2013 16th International Conference on Network-Based Information Systems

A proposal for the Deterrence of Information Leakage using Anti-virus Software

Takuya Kaneko[1], Takumi Nagaya[1], Keisuke Takemori[2], Yutaka Miyake[2], Masakatsu Nishigaki[1]

1 Graduate school of Informatics, Shizuoka University
3-5-1 Johoku, Naka, Hamamatsu, Shizuoka, 432-8011 Japan

2 KDDI R&D Laboratories, Inc.
2-1-15 Ohara, Fujimino, Saitama, 356-8502 JAPAN
{ takemori, miyake}@kddilabs.jp

***Abstract*— In recent years, the leakage of confidential files through the P2P (peer-to-peer) file sharing software has become a problem. Once sensitive files have been leaked to a P2P network, they are distributed rapidly and at random to a great many computers, bringing the danger of far-reaching damage. To deal with this, we propose a method to control the diffusion of sensitive file leakage on P2P networks by deleting the leaked sensitive files in the destination PC of the P2P user. In this method, we focus on the characteristic of anti-virus software that reacts to the unique patterns of a virus, automatically isolating and deleting a file concerned. By using the unique pattern of a virus to make the leaked sensitive file appear as a viral infection file, the anti-virus software installed on the destination PC will automatically delete the file. This is the concept of our method.**

Keywords-P2P file sharing software; data leakage; anti-virus software; unique pattern of a virus

I. INTRODUCTION

In recent years, the leakage of sensitive files through the P2P file sharing software (hereafter, P2P) has become a problem. To clarify the characteristics of the problem of sensitive file leakage via P2P, we point to the existence of viruses such as Antinny [1] which misuses P2P to leak files from PCs. Once infected by this type of virus, PC files are uploaded to P2P networks regardless of the user's intent, and sensitive files containing the personal information and internal data are leaked using P2P as the medium.

Another problem to be raised is the fact that it is not possible to identify the users who have obtained the sensitive files. Because P2P conducts end-to-end communication, it is not possible to manage logs of who accessed which files. Therefore, it can be said that those who obtain the leaked sensitive files cannot be specified, and the recovery of the leaked sensitive files is a highly difficult task. In addition, because the sensitive file, once leaked, diffuses to the entire P2P network, the danger of far-reaching damage is very high.

To deal with these problems, there are many enterprises that undertake measures such as controlling the leakage of sensitive files from internal networks and the transport of sensitive files outside the firm.

Nonetheless, there is no end in sight to the problem of leakage of sensitive files via P2P. It is thought that the sensitive files are leaked on the P2P network since users with low security awareness was infected with the virus such as Antinny after the sensitive file is taken out to a private PC in which the P2P software is easily installed. It is insufficient only to manage the exit of the sensitive information as long as even a single user with low security awareness is present in the organization.

Then, this proposed method to control the diffusion of sensitive files that are leaked out onto P2P networks by deleting the sensitive files thus leaked on P2P user's own PCs. It is expected that the leakage risk and the damage of the sensitive information can be greatly decreased by using this method in combination with exit management (measures to prevent the sensitive files from leaking from an internal network to the outside).

In this method, we focus on the characteristic of anti-virus software that reacts to the unique patterns of a virus, automatically isolating and deleting a file concerned. By using the unique pattern of a virus to make the leaked sensitive file appear as a viral infection file, the anti-virus software installed on the destination PC will automatically delete the file. Especially, in response to an increase in the virus damage and data leakage caused by viruses such as Antinny, the rate of enterprises and individuals implementing anti-virus software is increasing [2, 3, 4]. Thus, leveraging anti-virus software in measures to prevent data leakage holds great promise of effectiveness.

II. EXISTING MEASURES TECHNOLOGY AND ASSOCIATED PROBLEMS

Some measures taken in response to the problem of leakage of sensitive data include implementation of tools to automatically delete P2P software from PCs [5, 6] and implementation of batch processing thin client sensitive files on servers [7], implementation of authentication mechanisms for handling of sensitive files [8], prohibiting of copying of sensitive files onto external media [9], recording of access logs for sensitive files [8, 10]. In spite of this, there appears to be no end in sight for the problem of leakage of sensitive files via P2P. Unfortunately, this type

978-0-7695-5052-7/13 $26.00 © 2013 IEEE
DOI 10.1109/NBiS.2013.17

of exit management for sensitive data is insufficient as a strategy to stem the tide of data leakage at the present time. Moreover, another problem exists in that these types of exit management measures all significantly reduce usability.

