Digital forensics in the context of cloud computing data storage services

Paweł Olber[1]

[1] The Police Academy in Szczytno, p.olber@wspol.edu.pl

Summary

The role and capabilities of digital forensics in the cloud computing environment still remain an unsolved and insufficiently explored scientific area. The growing popularity of cloud-based data storage services makes recovery of evidence on the Internet a major challenge for forensic IT exminers. Existing legal barriers in most cases prevent direct access to cloud stored resources for the purpose of recovering data. In many situations, assistance from foreign law enforcement and judicial authorities is required. Findings made during forensic IT analysis that allow reconstructing user's activities and indicate the act of taking advantage of cloud computing data storage is a justification for undertaking international cooperation. Therefore, it is important to know potential sources of information that enable the reconstruction of the activity of the user in cloud computing services, as well as the awareness of the current state of knowledge and scientific research in this field.

Key words: computing cloud, file hosting service, digital forensics, IT examinations

Introduction

Cloud computing technology provides an unlimited access to computing resources such as servers and storage that are delivered with a minimal interaction with the service provider. Undoubtedly, one of the main advantages of cloud solutions is easy and quick access to resources that may be available on demand from anywhere in the world (Mell & Grance, 2011).

The growing popularity of cloud computing data storage services and the possibility of using these solutions for criminal purposes call for efforts on the part of scientists and practitioners in the field of forensic IT (Sharma, Arora, Sakthivel, 2018). One of the reasons for this situation is that cloud computing specifics exacerbate the difficulties in recovering and analysing digital evidence (Samy et al., 2018). A detailed review of the existing challenges and problems of digital forensics in the aspect of obtaining and examining data from cloud computing has been documented in the Report of the American National Institute of Standards and Technology: NIST (Herman et al., 2014). A large number of open challenges presented in the aforementioned document confirms that the role and possibilities of digital forensic in the cloud computing environment remain an unsolved and insufficiently explored area of scientific research (Martini, Choo, 2014). In order to overcome the persisting problems some scientists point to artifacts remaining in the memory of the data carriers recovered for the purpose of forensic examination as they may be important in the context of resources stored in the cloud computing environment.

The aim of the article is to present and summarise the current state of knowledge on the traces of cloud storage services users activities remaining in the memory of digital data carriers.

In view of the purpose of the article, the following research question was defined: what kind of traces of user activity of cloud computing data storage services stay in the memory of digital data carriers?

In connection with the posed research question a hypothesis was adopted according to which information on the activity of a user of Internet cloud remote file hosting service are stored in the memory of digital data carriers. That information can help to determine the content of resources in the cloud computing environment.

The research hypothesis had determined the structure of this publication. The results of previous research on forensic computing and cloud storage services were presented in separate parts of the article. Then, the existing possibilities of identifying traces of user activity in Dropbox and Google Drive were described in detail.

The research was carried out using the monographic method with critical literature analysis. In addition to that an empirical method was also applied to verify published results of Dropbox and Google Drive services research. Finally, the conclusions were formulated.

Literature review

Cloud computing services in the context of digital forensics are the subject of many scientific research undertakings. However, few scientists have studied the possibility of revealing traces of activity of users of cloud data storage services in the memory of local devices.

Dehghantanha and Dargahi (2017) endeavoured analyses of two services: CloudMe and Qihoo 360

Yunpan. CloudMe is a European file storage service established in 2012, owned by CloudMe AB. It offers secure cloud storage, file synchronization and remote data management client software. Qihoo 360 Yunpan, on the other hand, is a Chinese file storage service distinguished by the largest free online drive in the world. In the course of research, the authors used various operating systems: Windows 8.1, Android KitKat 4.4.2 and Apple iOS 8.0. As a result, they found that the analysis of the disk content, RAM memory, internal memory of mobile devices and secured network traffic allowed for the reading of authentication data, names of devices, as well as files stored on the on-line drives. These studies show that the use of cloud computing services leaves many traces in the memory of digital data carriers. They also demonstrate that the traces are formed as a result of performing various operations on files located in the cloud, but also during the installation / uninstallation of the client application. The researchers also highlighted the issue of data encryption in each of these services. They found that in the case of the Qihoo 360 Yunpan service, data security during the transfer was very low. The authors easily gathered and analysed network traffic data to recover required evidence. CloudMe protect users' privacy much better. Uninstalling the CloudMe client application, although it retained the configuration files, did not change the registry keys and uses data encryption during the transfer.

