A Forensic Methodology for the Identification of Illicit Data Leakage

Alessandro Simonetta[1], Luciano Fazio[2] and Maria Cristina Paoletti[1]

[1]Department of Enterprise Engineering, University of Rome "Tor Vergata", Via del Politecnico n.1, 00133, Rome, Italy
[2]Studio Giorgio ®, via Gallarate n.112, 20155, Milan, Italy

Abstract

The digital revolution had and is having profound impacts on modern society and, with it, we are witnessing the birth of new digital illicits, increasingly widespread both in Italy and in the USA. The most common case is the exfiltration of company data by unfaithful employees or former employees, who, for economic interests, act imprudently thinking that such activities are difficult to identify. This article deals with a methodology that allows you to find, in compliance with existing laws, such behaviors with the use of dedicated software tools. Furthermore, this innovation, which makes use of sophisticated data analysis techniques, must provide results that are immediately understandable and accessible even to a non-technical expert in the field such as a lawyer or a judge. For this reason, particular emphasis is given to the presentation of the found evidences through the formulation of a technical-legal report.

Keywords

computer forensics, digital proof, forensic tools, forensic analysis, civil illicit, civil proceeding, presentation of evidence

1. Introduction

When we use any electronic device, such as a computer or smartphone, our activities remain permanently recorded in the device's memory. These data are commonly called "digital traces" [1] and they can be created depending on the type of activity the user has carried out, such as:

- execution of a system program;
- read or copy a file;
- send or receive files via the Internet;
- print a file;
- access to a network resource;
- access to a remote or cloud system;
- execution of a query on a database.

However, in order for a digital trace to be used in any proceeding and, therefore, to take on probative value (thus becoming a *digital evidence*) it must be:

- *authentic*, it is necessary to have absolute certainty of the authenticity of the source from which it comes;
- *intact*, it is necessary to have a series of procedural precautions during its collection, in order not to alter its form or content in any way;
- *truthful*, obtained through the correct interpretation of the computer data;
- *complete*, through the certainty of having analyzed all the aspects connected to it, avoiding to leave out relevant information that could modify its status;
- *forensic*, obtained by respecting the laws in force [2].

In the field of private law, digital evidence is comparable to an IT document that is defined in the Italian Digital Administration Code, Legislative Decree 07/03/2005 n.82. In it we find the prerequisites that a digital document should have in order to be suitable for evidential evaluation. Unfortunately, because of the continuous technological evolution of the subject, the standards often fail to guarantee a perfect synchrony between the crystallization in legal terms and the change of technical standards [3].

Therefore, the lack of some key concepts such as the criteria for the identification, collection, acquisition, storage and transport of digital evidence has led to the use of the international standard ISO[1]/IEC[2] 27037:2012 (Fig. 1). For the management of digital evidence, the standard identifies four fundamental characteristics:

- *verifiability*: it must be possible for any involved party to evaluate the activities carried out in each phase of the life of a digital evidence;
- *repeatability*: it must be possible for any involved party to be able to reach the same results and, therefore, digital evidence, starting from the same

[1]https://www.iso.org/home.html
[2]https://www.iec.ch/homepage

SYSTEM 2021 @ Scholar's Yearly Symposium of Technology, Engineering and Mathematics. July 27–29, 2021, Catania, IT
alessandro.simonetta@gmail.com (A. Simonetta);
lf@albertogiorgio.com (L. Fazio);
mariacristina.paoletti@gmail.com (M. C. Paoletti)
https://www.albertogiorgio.com/ (L. Fazio)
0000-0003-2002-9815 (A. Simonetta)

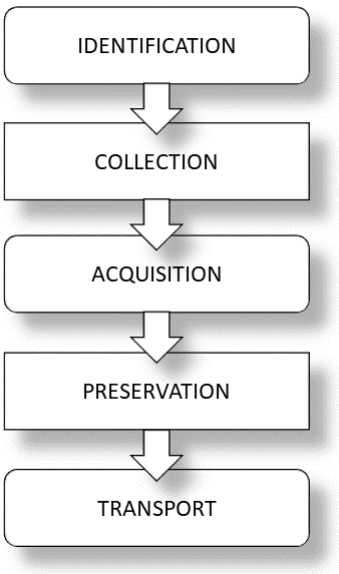

Figure 1: Scheme for the treatment of a digital evidence

conditions and following the same actions performed during the data analysis;