On the other hand, a method has been proposed as a countermeasure to files leaked to external networks, in which these files that have been leaked to external networks are retrieved by search. [11] This method preemptively adds an e-signature to the confidential file designating it as sensitive information, and using telecommunications equipment of the telecommunications provider a search is conducted to determine if the file marked with the embedded signature exists in the communication data, and forcibly deletes such files if they are detected. [11] Further, if a similar mechanism is mounted in the gateway where an internal network is connected with an external network, that can be operated as a method of exit management for preventing confidential files where the signature was added from leaking out to an external network. [12] However, a fundamental problem exists in these methods, in that the confidential file cannot be inspected for signatures in cases such as when the encrypted communication is conducted end-to-end. Moreover, the very presence of an e-signature can result in identifying the file to others as containing confidential information.

Another approach being considered is the method known as P2P network poisoning, which obstructs the access to confidential files that leak on the P2P network. This method obscures a genuine sensitive file by flooding the network with a large amount of dummy files of a similar form (for instance, same file name) when a sensitive file leaks onto a P2P network and uploading the genuine sensitive file in the dummy file. As a result, the possibility that a P2P user can access the genuine confidential file is suppressed, and the diffusion of the leaked sensitive file will be controlled as a result. However, this method is a retroactive measure undertaken after it is noticed that the secret file leaked. In the case of viruses such as Antinny, chances are great that discovery of the leak will be made too late, since the file uploads occur without regard for the intentions of the user. Moreover, uploading a large number of dummy files will have the effect of informing others that a sensitive file of the same name has been leaked onto the P2P network.

III. The Proposed Method

3.1 Concept

In the proposal method, diffusion of the leaked confidential file on the P2P network is controlled by automatically deleting the confidential file from the PCs on the P2P network. That is, this method is positioned as a countermeasure to contain sensitive files that have already leaked from an internal network to the outside. It is expected that the leakage risk and the damage of the sensitive information can be decreased further by using this method in combination with the exit management (measures to prevent sensitive files from leaking from an internal network to the outside).

In order to achieve deletion of the leaked confidential files, our method focuses on characteristics of the anti-virus software [14, 15, 16] to automatically isolate and to delete a file that matches the characteristic pattern of a virus. The anti-virus software is software that inspects targeted files to determine whether they contain certain patterns characteristic of viruses, and if these patterns are detected, automatically isolates and deletes the files in question. In this method, the "characteristic pattern" of a virus is preemptively embedded in the confidential file on the manager side (the bit string embedded is selected to be detected by anti-virus software, and does not contain any functions of the virus). By the embedding the characteristic pattern in the confidential file, it can be expected that even if a sensitive file is leaked the installed anti-virus software on the PC that receives the file will flag the characteristic pattern embedded in the file as a virus, and automatically delete the confidential file.

Using our method, there is reason to expect that the damage from leaks can be kept to a minimum even if sensitive material is leaked to a P2P network, because any such files can be automatically deleted on the receiving PC. Moreover, as seen in the documents [2, 3, and 4], with the ongoing increase in damage from viruses now and in the past, enterprises and individuals can be expected to increase their installation of anti-virus software, so this method can surmised to enjoy relative ease of implementation. Particularly on P2P networks where underground files tend to be disseminated, since the danger of encountering illicit files is high, P2P users (excluding those users with low security awareness who are at the source of sensitive information leakage) are thought to tend to have a higher rate of anti-virus software implementation than average users [17].

From the perspective of "deleting the leaked sensitive file", this method has the same concept as the method of document [11]. However, this method functions effectively even if the leaked confidential file is encrypted, unlike the method of identifying leaked sensitive files using the telecommunications equipment of the telecommunications provider. In file transfers using P2P networks as well, there are some cases where files are encrypted prior to sending. [18] Therefore, even if the communication is observed at the gateway etc., contents of the communicated file cannot be inspected. On the other hand, even if such files are encrypted during transfer through the communication circuits, it can be assumed that a third party when obtaining the leaked confidential files will need to decrypt them in order to read them, at which point the embedded characteristic patterns will be detected by the anti-virus software and the file automatically deleted.