Mohtasebi et al. (2017) conducted studies of three more services: SpiderOak, JustCloud and pCloud. The researchers located and described various forensic artefacts connected to the use of these services by three web browsers: Internet Explorer, Mozilla Firefox and Google Chrome, as well as client applications installed on Windows 8.1 and an iPhone 5S with iOS 8.1.1. The data revealed and retrieved during the study included e-mail addresses, the ID and name of the created account, as well as the names of uploaded and downloaded files.

Teing, Dehghantanha, and Choo (2018) investigated the possibility of revealing traces of CloudMe online disk user activity. Their analyses covered various operating systems: Windows 8.1 Professional, Linux (Ubuntu 14.04.1 LTS distribution), Apple Mac OS X Mavericks 10.9.5, as well as mobile devices: iPhone 4 with iOS 7.1.2 and HTC One with Android KitKat 4.4.4. The research included the installation / uninstallation of CloudMe client applications, as well as uploading, downloading, browsing, deleting, synchronizing and sharing resources. Those authors emphasise that during examinations of the CloudMe service, attention should be paid to the database files: Cache.db, db.sdb and logs, application configuration files, as well as the web browser cache. The analysis of web browsing history allowed them to identify unique web addresses that helped to determine the activities performed by the CloudMe user, such as logging in / logging out, accessing files / folders and data downloading. Even though the connection to CloudMe via the web browser was encrypted, the researchers recovered the contents of the application's root directory from the browser cache. The application directory included user files, metadata, and OpenSearch description files containing timestamp information and passwords for shared resources.

Ahmad et al. (2020) who illustrated the existing trace identification capabilities in Windows 7 Ultimate described the existing possibilities of revealing the traces of pCloud service users' activities in volatile RAM. The scientists verified the possibility of revealing information on user interaction with the online drive based on various scenarios including data transfer with opening and viewing of its contents. In the study, the content of the volatile RAM memory as well as the cache of the Google Chrome web browser were analysed. Consequently, in the case of the pCloud service, it was possible to read all credentials and all information about files stored on the file hosting service. The study in question confirms the results of the previous experiments (Dargahi, Dehghantanha, Conti, 2017), whose authors focused on the pCloud service and presented the possibility of detecting many traces in the following operating systems: Microsoft Windows, Android, iOS and Linux. They showed that it was possible to read the credentials of pCloud users and information about files stored on the disk.

According to the above literature review the analysis of the content of digital data carriers allows (in principle) the disclosure of information indicating the use of cloud computing data storage services by the user of the device / system and additional information, such as: the content of virtual resources, user identification data and credentials. The research projects focus on several different services, including: CloudMe, Qihoo 360 Yunpan, SpiderOak, JustCloud, pCloud, which do not exhaust the catalogue of existing remote data storage solutions. This list may be extended, for example, by two popular services: Dropbox and Google Drive, which were the subject of Horsman's research (2020). Due to the widespread popularity of these services and the timeliness of the study the results will be presented and verified later in the article.

Methodology

In order to make a preliminary assessment of the possibility of detecting traces of the Dropbox and Google Drive users activities, an analysis of the results of the above-mentioned studies was carried out. Horsman's experiments (2020) used the Microsoft Windows 10 operating system and the Google Chrome web browser (version 67.0.3396.99). The web browser cache was read with Nirsoft's ChromeCacheView v1.77. In the course of repeating the examination procedure of Horsman (2020), a newer version of the Google Chrome web browser (version 89.0.4389.90) and current Nirsoft programs: ChromeCacheView v2.25, ChromeHistoryView v1.42 were used.

Dropbox service

The research conducted by Horsman (2020) shows that in case of using the Dropbox service via the Google Chrome web browser a file called www.dropbox.com.html is saved in the web browser's cache. This file does not open in the web browser window. Source code analysis is required. That procedure allows reading basic information about the service user and identification of the contents of the virtual disk. The information stored in the www.dropbox.com.html file comprises the following data:

- User name (tag: „display_name":),
- Account identifier (tag: „id":),
- User's email address (tag: „email"),
- URL address of user's profile picture (tag: „photo_circle_url":).

A fragment of the code recorded in www.dropbox.com.html file is presented in Figure 1.