- *reproducibility*: it must be possible for any involved party to be able to reach the same results and, therefore, digital evidence, using different tools than the original ones, in order to be able to demonstrate that under certain conditions the original result is achieved regardless of the instrument used;
- *justifiability*: it must be possible for the operator who analyzed the data that led to digital evidence to justify every action and all the methods used to arrive at the result.

The use of the standard makes it possible to guarantee the integrity of the digital evidence from the acquisition phase and the subsequent analysis phase, and, at the same time, to obtain the admissibility characteristic of the evidence in a proceeding.

We remind you that starting from 2016 the GDPR (General Data Protection Regulation) [4] was launched, which came into force in Italy from May 2018. The introduction of a strict regulation on personal data, however, had no impact on the issue of the processing of digital evidence, since art. 9 (c.2 letter f) of the Regulation provides that the processing of personal data is lawful if it is necessary to ascertain, exercise or defend a right in court or whenever the judicial authorities exercise their judicial functions [5].

2. Data collection in the United States and in Italy

The preliminary stage to a civil proceeding in Italian law is the crystallization of the evidence in order to make it legally usable. This crystallization operation can take place through an acquisition by means of bailiffs or with the inclusion in deeds directly by the parties. In the latter circumstance, the parties involved are not obliged to present all the evidence (if these, for example, are not in their favor) but they must ensure, in any case, that the product complies with the regulations in force in terms of admissibility of the evidence.

In the United States procedural law, on the other hand, there is a preliminary phase to a proceeding called *discovery* (known in England as *disclosure*). During this phase, the parties can both obtain evidence relating to their own questions (*evidence gathering*), and investigate the opposing field to seek new information with the hope of obtaining further evidence admissible for the hearing (*evidence seeking*) [6].

In case of non-production of documents, and even worse, of incorrect or inadequate conservation of electronic documents, the consequences are serious and can compromise the subsequent procedural phase.

In the Italian legal system there is a similar mechanism (art. 210 cpc[3]) [7] but less effective, which is based on a diametrically opposite principle: the investigating judge, under certain limits (art. 118 cpc) and at the request of a party, can order the other party or a third party to show in court a document or other thing which it deems necessary for the trial. Both legislations, however, agree on the methods for the material collection of digital data, the so-called acquisition and preservation from voluntary and involuntary alterations.

In order for an acquisition to produce a digital data that can be used in any type of judicial procedure, in addition to being performed according to the standards already described, it must be accompanied by a document that describes all the handovers that the support object of acquisition undergoes between its identification, its possible seizure (where foreseen) and the crystallization of the data within it. This document is known as "*Chain of Custody*". It is the answer tested by practice to satisfy a rule of the discipline of the acquisition of evidence: the party interested in the acquisition of an object must present sufficient elements to make it appear that it corresponds to what is claimed to be [8].

At this point it is necessary to identify the suitable tool to physically carry out the data acquisition from a variety of possible candidate tools [9]. Once the tool has been decided, we move on to the data extraction phase from the digital source and to the creation of the so-called

[3]Italian Code of Civil Procedure

forensic image, in one of the possible formats available and in relation to the goal we want to achieve [10].

Before starting the analysis, it is necessary to confirm that what was collected and crystallized corresponds exactly to the original format. This is possible through the generation of the hash code of the two objects, which obviously must provide the same result. The use of this coding technique makes it possible to verify the exact correspondence of the two objects in any process phase.

3. Case study

Between 2018 and 2020, into the United States incidents caused (or involving) by internal staff increased by 47% [11]. The frequency of accidents varies according to the type of company. The Verizon 2021 Breach Investigation Report [12] provides an overview of the different types of incidents in the various types of companies involved. Companies in the *Health and Finance* sector recorded the largest number of incidents caused by the incorrect use of their employees' access privileges and suffered the largest number of data thefts. The exfiltration of data by the unfaithful employee in the United States, according to a 2020 statistic that involved 300 accidents in 8 different types of industrial sectors[13], was perpetrated for as many as 43% of the cases through forwarding to personal email accounts, while, for 16% of cases through the incorrect use of cloud sharing privileges. The remaining number of data exfiltration cases involve using USB devices (9%) and more. See Table 1.