3.2 Operation method

Figure 1 shows the mechanism of the proposed method. This method is composed of three major phases (1) - (3). Now among these, (2) and (3) are carried out in precisely the identical fashion as with conventional anti-virus software.

(1) Embedding of characteristic pattern by the manager

The manager of the sensitive file selects a characteristic pattern from a characteristic pattern list for anti-virus software and provided by the anti-virus software vendor, (hereafter, pattern file) and embeds it in the confidential file.

Once the characteristic pattern has been embedded, anti-virus software will react to the sensitive file with the embedded characteristic pattern by automatically isolating and deleting it. Therefore, even if this confidential file is leaked, the confidential file in question will be deleted by the anti-virus software, and thus it is expected that the diffusion of the secret file can be controlled as a result in PC whose users have installed anti-virus software.

However, because the anti-virus software is operating, the same confidential file must then also be deleted in PCs in the organization that legitimately created the file. To deal with this, the manager generates a whitelist (for instance, file name of the secret concerned file) at the same time, and notifies the anti-virus software of PCs in the organization of this whitelist. Even if a file is one in which the anti-virus software detects a characteristic pattern, it will not be deleted as long as it is registered to the whitelist. Further, we intend to continue work in the future to develop concrete methods for generating the characteristic pattern, the method of generating the whitelist, and the method of embedding the feature pattern.

(2) Distribution of pattern file by anti-virus vendor

The anti-virus vendor creates and manages the pattern file of the existing virus. The pattern file is updated every time the new type of virus is discovered. The latest pattern file is distributed to each client PC through the network.