Main window of the program

The www.dropbox.com.html file contains additional records that reflect user's activity on the Dropbox homepage. By default, on the homepage of the drive there is a list the last actions taken by the user. Each entry in the list of recent activities is incorporated in the structure of the www.dropbox.com.html file. The list of recent activities is saved under the "recent_activities" tags. The tags contain the interaction time (e.g. opening a folder / file), which is saved in UNIX format. If the user unfolds the last event that contains one or more files (usually graphic ones), a preview of the image is displayed for each file. That results in saving files with very specific names in the memory of the web browser, for example: size = 100x100size_mode = 4.jfif. In the event of finding files with similar names in the browser cache it should be concluded they indicate the above form of activity have taken place on the Dropbox drive.

Preview of drive contents

In the event that the Dropbox user displays all the files saved on the disk, the URL address is saved in the history of the web browser: https://www.dropbox.com/home. Additional information is saved in the www.dropbox.com.html file. In order to determine the entire contents of the user's virtual disk, read the information marked with "event_type": tags.

Preview of file contents

Horsman's research (2020) shows that when the graphic file content is displayed on the online drive it is cached by the web browser (stays in the cache) similarly to the situation when a file is previewed in the main window. The name of the file is very distinctive because it contains information about its size: size = 32x32size_mode = 5.jfif. The original file name is saved in the URL address and can be read from the browser's history: www.dropbox.com/home?preview=FILENAME.png. Additional information about viewed image files is saved in a text file named: is_xhr = trueactivity_context = 3activity_context_data =% 2FFILENAME.txt.


```
\{\"LOCALE\": \"GB\", \"prompt_ha_hiding\": true, \"_viewer_properties\":
{\"display_name\": \"GREY JOY\",
\"can_moderate_comments\": false,
\"deprecated_first_user_in_the_cookie_id\": 77837232,
\"is_reseller_session\": false,
\"is_team_assume_user_session\": false,
\"is_assume_user_session\": false,
\"_user_data\": [\{"initials_url\":\"https://ac.dropboxstatic.com/account_photo/get_initials?initials=GJ
\"user_root_permissions\":
\"edit\", \"has_never_set_password\": false,
\"id\": 77837232, \"sso_required\": false,
\"display_name\": \"Grey Joy\",
\"_authed\": true, \"home_ns_id\": 126648836,
\"lname\": \"JOY\", \"role\": \"personal\",
\"is_email_verified\": true,
\"fname\": \"GREY\",
\"cdm_path\": \"\",
\"email\": \"grey.joy@gooqlemail.com\",
\"is_paper_disabled\": false,
\"account_id\": \"dbid:AAAAz7mAv7FTO-BYzKWNpC1uj3FaJ1wVfBA\",
\"is_cdm_member\": false, \"nid\": \"01529833775757704936\",
\"is_dropbox_admin\": false,
\"paid\": 0,
\"root_ns_id\": 126648836,
\"photo_url\": null,
\"is_team_admin\": false,
\"familiar_name\": \"GREY\",
\"is_team\": false,
\"photo_circle_url\":"https://dl-web.dropbox.com/account_photo/get/dbaphid\%3AAACJ-_rJyCoDzFXXbB8MDaBqtS
\"DEFAULT_ROOT_NAME\": \"Dropbox\",
\"PERSONAL_ROLE_STRING\": \"Personal\"}
\}
```


Fig. 1. Fragment of www.dropbox.com.html file code.

Users' comments

A Dropbox user has an option of adding comments to files stored on the drive. Comments can also be added by other users who have access to the specific resource. Making a comment causes saving additional information, marked with the "comment": tag, in the text file is_xhr = trueactivity_context = 3activity_context_data =% 2FFILENAME.txt. If a third party responds to the comment, the account name of that third party will be saved in the following tags: display_name, lname and fname. However, the metadata does not include the e-mail address and account ID of the person responding to the comment, so it does not seem possible to identify the actual third party account.

File sharing and deletion

When a user goes to the tab containing the resources shared with other users it is reflected in the browsing history of websites: https://www.dropbox.com/share. Viewing a list of shared resources does not cache any additional information in the web browser. The same is true for deleted files. The Dropbox service stores deleted files for 30 days, and during that period the user can restore and view them. However, there will be no records in the web browser's cache that can be assigned to the data wipe action.

Results of author's own research

Upon repeating the research procedure of Horsman (2020) the author obtained different results. Namely, it has been found that when the contents of the Dropbox virtual drive are displayed https://www.dropbox.com/.html file is cached by Google Chrome and its name is encoded. This file is highlighted in Figure 2.

