal speeches, such as reactionaries, terrorist organizations, network fraudulent activities. These behaviors affect the reputation of the organization, bring economic losses to the organization, and even bring legal problems because these websites are usually illegal.

2.2 Bandwidth Abuse

When accessing websites, employees are often attracted by entertaining contents. Statistics show that most enterprise employees spend much time browsing non-work-related contents, which occupies much bandwidth and reduces work efficiency.

The egress bandwidth of most enterprises and public service units is less than 10M, but the bandwidth price is one of the most expensive. Employee PCs keep running download software to download audio and video files. The download occupies much bandwidth resources and the bandwidths for normal services are not ensured. In this case, the work efficiency of the organization decreases, and the normal work cannot be accomplished in a timely manner, bringing great loss to the organization.

2.3 Information Leaks

Unauthorized transmission of organizational confidential information and key information, known as information leak, also brings huge losses. There are a great number of such cases.

Confidential information may be leaked by ill-intentioned employees for economic benefits or honest and trust-worthy employees who are fooled to send important information to a third party or even competitor.

Therefore, more reliable methods are needed to prevent such behaviors.

3 Key Technology in Content Filtering

Predefined categories and real-time analysis are used in content filtering to control internet access.

Predefined categories, as the name implies, categorizes websites in advance. Content filtering needs only to query website categories, which ensures rapid response and high performance. The predefined category database can be updated dynamically in real time. Predefined categories resolves most security problems in web access.

Through in-depth analysis of web packets and packets of other protocols such as FTP, SMTP, and POP3, real-time analysis technology analyzes user behaviors and the content in transfer in real time to block non-work-related and risky behaviors.

The perfect combination of the two technologies enhances the efficiency and accuracy of content detection.

✧ URL Categorization

The URL filtering service can identify and block malicious websites so that it can minimize the possibility of infection by Trojan horses. The URL filtering function is the best means to block the phishing websites.

Huawei Symantec proprietary URL categories are predefined on devices and fall into keywords categories and URL categories. The URL category database contains more than 65,000,000 URLs in more than 10 languages, Chinese, English, French, Dutch, Russian, Spanish, Portuguese, Italian, Arabian, Farsi, Japanese, Ukrainian, Czech and Deutsch. The identification rate of the devices is higher than 85% and the accuracy is higher than 90%, which are among the best in the world.

Meanwhile, the technology also supports user-defined URL category to satisfy the special requirements of various customers.

✧ In-depth Protocol Analysis and Decoding, Multilayer and Granular Behavior Control

Multilayer and granular control are implemented based on protocols such as HTTP, FTP, SMTP, POP3 and webmail, behaviors such as upload, download,

mail receiving, and mail sending, as well as information such as the name, type and size of the file. The organization can deny network access completely, or allow browsing and downloading but block sending internal information outside, or allow sending a few common text information but deny sending word documents and source code files that may contain core confidential information. Such multilayer and granular access control ensures network security and information security of the organization.

✧ **Integrated Content Filtering**

Information and files are the content to be controlled. Network access can transmit such information with various methods by using various protocols. Configuration becomes easy through the abstraction and commonization of keywords, file names, and file types. For example, if the administrator wants to deny the sending of office documents and compressed packages outside, the user needs only to configure a file type pattern group and reference it in related protocols such as HTTP, FTP, and mails, but not perform the configuration repeatedly for each protocol.

✧ **Advanced Webmail Signature**

Signatures can be customized for specific webmail brands according to user requirements, making webmail content filtering more specific. It accurately identifies webmail services, including the mail sending, attachment upload, mail card sending, postcard sending, and automatic reply configuration, providing strong evidence for mail audit and backtracking. The signature file keeps synchronized with the webmail server. When the software of the webmail server is upgraded, you need only to update the webmail signature file to implement content filtering on the latest webmail brands.

3.1 URL Filtering

URL filtering controls users' HTTP requests by allowing or denying users to access certain network resources, therefore regulating online behaviors.

