Original
Received February 2, 2012
Accepted for Publication April 24, 2012

Method for Detecting Remaining Files that Contain Copied Data by Monitoring Clipboard and Directory

Chikako ISHIZAWA※, Ryo SATO※ and Makoto NISHIDA※※

※Department of Computer Science and Engineering, Graduate School of Engineering and Resource Science, Akita University, 1-1 Tegata Gakuen-Machi, Akita 010-8502, Japan
※※Akita University, 1-1 Tegata Gakuen-Machi, Akita 010-8502, Japan
E-mail : ishizawa@ie.akita-u.ac.jp

This paper focuses on information leaks caused by the human mistake of forgetting to delete copied files from portable storage media. We propose a processing method that obtains relevant logs and detects remaining files in order to determine the remaining files that contain copied data, even if only a part of the data in the file on the portable storage medium has been copied. The proposed processing method for obtaining the relevant logs monitors the state of both clipboard and directory and creates logs. Information recorded in a log includes the type of change that occurred in a directory, the date of the recording, and the name and path of the changed file on a personal computer. The proposed processing method for detecting the remaining files comprises three steps. First, a log entry for the clipboard is searched sequentially, starting at the head of the log file. Next, the entry indicating the path and name of the file containing the copied data is identified. Finally, it is ascertained that files containing copied data remain if an entry indicating the deletion of such files is not found. Various file operations were tested on Microsoft Windows XP and Windows 7. Our experimental results suggest that the copy operation performed on a portion of data was distinguished by using the change logs of the clipboard and the directories. The remaining files that contained copied data were correctly detected.

Key Words : Information leaks, Log analysis, File operation, Clipboard, Directory

1. INTRODUCTION

An increasing number of people are carrying data on portable storage media such as universal serial bus (USB) flash drives. Such portable storage media are small and easy to lose. In order to prevent data on the portable storage medium being read by any (unauthorized) person, prudence dictates that such data be encrypted [1].

Data leaks from portable storage media to the Internet via personal computers occur [2][3]. For example, information can be leaked if a personal computer to which a file is copied is infected with a computer virus and the copied file is not deleted after use. Even if a file is copied to a virus-free personal computer and is used correctly, remains of the file pose a data leak threat in portable storage media. In order to prevent such leaks, USB flash drives that do not allow the files they contain to be copied were developed [4]. However, such a USB flash drive cannot be used when a file needs to be copied. Therefore, it is essential that those files that were copied from the portable storage medium be identified and subsequently deleted.

Many techniques that trace a secret file in a personal computer use the hooks of file system API (application programming interface) [5]-[8]. Although an API hook can be used on Linux or Windows XP, it cannot be used on Windows Vista and Windows 7 [9].

We have already proposed a method for detecting the remaining files copied from a portable storage medium in an effort to prevent the human mistake of forgetting to delete the copied file [10]. In our previously proposed method, copied files are identified based on the change history of a state of a directory in a storage unit.

The previously proposed method consists of three integrated processes: a process to obtain logs of change in a directory, a process to detect the removal of the portable storage medium, and a process to detect the remaining files. The remaining files were detected accurately when the files in the portable storage medium were directly copied to the personal computer. More specifically, even if the copied file was updated, moved, deleted, or renamed, it was still possible to detect the remaining files. However, since the state of the directory did not change when some data in a file were copied, the remaining files were not able to be detected.

This paper proposes a novel processing method that obtains relevant logs and detects remaining files by using the copy operation on a part of the data. The proposed method uses the change history of a state of a directory and clipboard. The operating system of the personal computer used in this study was Microsoft Windows. A computer virus, a rootkit, and reading data without using an operating system do not naturally exist in a personal computer and a portable storage medium. We assume that a user wants to prevent leakage and does not deliberately alter a log.

This paper consists of six sections. Section 1 outlines the background and provides the objective of this study. Section 2 explains the state changes that occur in a personal computer when the copy operation is executed on a portion of data. Section 3 outlines