(3) Deletion of confidential file in target PC using anti-

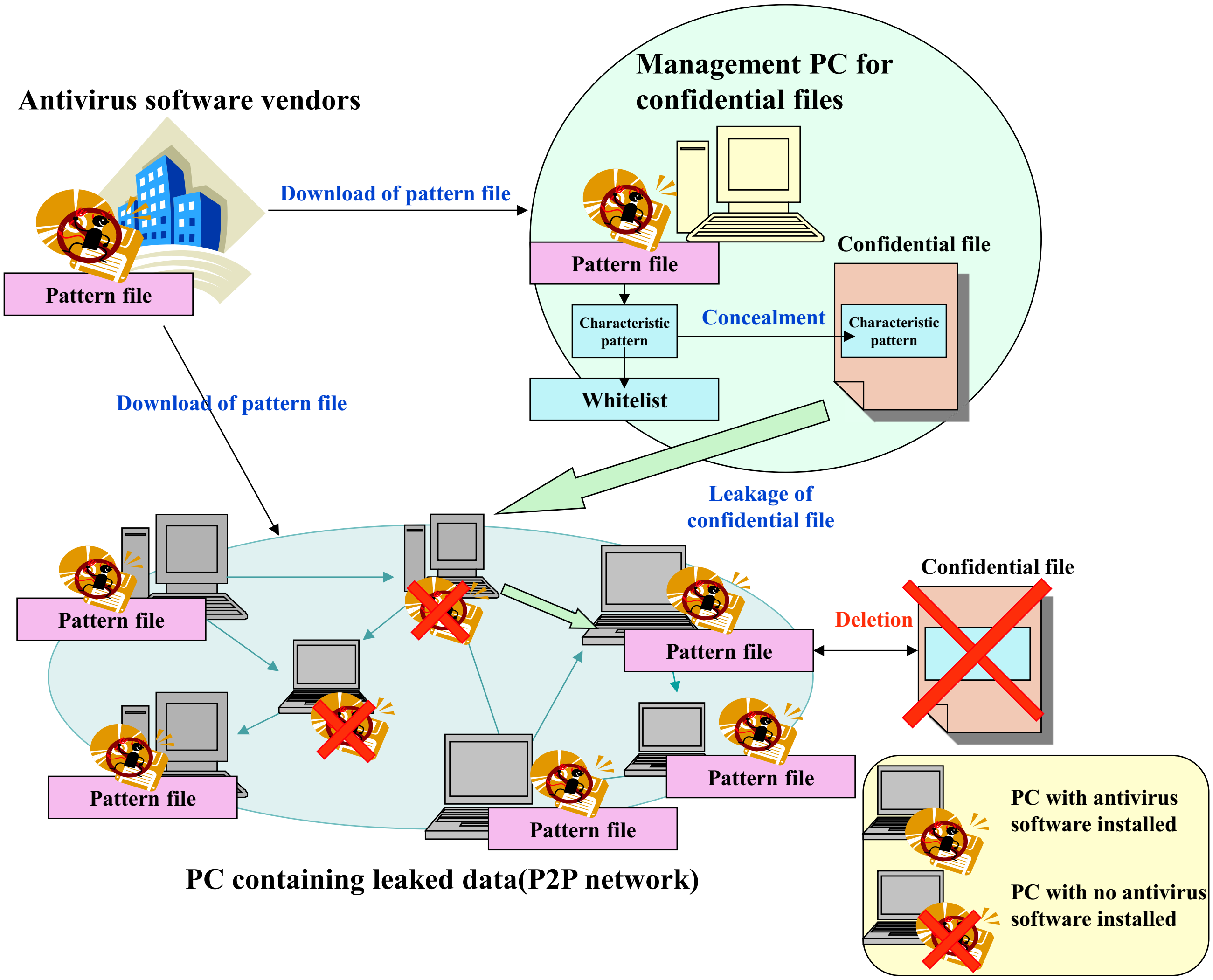


Figure 1 Mechanism of proposal method

virus software

PC in which the anti-virus software is installed inspects the virus of all received files by using the pattern file. If any of the characteristic patterns from the pattern file are detected, the file in question is categorized as a virus file, and is automatically isolated and deleted.

3.3 Measures against leakage of confidential files to non-users of anti-virus software

Because this method is an information leakage countermeasure that uses anti-virus software, confidential files will naturally not be deleted in PCs of users who have not installed anti-virus software. For this reason, confidential files are leaked and fall into the hands of these users. Here we propose a method of improving the deterrent effect to control confidential file leakage by focusing on the file retrieval function of P2P, and providing a mechanism that causes users who have not installed the anti-virus software to hesitate in the download of the confidential file for the in this method.

Many files are uploaded to a P2P network, and when the P2P user has acquired summary information in a file called the "key" file, the download of that file becomes possible. The P2P software regularly exchanges keys with surrounding PC, and at any given time several thousands or tens of thousands of keys are always saved in each P2P user's PC. [18] The user selects a desired file from among this large volume of keys, and downloads it. In other words, the user needs to search for the desired file to be obtained from among these keys. Here, when the user searches for a file, a mechanism is employed to cause the user to hesitate to download the file.

Specifically, the function of the anti-virus software is expanded to add a feature that not only isolates and erases file A if it determines that A downloaded by the P2P software is a virus, but also uploads a warning file named "A-is-a-virus.txt" or similar. Let's assume that secret file "Confidential.pdf" has now leaked onto the P2P network. Here, the manager in the organization has previously embedded the characteristic pattern under Confidential.pdf. Therefore, and when the user who is using the anti-virus software with an additional function downloads it, Confidential.pdf is deleted from the user's PC, and in addition the warning file named "Confidential-pdf-is-a virus.txt" is uploaded to the P2P network. As a result, in addition to the Confidential.pdf file key being distributed on the P2P network, the file key "Confidential-pdf-is-a virus.txt" is also distributed. That is, on the PC of the user who is participating in the P2P network, the two keys "Confidential.pdf" and "Confidential-pdf-is-a virus.txt" are stored together. Therefore, when the P2P user retrieves the file "Confidential.pdf" relying on the file name search, the warning file named "Confidential-pdf-is-a virus.txt" will appear along with the file "Confidential.pdf" from among the keys on the PC. Naturally, the anti-virus software with the additional feature will not limit itself to uploading warning files only when confidential files have been