It has been determined that https://www.dropbox.com/.html does not contain any data identifying a Dropbox user. The examinations confirmed that the activity of the service user has its reflection in the website browsing history. If the Dropbox user views the entire disk content, the URL: https://www.dropbox.com/home is saved in the history of the web browser (according to Horsman's findings). Similarly, the web browser history stores web addresses that indicate other user activities, such as: viewing shared resources, opening a list of file requests, and opening a list of deleted files.

The author's examinations confirmed that displaying the content of the graphic file in the Dropbox service results in saving it in the cache of the web browser in JFIF format. In the example under discussion, the graphic file was saved under the name: fv_content = true & size_mode = 5.jfif.

The original name of the file was incorporated in the URL address and it can be read in the web browsing history. In the discussed example it is the file named: *1.JPG.*

The conducted study did not confirm the findings of Horsman (2020), according to which viewing the content of an image file and adding a comment in the Dropbox service causes saving additional information in text files.

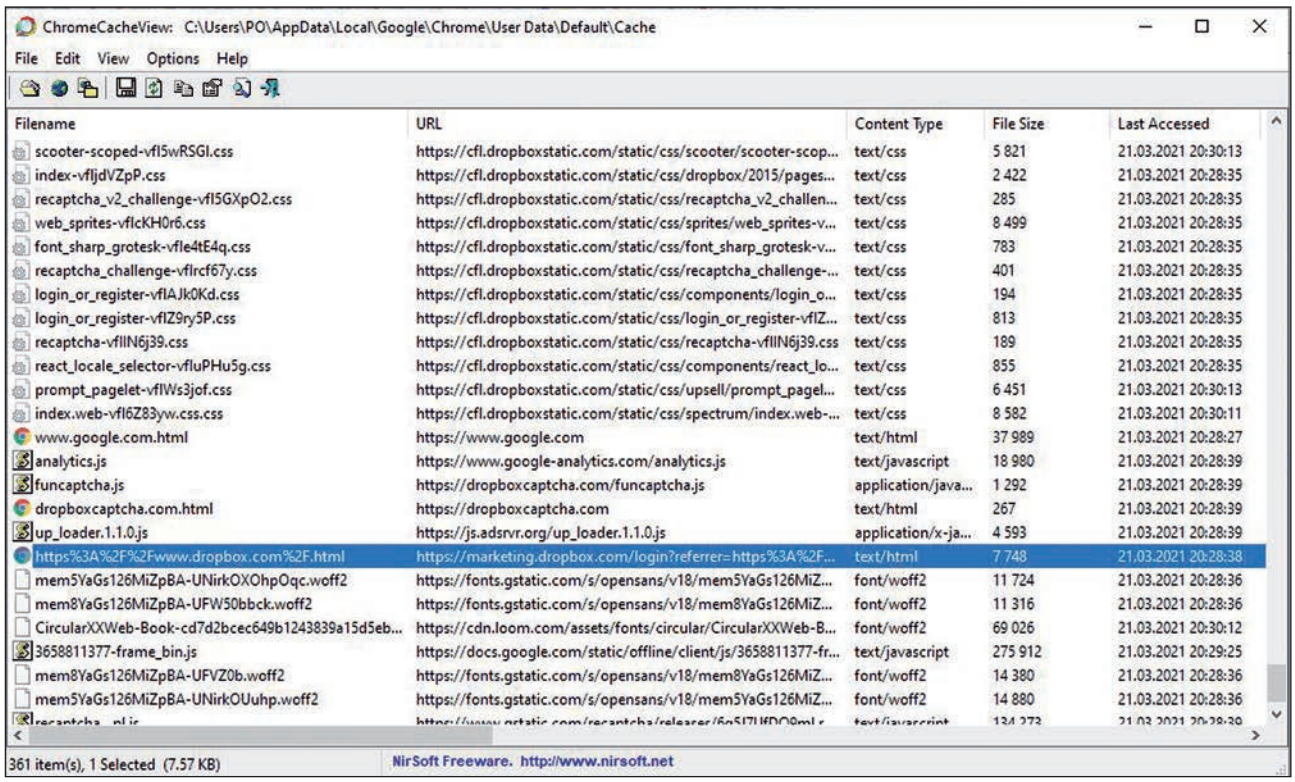

Fig. 2. ChromeCacheView window with the contents of the browser cache.