3.1.1 Introduction to URL

Each web page on the Internet has a unique identifier, which is called Uniform Resource Locator (URL).

The common format of an URL is protocol://hostname[:port]/path[?query]. Table 3-1 shows the parameter description.

Table 3-1 URL parameters

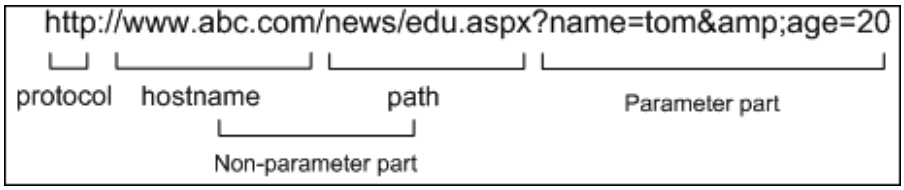
Field	Description
protocol	Indicates the application protocol. HTTP is the most commonly used one. You do not need to enter http:// when the protocol is HTTP. Notes: The URL filtering function supports only HTTP URL requests.

Field	Description
hostname	Indicates the domain name or IP address of the web server.
port	(Optional) Indicates the communication port. Application protocols have default port numbers, for example, the default port for HTTP is 80. When the web server uses a non-standard port, the port number in the URL cannot be omitted.
path	Indicates a directory or a document location on the host. The path is a character string separated by zero or multiple slashes (/).
?query	Optional. Used for transmitting parameters to dynamic web pages.

In Figure 3-1, http://www.abcd.com/news/education.aspx?name=tom&age=20 is used as an example:

- www.abcd.com indicates the host name.
- news/education.aspx indicates the path.
- ?name=tom&age=20 indicates the parameters.

Figure 3-1 URL formats



3.1.2 Matching Order of URL Filtering

URL filtering filters only the URL requests of HTTP.

By analyzing the URLs of the network access requests from users, the device compares the URLs with the blacklist, whitelist, and URL categories, and then decides to permit or deny the requests according to policies.URL filtering observes the following matching order: whitelist, blacklist, user-defined URL categories, and predefined URL categories. After a URL matches a certain matching item, no further matching is performed.

- Blacklist and whitelist

The device matches the URL of the network access request with the blacklist and whitelist. If the whitelist is matched, the HTTP request is allowed. If the blacklist is matched, the HTTP request is blocked and a web page is pushed.

If the URL of an Internet access request matches the whitelist, no further operation (including AV and IPS) is performed. The whitelist helps improve the matching efficiency.

- User-defined URL categories

User-defined URL categories are configured and maintained by users. URLs with the same signatures are grouped by user-defined URL categories. Users can configure policies to permit or deny the access to URLs in all categories. Compared with pre-defined URL categories, user-defined URLs enable more refined control of users over URLs.

- Predefined URL categories (remote URL filtering)

Pre-defined URL categories are provided and maintained by the security service center. Predefined URL categories cover various URL categories, such as education URLs and news URLs. Users can apply policies by service to permit or deny the access to URLs in all categories. To apply pre-defined URL categories, you need to set up a connection to the security service center. A large number of common URLs are already grouped in predefined URL categories so that users can easily control the accessible and inaccessible URL categories.

Filtering policies based on URL categories implement access control on the users in specified address sets in a certain time segment. The combination of the URL module and firewall policies implements URL category filtering by time and address.

- Time object

The USG supports URL filtering of HTTP requests in specific time segment.

- Address object

The USG supports URL filtering of HTTP requests initiated by specific addresses or address groups.

3.1.3 URL Matching Mode

In URL filtering, the blacklist, whitelist, and user-defined category match URLs based on common pattern groups. For example, user-defined category **uc:game** references common pattern group **game** that contains patterns **aaa.com** and **bbb.net** and the matching mode is any. All URLs containing **aaa.com** or **bbb.net** are added to user-defined category **uc:game**.