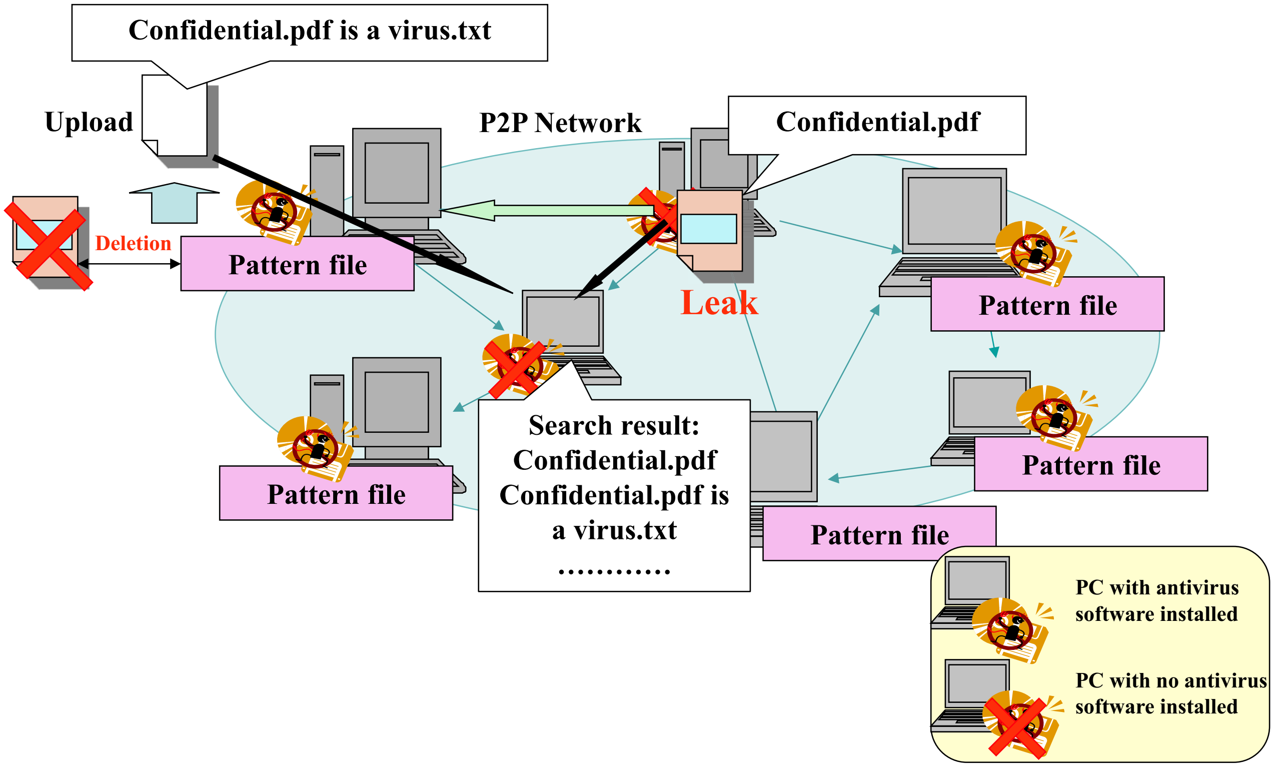


Figure 2 Measures for non-users of anti-virus software

discovered, but will also upload warnings in the event that real virus files have been discovered. Because so many illicit files are present on P2P networks, the fact that the key search resulted in discovery of a warning file stating "Confidential-pdf-is-a-virus.txt", users will be convinced that (although it might be a confidential file, probably it is more likely that) chances it is an actual virus file are high. Therefore, it is thought that many users will hesitate to download Confidential.pdf when the warning file is discovered by key retrieval.

All these elements are brought together as shown in Figure 2. By utilizing this type of mechanism, many users belonging to the P2P network will have anti-virus software installed, and those that don't use such software can be expected to hesitate to download the leaked confidential file.

3.4 Management of characteristic pattern by anti-virus software vendor

In 3.2 we indicated the operation method in which the manager of a confidential file would embed "a characteristic pattern of an existing virus", but it is also possible for the confidential file manager to designate and apply for registration with anti-virus software vendors of a new bit string as a characteristic pattern, and to then embed this in the confidential file.