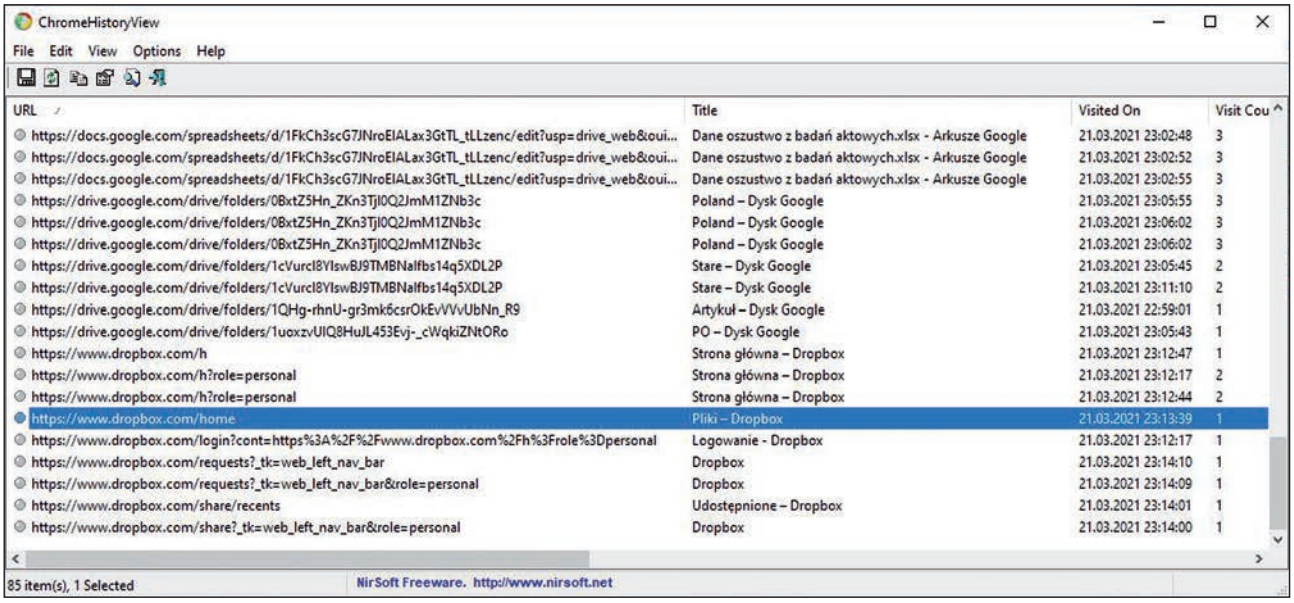

Fig. 3. Web browsing history.

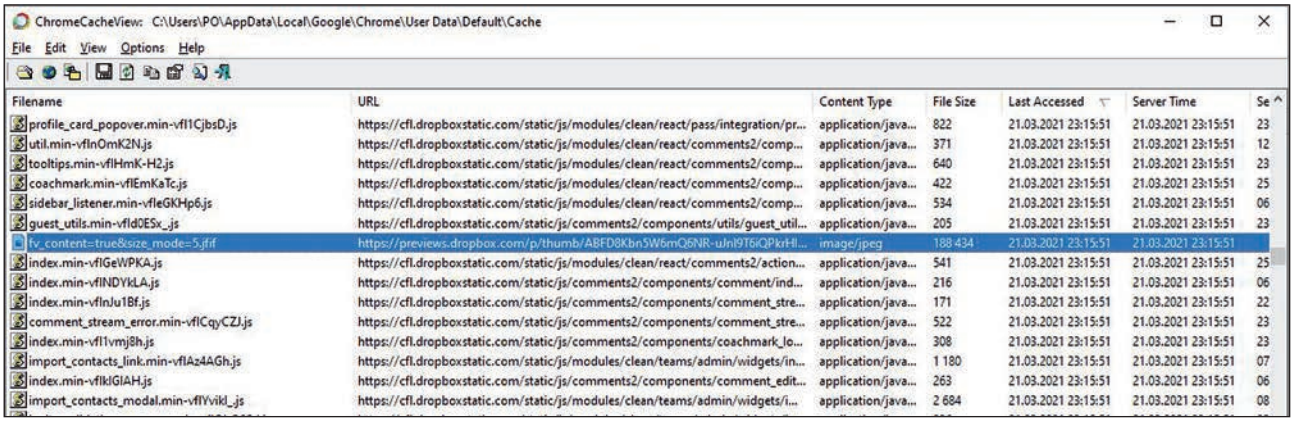

Fig. 4. Indication of fv_content=true&size_mode=5.jfif file.