3.1.4 URL Category Database

URL service platform consists of the categorization platform, central database, query server cluster, and upgrade server. Among them, the categorization platform obtains URL page data from the Internet and categorizes URLs using proprietary intelligent automatic categorization technology. The categorization result data is saved in the central database. The upgrade server automatically monitors the data updates in the central database, and instructs the query server cluster on the service platform and local URL query server of the customer to perform data updates.

The customer can obtain the URL category services either in local query mode or remote query mode.The local query mode requires that a local URL query server is deployed in the local network of the customer and provides URL category query services for various devices. Different from the local query mode, for customers that use remote query, the devices access the URL

category service platform directly through the Internet to obtain the URL categories.

3.1.5 URL Hotspot Database

After the URL filtering service is enabled, the device queries the nearest URL query servers. Standard pre-installed URL category database which stores the information about hotspot websites is provided to reduce query latency and improve user experience.

The update file of the standard pre-installed URL category database is less than 10 MB. The device does not require frequent updates. Because hotspot database query is performed locally, the query rate is enhanced greatly.

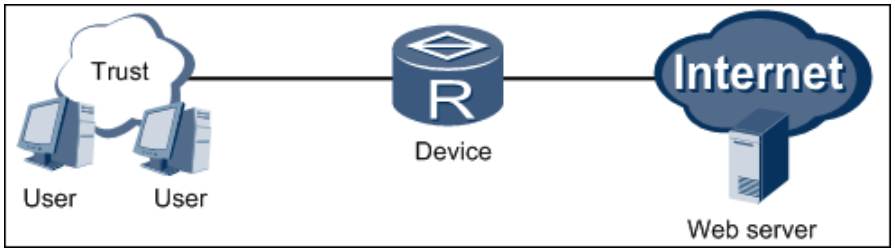
3.2 Search Keyword Filtering

Search keyword filtering filters out the keyword in the specific search engine, controls the search content of intranet users, and prevents unauthorized access to sensitive information.

3.2.1 Application Scenario

Currently, search engines such as Google, Yahoo, Bing, and Baidu support keyword filtering. The following figure shows the networking for search keyword filtering.

Figure 3-2 Networking for search keyword filtering



Once search keyword filtering is enabled on a device, the device resolves HTTP packets in the following way when users use search engines:

- If the HTTP packet matches the keyword policy, the filtering policy is executed (block the packet or generate alarms). Meanwhile, the device can push web pages to inform the user.
- If the HTTP packet does not match the keyword policy, the HTTP packet is allowed through, and the user can search for the keywords.

3.2.2 Matching Method

Search keyword filtering supports keyword matching. For example, when the administrator configures **Violence** as an illegitimate search keyword, you cannot obtain any result by entering a word, phrase, or sentence that contains the keyword. When the user's request is denied, the device pushes web information to the user. The web information can be customized.

3.3 Web Content Filtering

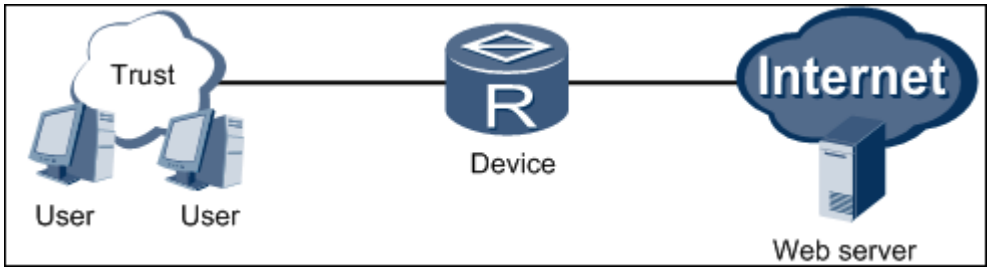
Web content filtering controls the web page content that can be accessed by users.

3.3.1 Application Scenario

Web content filtering filters out HTTP web page content based on the following items:

- Web browsing keyword: filters according to the Web page content.
- HTTP POST filtering: filters according to HTTP POST operations.
- HTTP POST file filtering: filters according to the operation on HTTP POST files.
- Uploaded/Downloaded file name keyword: filters according to the name of the uploaded or downloaded file.
- Uploaded/Downloaded file type keyword: filters according to the type of the uploaded or downloaded file.
- File size filtering: filters according to the size of the uploaded/downloaded file.