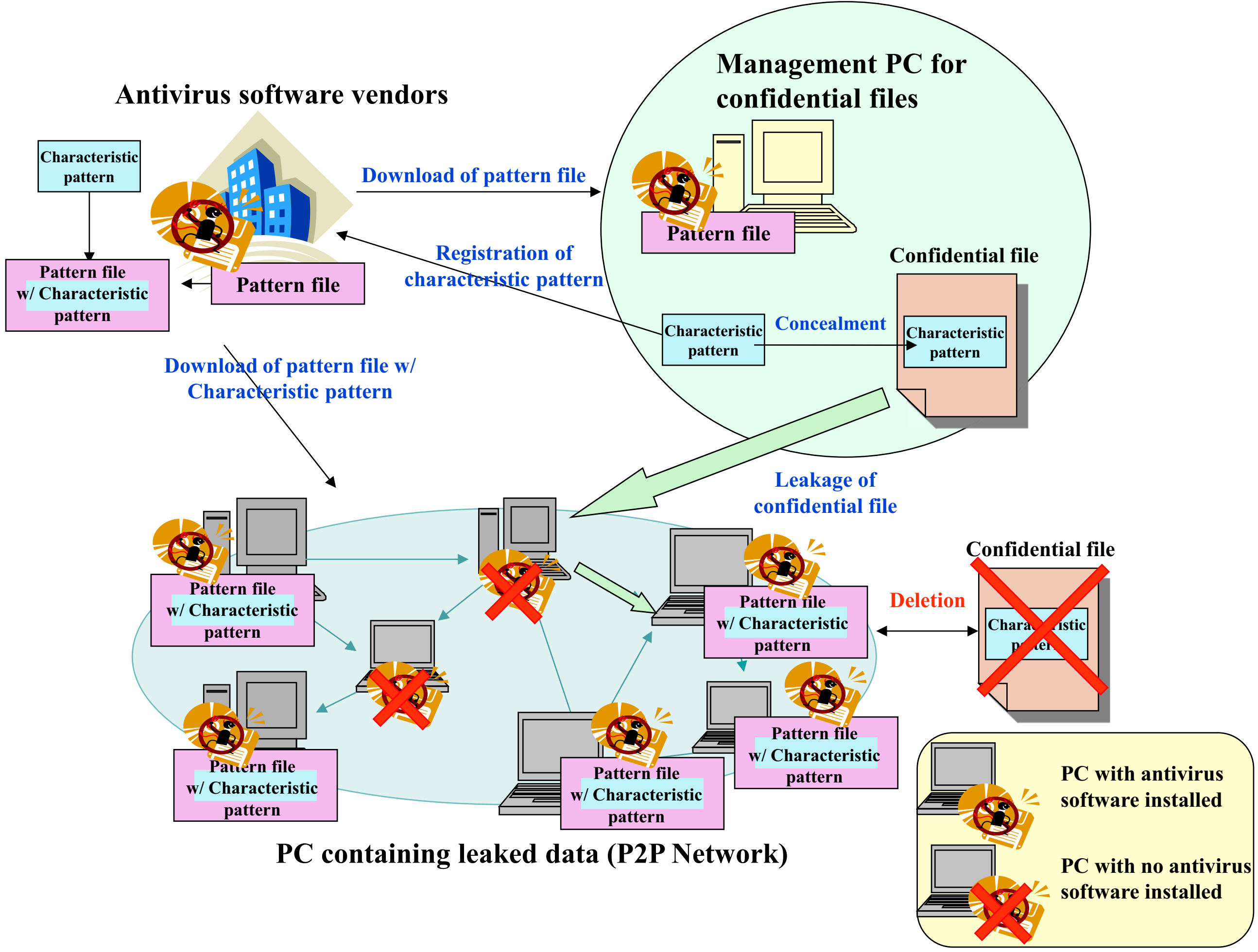
In this case, the new characteristic pattern sent to the anti-virus software vendor by the manager of the confidential file is added to update information on the pattern file, and supplied to the anti-virus software of all PCs. However, this pattern is not included in the updated information of the pattern file supplied to the anti-virus software for PCs in the organization with the confidential file where the feature pattern was registered. As a result, this file will not be identified as a virus in PCs in the organization with this confidential file in its possession. Figure 3 demonstrates this.

IV. THE BUSINESS MODEL'S IMPROVEMENT

4.1 Problems

Problem (1):

Recently, the leakage of the confidential files through the



Antivirus software vendors
Characteristic pattern
Pattern file w/ Characteristic pattern
Pattern file
Download of pattern file
Registration of characteristic pattern
Management PC for confidential files
Pattern file
Confidential file
Characteristic pattern
Concealment
Characteristic pattern
Download of pattern file w/ Characteristic pattern
Leakage of confidential file
Confidential file
Deletion
Characteristic pattern
Pattern file w/ Characteristic pattern
Pattern file w/ Characteristic pattern
Pattern file w/ Characteristic pattern
Pattern file w/ Characteristic pattern
Pattern file w/ Characteristic pattern
PC containing leaked data (P2P Network)
PC with antivirus software installed
PC with no antivirus software installed

Figure 3 Method for new characteristic pattern registration

P2P file sharing software (hereafter, P2P) etc. has become a problem. Due to the nature of P2P networks, it is difficult to eradicate sensitive information from a P2P network once that information has been leaked to that P2P network.

