Forensic Science International: Digital Investigation

Digital forensic logistics: the basics of scientific theory

--Manuscript Draft--

Manuscript Number:	
Article Type:	Research Paper
Section/Category:	Future Challenges
Keywords:	digital forensics; logistics; algorithmization; modeling; big data; scientific theory; investigation
Corresponding Author:	Dmitry V. Bakhteev, Ph.D. Ural State Law University YEkaterinburg, Sverdlovskaya RUSSIAN FEDERATION
First Author:	Sergey V. Zuev, Doctor of Law
Order of Authors:	Sergey V. Zuev, Doctor of Law
	Dmitry V. Bakhteev, Ph.D.
Abstract:	Investigations of complex crimes with digital evidence increasingly require the use of modern digital devices and computer programs. Working with big data involves the accumulation, processing, and analysis of forensic information for further algorithmization and modeling of investigative actions, as well as the automation of the organizational activities of investigators. The article substantiates the need for the use of digital forensic logistics to optimize information flows and build the most effective analytical human and computer processing, not excluding the use of artificial intelligence systems. Digital forensic logistics is a sub-branch of digital forensics in the collection, identification, storage, verification, and analysis of data, as well as the generation of electronic evidence for evidence in court. The article provides the main directions of digital forensic logistics, including the logistics of evidence in criminal cases; logistics of the general organization of crime investigation; logistics planning (selection of tools and methods of investigation); logistics of putting forward versions of events; logistics of decisions in criminal matters.

Powered by Editorial Manager® and ProduXion Manager® from Aries Systems Corporation

Modern forensics needs a system of logistic (data processing) methods

Digital forensic logistics determines means of obtaining and collecting data

Digital forensic logistics supports the whole investigation process

Main areas of development of this theory are blockchain, AI and electronic assistants

Manuscript File Click here to view linked References

Digital forensic logistics: the basics of scientific theory

Sergey Zuev [1]
Dmitry Bakhteev [2]

Abstract

Investigations of complex crimes with digital evidence increasingly require the use of modern digital devices and computer programs. Working with big data involves the accumulation, processing, and analysis of forensic information for further algorithmization and modeling of investigative actions, as well as the automation of the organizational activities of investigators. The article substantiates the need for the use of digital forensic logistics to optimize information flows and build the most effective analytical human and computer processing, not excluding the use of artificial intelligence systems. Digital forensic logistics is a sub-branch of digital forensics in the collection, identification, storage, verification, and analysis of data, as well as the generation of electronic evidence for evidence in court. The article provides the main directions of digital forensic logistics, including the logistics of evidence in criminal cases; logistics of the general organization of crime investigation; logistics planning (selection of tools and methods of investigation); logistics of putting forward versions of events; logistics of decisions in criminal matters.

Keywords: digital forensics, logistics, algorithmization, modeling, big data, scientific theory, investigation

Introduction

The effectiveness of solving and investigating crimes committed using digital technologies depends on the solution of the logical, algorithmic, automatic, mathematical, and legal problems of working with big data (accumulation,

[1] E-mail address: zuevsv@susu.ru (S. Zuev). URL: https://www.famous-scientists.ru/14214/
[2] E-mail address: dmitry.bakhteev@gmail.com (D. Bakhteev).

processing, analysis of information) coming from various sources, for further optimal use of the results obtained in criminal cases.

Walder H. and Hansjakob T. (2016) believe that in some cases, crime investigation often consists of 10% forensics and 90% painstaking work with data arrays. For example, if each of the suspects in a case stores several thousand WhatsApp messages on their phone, then certain tools are needed to analyze the message data which may contain forensic information. If there are five suspects for 20 criminal acts, it will be extremely difficult to analyze the available information if such activities are not properly structured and organized.

Solving such issues requires scientifically-based recommendations on working with forensically-significant information flows subject to analytical human and computer processing, including through the use of modern technical means and special software, not excluding the application of artificial intelligence systems.

Theory

Software makes it possible to collect information, upload it to electronic devices, download it from electronic devices, change the data expression format, and analyze processed data without the need for human labor. The subject of forensic knowledge (the investigator) receives results according to predefined parameters and algorithms for the movement of information.