The following figure shows the networking of web content filtering.



The following uses the uploaded file name keyword as an example to describe the processing procedure of web content filtering:

1. The user initiates an operation to upload an attachment, which triggers the sending of a HTTP packet.
2. When the HTTP packet reaches the device, the device implements the filtering policy:

- If the HTTP packet matches the keyword policy, the filtering policy is executed (block the packet or generate alarms). Meanwhile, the device can select to push web pages to inform the user of the violation.
- If the HTTP packet does not match the keyword policy, the HTTP packet is allowed through, and the user can upload the attachment.

3.3.2 Matching Method

Web browsing keyword filtering and file name filtering support keyword pattern group matching (any matching). For example, if the administrator configures keyword **Violence** as an illegitimate word in web browsing, users cannot access the websites that contains **Violence**.

File type filtering supports file type pattern group matching (exact matching).For example, if the administrator configures .mp3 files as illegitimate files in file download, users cannot download .mp3 files.

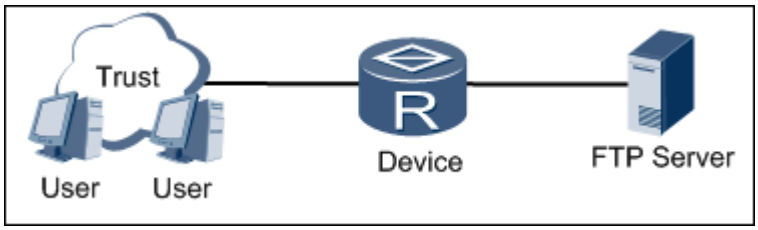
3.4 FTP Content filtering

The FTP filtering function controls uploaded or downloaded FTP files and FTP operations.

3.4.1 Application Scenario

FTP provides file transfer services between the local PC and the remote PC. FTP is widely applied in service operations such as version upgrade, log download, file transfer, and configuration storing. Unrestricted FTP operations may bring uncontrollable threats to the network. FTP filtering controls FTP operations (upload, download, deletion), names of the uploaded or downloaded files, file types, and file sizes.

Figure 3-3 Application scenario of FTP filtering



When the FTP filtering function is enabled on the device, the device resolves the FTP operations such as upload, download and deletion.

- If the FTP operation matches the FTP filtering policy, the policy is executed (block the operation or generate alarms).
- If the FTP operation does not match the FTP filtering policy, the FTP operation is allowed.

3.4.2 Matching Method

FTP file names support keyword pattern group matching (any match). For example, if the administrator configures keyword **Violence** as an illegitimate word in FTP download, users cannot upload or download a file whose name contains **Violence**.

FTP file type filtering supports file type pattern group matching (exact match).For example, if the administrator configures .mp3 files as illegitimate files in file download, users cannot download .mp3 files.

3.5 Mail Content Filtering

When an intranet user sends or receives mails through the webmail or SMTP/POP3 client, the mail filtering function monitors the mail address,

subject (title), body, attachment size and number, attachment name, or attachment extension to prevent data leaks.

Mail filtering prevents internal confidential information from being leaked and intranet users from sending and receiving mails containing sensitive information. This avoids law violation and malicious content.

3.5.1 Application Scenario for Webmail Mail Filtering

Webmail refers to a service or technology for sending and receiving mails by using network browsers. For webmail, mail clients are unnecessary. As long as you can access the Internet and have a webmail account, you can send and receive mails from the mail server (email) by using network browsers.

Webmail mail filtering monitors the mail address, subject, body, attachment size, attachment name, or attachment extension when an intranet user sends or receives mails using the email box.