Table 1
Typical unfaithful employee behavior in US

Behavior	%
E-mail forwarding to personal e-mail account	43.75
Misusing cloud collaboration privileges	16.07
Data aggregation - downloads	10.71
Using unauthorized/unencrypted USB devices	8.93
Data snooping using sharepoint	8.04
Data exfiltration using external sites	6.25
E-mails sent to non-business domains	3.57
E-mails sent to competitor domains	2.68

It is interesting to note that the main reasons that induce employees to make such a gesture [11, 14, 15] are economic (64%), followed by espionage (17%), entertainment (17%) and issues of resentment (14%). So, if the economic leverage is so strong, it will be even higher for a former employee who will feel free from constraints in leaving the old company.

For this reason we will analyze the case study of the former employee who, after moving from one company to another, uses documents owned by the old company in the new one (e.g. customer list, company secrets, confidential information, source code or banks data).

To the ex employee could be challenged various offenses, for example, for having violated the contractual rules that bind him to the old company, or the rules in force in the field of copyright protection, of company jurisprudence or unfair competition (art. 2598 cc[4]).

According to the data provided courtesy of Studio Giorgio®[5] on over 100 cases handled in Italy, the exfiltration techniques used by the former employee are the same compared to those used in the US (Table 2): sending emails to personal mailboxes is the tool used for 50% of cases, while external USB devices are used for over 30% (much higher than 9% of the US statistic).

Table 2
Typical unfaithful employee behavior in IT

Behavior	%
E-mail forwarding to personal e-mail account	51.75
Using unauthorized/unencrypted USB devices	30.80
Data exfiltration using external sites	9.85
Others	7.60

3.1. Forensic analysis software platforms

To prove wrongdoing by a former employee the company has the right, by virtue of the clauses normally required for the use of company tools (PC, telephone, e-mail box, storage disks, ...), to access the information contained therein. On the market there are various [16] software platforms that allow you to support the digital forensic expert in all activities, starting from the creation of forensic images [17]:

- AccessData FTK (Forensic ToolKit)[6]
- X-ways Forensic [7]
- EnCase [8]
- Magnet AXIOM[9]

All data analysis platforms have peculiarities that distinguish them from each other and, therefore, pros and cons, but all strive to provide a comprehensive solution for the analysis of the most common hardware/software environments.

The aforementioned tools allow you to process a huge amount of data but, before allowing full use of their functions, they need to have the computer used for their operation carry out a preliminary data processing phase.

[4] Italian Civil Code
[5] https://www.albertogiorgio.com/
[6] https://www.exterro.com/forensic-toolkit
[7] http://www.x-ways.net/forensics/
[8] https://security.opentext.com/encase-forensic
[9] https://www.magnetforensics.com/products/magnet-axiom/

During the pre-processing phase, the entire content of the data extracted from the original finds is read (in the form of a forensic image) and, by means of machine learning techniques [18][19][20], now increasingly used in various scientific contexts [21][22][23][24][25][26], the data and images are classified and indexed [27][28], thus creating the most common artifacts from the source system. This phase is typically onerous from a computational point of view and requires a fair amount of time, which often clashes with the need for speed of an analysis. For this reason, new computing architectures are being studied, such as quantum computing [29] or computing solutions based on multi-valued algebra (MVL) [30][31].

Once this phase has been completed, the analysis software allow access to the statistically most relevant behaviors of the former employee, such as:

1. USB devices connected in the last working period of the former employee [32];
2. files accessed and possibly copied to an external device or remotely, in the last working period;
3. cloud storage services used without authorization from the company;
4. emails containing company information sent to personal email addresses;
5. printing of company documentation.