Problem (2):

In current Internet businesses, one popular business model involves offering content to users free of charge, and instead securing earnings through advertising income. Because most current business-oriented anti-virus software is sold to users in a fee-based format, users experience an economic burden.

4.2 Solutions

With respect to (1), it is difficult to achieve actual erasure of sensitive information from a network once it has been leaked, but using alternate methods, one can make it impossible for the end user to view the sensitive data.

To achieve this, a special signature from the confidential information file is added to the virus definition file of the anti-virus software. Furthermore, a pattern (hereafter called a pseudovirus) is embedded in the confidential information file that will be detected as a virus by anti-virus software. As a result, even if the sensitive information is leaked, and even if this information falls into the hands of end users, the file will be judged to be a virus and automatically deleted (isolated) by the anti-virus software.

Here, in the enterprise in possession of the confidential information, in order to prevent the confidential information file from being mistakenly deleted in the internal corporate LAN, the embedded virus pattern is managed as a whitelist.

Here it is important to note that in this proposed method, the problem of (2) is simultaneously resolved.

When this proposed method is operated, the phase in which the enterprise requests the anti-virus software vendor to add the signature of the sensitive information to the definition file of the anti-virus software (Or, when the pseudovirus pattern is embedded in the sensitive information) becomes indispensable. In that case, the anti-virus software vendor charges the enterprise according to the number of files and the period, etc.

As a result, it becomes possible for the anti-virus software vendor to increase earnings through revenues from the enterprise, and to offer the user the anti-virus software free of charge. From the perspective of the anti-virus software vendor, the greater the share of the market enjoyed by their product, the more effective the deletion of data will become, and the number of orders received from enterprises will be maximized. That is, by switching to this type of business model there will be no change in the psychology or incentive on the anti-virus software vendor side to maximize the number of users employing their products, and other than the source of revenues, they will be able to carry on their business activities as before.

4.3 Effects

The proposal technique, aims at:

- protection of sensitive information, and
- creation of a new business model (offering anti-virus software free of charge), and

proves to be a method that is able to achieve both above aims.

4.3.1 Innovation

By prior registering of the signature of the sensitive information in the anti-virus software on the client side, in the unlikely event that sensitive information is leaked, it will be automatically deleted on the client side (there is no time lag between noticing the leak and initiating countermeasures).

Rather than selling anti-virus software to earn profits, earnings can be realized through services offered using the anti-virus software, making it possible to create a new business model.

4.3.2 Originality

The concept of deleting the leaked secret information on the ordinary user side is unique to this proposed method.

It is a solution that accomplishes both the protection of the secret information and the creation of a new innovative business model simultaneously, making it highly original.

4.3.3 Utility

Exit management is the generic form of countermeasure for leakage of confidential information. However, the current situation is that exit management is not sufficient as a countermeasure to fully prevent leakage of information. Because the proposed method accomplishes deletion at the remote location where the leaked data is received, it solves the problems that mere exit management is unable to completely deal with.

Because there is a high possibility of malware infection on the P2P network, the likelihood of P2P users installing anti-virus software is also high. On the other hand, there are many exposure viruses including Antinny that leak information onto the P2P networks. Therefore, it can be expected that this will be especially effective in properly deleting the confidential information leaked to the P2P network from the P2P user's PC in comparison with users who expect to profit.

By operating the proposed method, there is no addition of correction or operation on the user side PC or software.

4.3.4 Feasibility

As long as the mechanism that signature to the sensitive information is made on the anti-virus software vendor side (or the pseudovirus pattern is embedded in the sensitive information) can be arranged, ease of implementation can be expected for this method, because it doesn't differ from the function of other conventional anti-virus software operating schemes (which obtain regular automated updates of the latest virus definition files and delete (isolate) data with matching signatures if found to be present or intruding on PCs).

If the whitelist in the pseudovirus pattern is leaked, this gives rise to the fear that it might be used by illegal operators. Therefore, the whitelist of pseudovirus patterns becomes sensitive information in the enterprise as well. For enterprises, this means the current method will add one piece of sensitive information to manage, but this is a small price to pay for the benefits to be gained.