Google Drive service

In the cache of the Google Chrome web browser limited information is available regarding the user's interaction with the Google Drive service. Research by Horsman (2020) shows that Google Chrome web browser cache does not contain files revealing the identity of the user or describing the contents of the virtual Google Drive. This means the information about the Google Drive service is not cached, unlike the Dropbox service. Only graphic files viewed from the Google Drive level are saved in the browser memory. The browsed files are saved under a specific name, for example: W1366-H662. This property does not apply to text documents, spreadsheets and multimedia presentations viewed at the disk level.

According to Horsman (2020) web addresses stored in the history of browsing websites and related to the user's activity contain limited information. For example, previewing the contents of an image file on a virtual drive does not change the URL: https://drive.google.com/drive/my-drive. On the other hand, displaying the contents of a specific folder alters the web address in a way that does not allow the resource name to be determined: https://drive.google.com/drive/folders/0By-CihkhmywOek1Gak4ySlhnQkk.

Results of author's own research

As a result of his own research the author found that Horsman's (2020) conclusions regarding the Google Drive service need to be complemented in terms of the possibility of identifying the names of folders stored on a virtual drive. Each time the contents of a directory are displayed, the URL changes, which does not allow the resource name to be determined. However, displaying web browsing history through ChromeHistoryView v1.42 allows to determine the names of Google Drive directories that have been opened by the user. Directories' names are included in the website titles, as shown in Figure 6.

The conducted research confirmed that the preview of the content of the graphic file on the Google Drive

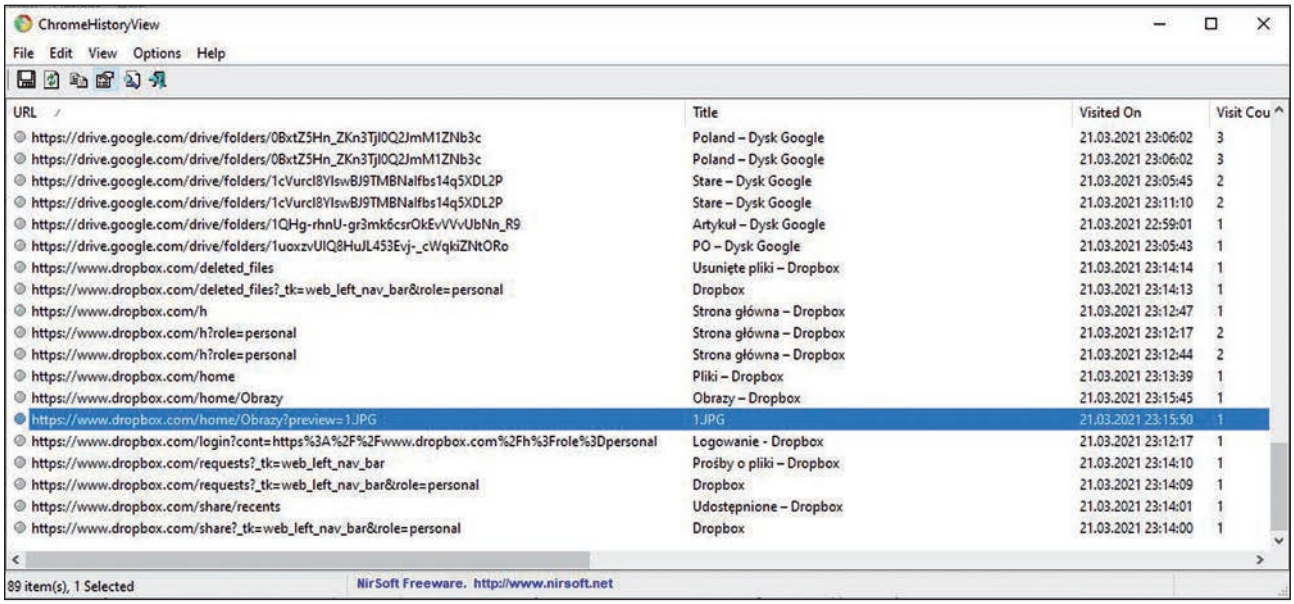

Fig. 5. Indication of 1.JPG file.

causes that it is cached and saved in the cache of the Google Chrome web browser. Figure 7 shows the graphic file w1280-h881-ft.jfif, the content of which has been displayed on the disk.