Obtained this minimum set of information, it is possible to have sufficient elements to have the legitimate suspicion (if not proof) of the export of confidential and protected company data. However, this approach is not comprehensive because it considers the activities performed by employees on corporate devices and tools.

As the internal network can be a valid vehicle for disseminating data that can also be used with non-company workstations, the analysis can also be extended to this potentially available class of activity.

For example, the monitoring and logging tools of network activities (if present) allow to detect, even afterwards, the data read/copied by a specific user within the company storage, such as QRadar Risk manager, CA Spectrum and Netwrix Auditor [33].

It is important to underline that, the initial phase of crystallization of the entire amount of data, to which the former employee had access during the employment relationship (forensic image), is fundamental both as a term of comparison for the research of the exported data, and to demonstrate the origin of the data for which protection is requested.

All these analyzes are based on the employee using a digital data transfer. There are also analog modes (which leave no trace) and are more difficult to detect, such as a screen photograph.

However, it should also be considered that enterprise-level companies should adopt solutions to protect data dynamically also based on the type of request. In fact, there are database access monitoring software that can detect "suspicious" activities that cannot be performed.

Finally, the expert will draw up a technical report aimed at showing the evidence found.

3.2. Presentation of the evidence

The presentation of the results of a forensic analysis is crucial to understand the behavior of the former employee.

The presentation takes place by means of the drafting of a technical-legal report, that brings together what was found with any specific violations identified.

Fig. 2 shows the structure of an expert report in its fundamental sections. The aim is to highlight the evidence found, using a language suitable for understanding even for a non-technical reader such as a lawyer or a judge.

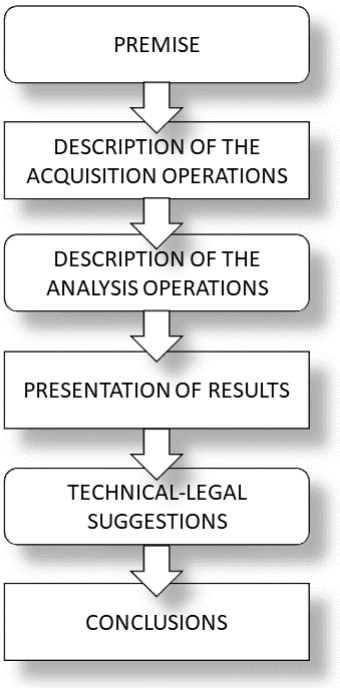

Figure 2: Scheme for the presentation of the evidence

The *premise* shows who conferred the assignment, the objective of the assignment and any other useful element to motivate the choices in the methodology adopted. Sometimes it is useful, already at this stage, to provide the reader with an anticipation of the evidence subsequently found.

The following sections are all intended for an expert in the field, therefore they technically describe the development of the various operations.

The *description of the acquisition operations* section contains all the elements necessary to verify the integrity and authenticity of what has been acquired, at any time after the presentation of the technical report. It is essential to enter in detail the picture of the acquisitions performed, describing each intervention up to the creation of the forensic image.

The *description of the analysis operations* section describes all the technical methods implemented for the analysis of the forensic images of the acquired finds. In it, it is important to indicate the tools used to process the data, but also the logical process used to examine them.

We then arrive at the (technical-legal) section of *presentation of results*. The goal is to describe, in a clear, linear and precise way, every evidence found and every element useful to describe the events, creating a sort of [34] timeline of them. It is also useful to insert fragments of data, screenshot or reports extracted from forensic software, in order to match the resulting evidence with the underlying objective data.

The *technical-legal suggestions* section contains any proposals or indications so that whoever is entitled to a legal proceeding can act in the most appropriate and technically correct manner. Furthermore, it is appropriate to remember the presence of the touchstone created (in the form of a forensic image) in order to identify the material presumably extracted in the subsequent stages of the procedure to be established.

Finally, in the *conclusions*, it is appropriate to enclose a broad summary of the previous sections, confirming the evidence found (anticipated in the introduction), providing a cross-section of the events and actions carried out by the former employee.

4. Conclusions

The Italian judicial system, in the civil field, does not have a specific reference standard for the management of digital evidence, for this reason the methodologies applied by professionals in the sector refer to international standards, such as ISO/IEC 27037:2012.