3.5.2 Application Scenario for SMTP/POP3 Mail Filtering

Simple Mail Transfer Protocol (SMTP) transfers emails on the Internet. Post Office Protocol (POP3) allows clients to obtain mails from the mail server. A client needs to be installed on the user PC to transfer mails through SMTP/POP3, for example, Outlook and Foxmail.

SMTP/POP3 mail filtering monitors the mail address, subject, body, attachment size, attachment name, or attachment extension when an intranet user sends or receives mails using the mail client.

3.5.3 Filtering Method

The SMTP/POP3 mail is sent and received together with the attachment. If the attachment does not meet the policy requirement of the device, the entire mail is blocked; differently, attachment upload and mail sending through webmail are independent. If the attachment to be uploaded through webmail does not meet the policy requirement of the device, the upload is blocked, which does not affect the sending of the mail body.

4 Value of Content Filtering

4.1 Ensuring Content Security and Guarding Office Network

Using the pre-installed URL category database and URL category query service that contains 65,000,000 websites, the device blocks malicious URLs that may bring hidden troubles, and pornographic and retroactive URLs that are not allowed by the organization. In this way, intranet employees' active or passive accesses to illegitimate websites are reduced, which secures the office network.

4.2 Regulating Online Behaviors and Improving Work Efficiency

On the one hand, the organization wants to provide convenient Internet access for employees so that they can obtain latest information and required documents. Besides, more flexible and human centric management is required.

On the other hand, network abuse brings huge loss in productivity and dramatic decrease in work efficiency. For example, if an employee spends one hour per day in browsing non-work-related websites, 1.5-month salary is paid for nothing each year. On the whole, the organization pays a remarkable amount of money for nothing each year.

The content filtering function serves the two ends.

The URL category database that contains more than 65,000,000 websites and possesses an analysis capacity of more than 1,000,000 websites per day can categorize nearly all websites accessed by employees. Integrated with user-defined keyword pattern groups, file name pattern groups, and file type pattern groups, URL filtering enables the organization to provide convenient accesses to related information, news, and technical websites, while entertaining and retroactive websites are inaccessible. Even entertaining contents on legitimate websites can be prevented from download, which ensures employees' work efficiency.

4.3 Controlling Information Sending and Preventing Information Leaks

Besides network abuse and reduced work efficiency, information leak is another problem that brings huge economic losses to the organization.

Unsatisfied with the organization in terms of the salary, interpersonal relationship, or unfair treatment, or lured by huge economic benefits, or deceived by malicious persons because of low vigilance, employees may provide confidential information such as intellectual property rights, quotation, internal policies for third parties or even competitor. Confidential information can be leaked through BBS, blogs, FTP file transfer, mail clients, and webmail.

For information leaks through BBS and blogs, content filtering parses HTTP in-depth and filters employees' posting behaviors by auditing the keywords in the posts so that the content containing the related keywords cannot be sent to external networks.

In typical examples of information leaks through FTP files, employees send design documents or source code documents to third parties or even competitors. By restricting keywords, file types, and file sizes, content filtering effectively controls the sending out of important documents through FTP and generates logs for auditing.

For information leaks through mails, content filtering controls the sending out of important information by restricting keywords, file types, and recipients.

5 Application

The bandwidth of small- and medium- sized enterprises is usually less than 10M. Some enterprises deploy firewalls on the network. The enterprises are faced with the following threats:

- P2P and video traffic blocks normal services.
- Entertaining websites, IM, and online games affect work efficiency.
- PCs affected by viruses due to vulnerabilities in the browser and files result in the loss of private and identity data.
- Internal employees give out confidential information in various methods.

Content filtering system is deployed on the ingress of the intranet. That is, the system is connected between the intranet and Internet access device in in-line mode. If a firewall is deployed, the content filtering device is usually deployed inside the firewall (at the enterprise side).Meanwhile, based on the multilayer user management of the device, various content filtering policies can be configured for employees. For example, entertaining websites are not accessible in work time; non-technical websites are not accessible to R&D departments; employees in the R&D department are not allowed to send out documents; marketing personnel can send mails, but the mails cannot contain illegitimate keywords or be of illegitimate types.