Displaying the content of text documents, spreadsheets and multimedia presentations saved on a virtual Google Drive does not result in saving additional information in the browser cache. However, it is possible to define the names of files opened by the user. Such information is available in the ChromeHistoryView program window, as shown in Figure 8.

Results and discussion

The analysis of published studies and the repeated procedure show that user interactions with Dropbox and Google Drive services result in saving information in the memories of local devices. Most information is saved when the Dropbox service is used. In this case, however, author's own research led to obtaining results different from the results of Horsman (2020). These differences are most likely due to the use of different versions of Google Chrome browser and free programs: ChromeHistoryView v1.42 and ChromeCacheView v2.25. The possibility of determining the content of the Google Drive based on the titles of the websites displayed via the above-mentioned applications, which was not described in the Horsman study (2020) should also be mentioned.

Summary

It can be assumed that the popularity of cloud computing storage services will continue to grow, which is due to the many advantages of these solutions. Therefore, there is a growing probability that a lot of data significant for criminal proceedings will be located in cloud storage including the virtual drives. However, in most cases the existing legal barriers make it impossible to gain direct access to this type of resources. In these circumstances it is necessary to use the assistance of foreign law enforcement and judicial authorities. The justification for applications addressed to foreign entities may be the findings made by experts in digital forensics. Those analyses may confirm the use of cloud-based data storage services, as well as help in determining credentials and the content of resources, which may be of importance for criminal proceedings.

Although the presented publication has mainly a character of an informative review it does contain

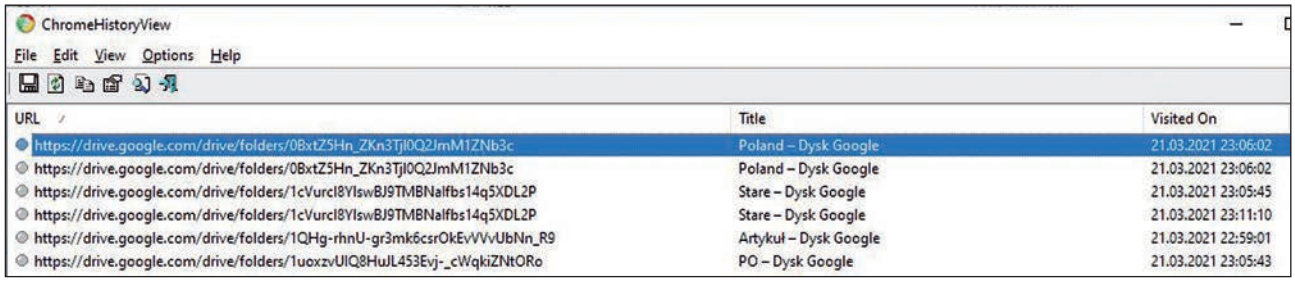

Fig. 6. Virtual disk folder names in web page titles.

Fig. 7. W1280-h881-ft.jfif graphic file saved in the browser cache.

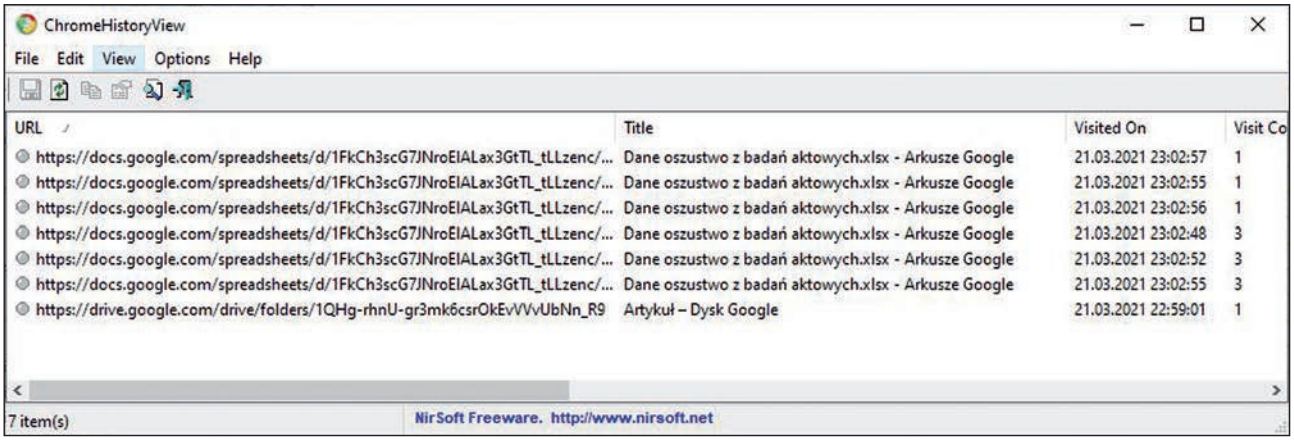

Fig. 8. Names of files opened by a Google Drive user.

a description and the results of author's own research confirming the adopted hypothesis. The own research was carried out using the latest versions of the application, which explains the differences in the obtained results. This example also confirms that each IT examination is of unique kind.