Furthermore, the methods of introducing digital data into civil proceedings in Italy differ considerably from what happens overseas, however the method of acquiring and analyzing the evidence remains valid in both doctrines.

The statistics collected in the USA have shown an ever greater growth in data exfiltration from companies, identifying the unfaithful employee as the cause of greater frequency. While, in Italy, we observed the same trend for the former employee who fraudulently commits the same offense using the same techniques.

This article describes the methodology to be adopted to protect the company in the event of data exfiltration by a former employee and what is the correct procedure that the forensic technician must follow to identify any offenses. The method of presenting the results allows a non-technical reader to have all the necessary tools available to be able to act in the best possible way during all the subsequent phases of the procedure, that will be established against the former employee.

References

[1] R. Brighi, Informatica forense, algoritmi e garanzie processuali, Ars interpretandi, Rivista di ermeneutica giuridica (2021). doi:10.7382/100798.

[2] V. G. Calabro, La fragilità delle tracce digitali, Master breve in Diritto e Tecnologie Informatiche (2009). doi:10.13140/RG.2.2.27355.62240.

[3] C. Galli, A. Giorgio, Others, L'acquisizione forense delle prove in materia di violazione dei segreti aziendali, in: Il nuovo diritto del know how e dei segreti commerciali, Wolters Kluver, 2018.

[4] European Union, Regulation 2016/679 (General Data Protection Regulation), 2016. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679.

[5] G. Barrera, Il trattamento ai fini di ricerca dei dati personali relativi a condanne penali e reati. a proposito di gdpr, Rivista di studi e ricerche sulla criminalità organizzata (2019). doi:10.13130/cross-11272.

[6] M. Gradi, L'obbligo di verità delle parti, ISBN 9788892114036, G. Giappichelli Editore, 2018.

[7] L. Dittrich, L'esibizione delle prove, in: Diritto Processuale Civile, Utet Giuridica, 2019.

[8] L. Bartoli, La catena di custodia del materiale informatico: soluzioni a confronto, Universidad de La Laguna. Servicio de Publicaciones, España (2016). URL: http://riull.ull.es/xmlui/handle/915/6247.

[9] M. Faiz, W. Prabowo, Comparison of acquisition software for digital forensics purposes, 2018. doi:10.22219/KINETIK.V4I1.687.

[10] E. Akbal, S. Dogan, Forensics image acquisition process of digital evidence, International Journal of Computer Network & Information Security (2018).

[11] Insider threat statistics you should know, 2021. URL: https://www.tessian.com/blog/insider-threat-statistics/.

[12] Verizon 2021 breach investigations report, 2021. URL: https://www.verizon.com/business/en-sg/resources/reports/dbir/.

[13] Most common data exfiltration behaviors during insider threats in the united states in 2020, https://www.statista.com/statistics/1155846/most-common-data-exfiltration-insider-threat-types-usa/, 2020.

[14] R. Avanzato, F. Beritelli, M. Russo, S. Russo, M. Vaccaro, Yolov3-based mask and face recognition algorithm for individual protection applications, in: CEUR Workshop Proceedings, 2020, pp. 41–45.

[15] G. Capizzi, C. Napoli, S. Russo, M. Woźniak, Lessening stress and anxiety-related behaviors by means of ai-driven drones for aromatherapy, volume 2594, 2020, pp. 7–12.

[16] Popular computer forensics, 2021. URL: https://resources.infosecinstitute.com/topic/computer-forensics-tools/.

[17] K. Ghazinour, D. M. Vakharia, K. C. Kannaji, R. Satyakumar, A study on digital forensic tools, IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI) (2017) 3136–3142. doi:10.1109/ICPCSI.2017.8392304.

[18] R. M. A. Mohammad, M. Alqahtani, A comparison of machine learning techniques for file system forensics analysis, Journal of Information Security and Applications 46 (2019) 53–61. URL: https://www.sciencedirect.com/science/article/pii/S2214212618307579. doi:10.1016/j.jisa.2019.02.009.