4.3.5 Further recommending aspects

The anti-virus software can be offered to the user free of charge because the profit is obtained through services using the anti-virus software, and thus a further increase in the number of users and expansion of market share can be expected. Here, new characteristic patterns sent to the anti-virus software vendor by the confidential file manager will be added to the updated information in the pattern file, and supplied to the anti-virus software of all PCs. However, for the PCs in the organization that registered the characteristic pattern and that possesses the confidential file, the anti-virus software provided and containing updated pattern list will not contain the pattern for that particular file. As a result, the file in question will not be identified as a virus on PC in the organization that possesses the confidential file. Figure 3 shows this.

V. SUMMARY

In this paper, we proposed a method to control the distribution of confidential files on P2P networks by embedding the characteristic pattern to trigger anti-virus software to delete the file, thus controlling the spread of the confidential file on the P2P network. It is expected that the leakage risk and the damage of the sensitive information can be kept to a minimum by using this method in conjunction with exit management (measures to prevent secret files from leaking from an internal network to the outside) because this method deletes leaked confidential files on the remote PCs where they are received. In particular, information leaks via P2P result in rapid dissemination of the leaked confidential files, making the need for “a method that can delete leaked confidential files” very great indeed. Moreover, it can be said that one great advantage of this method is its ability to delete leaked confidential files easily even when the communication route utilizes encryption.

Effective methods for embedding characteristic patterns and the generation method for whitelists in the proposal method etc. will be examined in future studies. Moreover, we intend to verify effectiveness and validity when the proposal method is implemented in simulations.

REFERENCE

[1] W32.HLLW.Antinny | Symantec:
https://www.symantec.com/security_response/writeup.jsp?docid=2003-080817-4045-99

[2] The Ministry of Public Management, Home Affairs, Posts and Telecommunications: telecommunication white paper, 2001 (in Japanese):
http://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h13/index.html

[3] The Ministry of Public Management, Home Affairs, Posts and Telecommunications: telecommunication white paper, 2003 (in Japanese):
http://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h15/index.html

[4] The Ministry of Public Management, Home Affairs, Posts and Telecommunications: telecommunication white paper, 2005 (in Japanese):
http://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h17/index.html

[5] IPA/ISEC in JAPAN: virus and UCA incident report for Mar. 2006:
https://www.ipa.go.jp/security/english/virus/press/200603/E_PR200603.html

[6] Software Sales | Global IT Solution | Fuji Infox-net Co., Ltd.:
http://www.infoxnet.co.jp/eng/global/software.html

[7] Sun Ray Clients and Oracle Desktop Virtualization Clients:
http://www.oracle.com/us/technologies/virtualization/sun-ray/overview/index.html?origref=http://www.oracle.com/us/technologies/virtualization/sun-ray/overview/index.html

[8] File Access Control:
http://en.quality.co.jp/products/DKSP/index.html

[9] USB Copy Protection: USB Drive, USB Disk, Flash Drive Copy Protection Software for Windows 7/8/XP:
http://www.kakasoft.com/usb-copy-protect/

[10] PISO | Insight Technology, Inc.:
http://www.insight-tec.com/products-2/piso?lang=en

[11] Slashdot: The development of the technology that the Ministry of Public Management, Home Affairs, Posts and Telecommunications eliminate a specific file from the net is planned:
http://slashdot.jp/security/article.pl?sid=06/08/23/0055230

[12] Yoshihiro Yano, Kiyoko Nishimura: “Information leakage prevention system”, 2003-333420 patent pending, 2005-100123 patent pending

[13] A Survey of Peer-to-Peer Network Security Issues:
http://www.cse.wustl.edu/~jain/cse571-07/ftp/p2p/index.html#poison

[14] Titanium Antivirus + Antispyware – Internet Security – Trend Micro USA:
http://www.trendmicro.com/us/home/products/titanium/antivirus-plus/index.html

[15] Norton AntiVirus 2013 – Virus and Spyware Protection | Norton:
https://us.norton.com/antivirus/

[16] Anti virus Software, Virus Protection Scan, Antivirus Plus 2013 | McAfee
http://home.mcafee.com/store/antivirus-plus

[17] General corporate judicial person Content Overseas Distribution Association: “The investigation report about use of file exchange software (the news flash version)”, 2011

[18] Isamu Kaneko: “The Technology of Winny”, Ascii Corporation, October 2005