In future scientific research on the issue of detecting traces of user's activity in cloud computing services, one ought to pursue applying more advanced programms designed for forensic IT examinations, such as: X-Ways Forensics, NetAnalysis and HstEx.

Sources of figures:
Fig. 1: elaborated by author basing on: Horsman, 2020
Figs. 2–8: author

Bibliography

1. Ahmad, N.H., Hamid, A.S.S.A., Shahidan, N.S.S., Ariffin, K.A.Z. (2020). Cloud forensic analysis on pCloud: From volatile memory perspectives. In: M.H. Miraz, P. Excell, A. Ware, S. Soomro, M. Ali (ed.), *Emerging Technologies in Computing, Third EAI International Conference, iCETiC 2020*. Cham: Springer.
2. Dargahi, T., Dehghantanha, A., Conti, M. (2017). Investigating storage as a service cloud platform: pCloud as a case study. In: A. Dehghantanha, K.-K.R. Choo (ed.), *Contemporary Digital Forensic Investigations of Cloud and Mobile Applications*. Amsterdam–Boston et al.: Elsevier, https://doi.org/10.1016/B978-0-12-805303-4.00012-5.
3. Dehghantanha, A., Dargahi, T. (2017). Residual cloud forensics. In: A. Dehghantanha, K.-K.R. Choo (red.), *Contemporary Digital Forensic Investigations of Cloud and Mobile Applications*. Amsterdam–Boston et al..: Elsevier, https://doi.org/10.1016/B978-0-12-805303-4.00014-9.
4. Herman, M., Iorga, M., Salim, A.M., Jackson, R.H., Hurst, M.R., Leo, R., … Sardinas, R. (2014). *NIST Cloud Computing Forensic Science Challenges*. Gaithersburg: National Institute of Standards and Technology.
5. Horsman, G. (2020). What's in the cloud? – An examination of the impact of cloud storage usage on the browser cache. *Journal of Digital Forensics, Security and Law*, *15*(1).
6. Martini, B., Choo, K.-K.R. (2014). Cloud forensic technical challenges and solutions: A snapshot. *IEEE Cloud Computing*, *1*(4), https://doi.org/10.1109/MCC.2014.69.
7. Mell, P.M., Grance, T. (2011). *The NIST Definition of Cloud Computing*. Gaithersburg: National Institute of Standards and Technology.

8. Mohtasebi, S.H., Dehghantanha, A., Choo, K.-K.R. (2017). Cloud storage forensics. In: A. Dehghantanha, K.-K.R. Choo (ed.), *Contemporary Digital Forensic Investigations of Cloud and Mobile Applications* (s. 205–246). Amsterdam–Boston et al.: Elsevier, https://doi.org/10.1016/B978-0-12-805303-4.00013-7.
9. Samy, G.N., Shanmugam, B., Maarop, N., Magalingam, P., Perumal, S., Albakri, S.H. (2018). Digital forensic challenges in the cloud computing environment. In: F. Saeed, N. Gazem, S. Patnaik, A.S. Saed Balaid, F. Mohammed (red.), *Recent Trends in Information and Communication Technology*. Cham: Springer, https://doi.org/10.1007/978-3-319-59427-9_69.
10. Sharma, P., Arora, D., Sakthivel, T. (2018). Mobile cloud forensic: Legal implications and counter measures. W: S.C. Satapathy, A. Joshi (red.), *International Conference on Information and Communication Technology for Intelligent Systems*. Cham: Springer, https://doi.org/10.1007/978-3-319-63673-3_64.
11. Teing, Y.-Y., Dehghantanha, A., Choo, K.-K.R. (2018). CloudMe forensics: A case of big data forensic investigation. *Concurrency and Computation: Practice and Experience*, *30*(5), https://doi.org/10.1002/cpe.4277.

Translation Ewa Nogacka