[19] C. Napoli, G. Pappalardo, E. Tramontana, A mathematical model for file fragment diffusion and a neural predictor to manage priority queues over bittorrent, International Journal of Applied Mathematics and Computer Science 26 (2016) 147–160.

[20] S. Spanò, G. C. Cardarilli, L. Di Nunzio, R. Fazzolari, D. Giardino, M. Matta, A. Nannarelli, M. Re, An efficient hardware implementation of reinforcement learning: The q-learning algorithm, IEEE Access 7 (2019) 186340–186351. doi:10.1109/ACCESS.2019.2961174.

[21] A. A. Jaber, R. Bicker, Fault diagnosis of industrial robot gears based on discrete wavelet transform and artificial neural network, Insight 58 (2016) 179–186. doi:10.1784/INSI.2016.58.4.179.

[22] M. Wozniak, D. Polap, G. Borowik, C. Napoli, A first attempt to cloud-based user verification in distributed system, in: 2015 Asia-Pacific Conference on Computer Aided System Engineering, IEEE, 2015, pp. 226–231.

[23] A. A. Jaber, A. Saleh, H. F. M. Ali, Prediction of hourly cooling energy consumption of educational buildings using artificial neural network, International Journal on Advanced Science, Engineering and Information Technology 9 (2019) 159–166. doi:10.18517/IJASEIT.9.1.7351.

[24] A. A. Jaber, K. M. Ali, Artificial neural network based fault diagnosis of a pulley-belt rotating system, International Journal on Advanced Science, Engineering and Information Technology 9 (2019) 544–551. doi:10.18517/IJASEIT.9.2.7426.

[25] F. Bonanno, G. Capizzi, L. G. Sciuto, A neuro wavelet-based approach for short-term load forecasting in integrated generation systems, in: 2013 International Conference on Clean Electrical Power (ICCEP), 2013, pp. 772–776. doi:10.1109/ICCEP.2013.6586946.

[26] F. Bonanno, G. Capizzi, G. Lo Sciuto, C. Napoli, Wavelet recurrent neural network with semi-parametric input data preprocessing for micro-wind power forecasting in integrated generation systems, 2015, pp. 602–609. doi:10.1109/ICCEP.2015.7177554.

[27] G. C. Cardarilli, L. D. Nunzio, R. Fazzolari, D. Giardino, A. Nannarelli, M. Re, S. Spanò, A pseudo-softmax function for hardware-based high speed image classification, Scientific Reports 11 (2021). doi:10.1038/s41598-021-94691-7.

[28] G. Capizzi, G. Lo Sciuto, C. Napoli, E. Tramontana, M. Woźniak, A novel neural networks-based texture image processing algorithm for orange defects classification, Int. J. Comput. Sci. Appl. 13 (2016) 45–60.

[29] S. K. Sharma, M. Khaliq, The role of quantum computing in software forensics and digital evidence: Issues and challenges, in: Limitations and Future Applications of Quantum Cryptography, IGI Global, 2021, pp. 169–185. doi:10.4018/978-1-7998-6677-0.ch009.

[30] A. Simonetta, M. C. Paoletti, M. Muratore, A new approach for designing of computer architectures using multi-value logic, International Journal on Advanced Science, Engineering and Information Technology (in press).

[31] A. Simonetta, M. C. Paoletti, Designing digital circuits in multi-valued logic, International Journal on Advanced Science, Engineering and Information Technology 8 (2018) 1166–1172. URL: http://ijaseit.insightsociety.org/index.php?option=com_content&view=article&id=9&Itemid=1&article_id=5966. doi:10.18517/ijaseit.8.4.5966.

[32] A. Neyaz, N. Shashidhar, Usb artifact analysis using windows event viewer, registry and file system logs, Electronics (2019). doi:10.3390/electronics8111322.

[33] A. Khurat, P. Sangkhachantharanan, An automatic networking device auditing tool based on cis benchmark, 18th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON) (2021). doi:10.1109/ECTI-CON51831.2021.9454830.

[34] V. Calabrò, P. D. Checco, B. Fiammella, La timeline: aspetti tecnici e rilevanza processuale, IISFA Memberbook (2011).