Like many other terms in the sciences of the criminal law cycle, the term "logistics" originally refers to the field of military affairs and denotes the totality of knowledge and their derivative actions in the procurement, maintenance, and transportation of military equipment, equipment and people[3]. Information logistics can be understood in two ways: as “an area of organization logistics that studies and solves the problems of organizing and integrating information flows for

[3] Logistics. Mirriam-Webster: dictionary. For more, see https://www.merriam-webster.com/dictionary/logistics.

making managerial decisions in logistics systems" (Alesinskaya, 2010) (as a kind of add-on to "analog" production processes), or as a field of scientific knowledge reflecting the applied aspects of working with digital information, including in large volumes. Digital logistics can be considered part of the latter group, the development of which is largely the result of digitalization and the increase in the computing power of modern computer technology. Digital logistics is closely related to such areas of computer science as data science (machine science, deep learning, computer vision, etc.) The methods and technological solutions of digital logistics can be used in all areas of daily life, including in the specific tasks of solving and investigating crimes. Forensics, by setting objectives and criteria for approaches to information processing to solve legal problems, is simply a segment of this area, which allows us to speak about a subspecies of digital logistics—digital forensic logistics.

In a narrow sense, digital forensic logistics is a system of methods for collecting, processing, storing, and investigating electronic information relevant to solving and investigating crimes during criminal investigations based on logical rules and algorithms of a single digital environment.

Place in digital forensics

Digital forensic logistics is seen as a sub-branch of digital forensics, which in turn represents the process of collecting, identifying, storing, verifying, and analyzing data or information to be presented in court as evidence (Khan and Varma, 2020). The form in which such methods are implemented is determined by criminal procedure legislation, and the content is based on the forensic concepts of recognition, classification, the mechanism of traceability, and the theory of forensic identification.

In digital forensics, all traces of a crime (both material—that which is left on objects and documents, and testimonial—reflected in the minds of individuals) are information that forms the corresponding flows coming into a single digital

environment (usually in the form of a network or network cluster in the form of a message, publication, logs, etc.). Logistic operations performed by the investigator or software allow for the selection of optimal means of identifying the person who committed the crime, as well as achieving other goals of the criminal process. Automated identification of suspects is typical, for example, of a digital system called SISC, which is based on the categorization of attributes of ordinary criminals stored in a database, a decision tree, logistic regression, and chi-square analysis methods (Taha and Yoo, 2018). Another example of such software is the UK-developed VALCRI (Visual Analytics for Sensemaking in Criminal Intelligence) project, which allows for the optimization of search, finding and evidence-based activities (Islam, 2017) in a graphical (relatively easy to understand) form.

Related work

The modeling method plays a large role in digital forensics. There are various models of logistic regression, including binary, proportional, ordered, partially ordered, and disordered regression procedures for categorical answers, which are also used in social science (Hilbe et al., 2009).

Modeling includes a description of typical and specific criminalistic characteristics of crimes, a set of tools and methods of proof, forecasting decisions in criminal cases, identifying investigative errors, etc.

Digital forensic logistics is closely related to the process of organizing criminal investigations, building forensic versions of events and deriving logical consequences from them, planning investigative actions, and intelligence-gathering operations. For example, during a search of a suspect's home, an empty hard drive is found. When recovering deleted information, several hundred thousand text files (documents, network and system logs, etc.) are detected, the manual processing of which is either impossible or impractical from the point of view of saving time. In this case, digital forensic logistics methods may be applied.

As a rule, all such methods are based on dichotomous division as the operation of dividing the class of objects of forensic knowledge into subclasses, subclasses into groups, etc. The typical work with trace information is constructed in a similar way: it can be represented as a sequence of closed questions, for instance: “Is this person involved in the event being investigated?”, “Does this trace relate to the crime event”, “Could a change in digital information have occurred at the indicated time”, and the responses to them. Answers to such questions can be obtained by investigative means, or by operational and expert means.

In the process of establishing a criminal event and proving the guilt of a suspect, the investigator receives information and, managing the information flow, employs the most acceptable (optimal) set of procedural (investigative actions) and non-procedural means (intelligence-gathering operations) of establishing the truth in a criminal case. Logistic operations permeate the entire digital communications system, relying on existing and newly-received data. The criminal justice evidence information system itself is the result of such operations.

Digital forensic logistics, by modeling typical forms of criminal investigations, determines not only the possible means of obtaining information (information flows), but also determines the course of the investigation, and provides access to wide electronic departmental control through certain information resources. Its application should be based on a systematic approach, including the use of forensic records, the electronic interaction of the investigator with other units and services of law enforcement agencies, as well as with state institutions and officials, and a system for working with citizens.

Subject of digital forensic logistics

The subject of digital forensic logistics is any forensically-significant information recorded in electronic form, including data obtained as a result of

investigative and intelligence-gathering activities for the detection and investigation of crimes, both structured and unstructured.

Digital forensic logistics includes several areas: logistics of evidence in criminal cases; logistics of the general organization of crime investigation; logistics planning (selection of tools and methods of investigation); logistics of putting forward versions of events; logistics of decisions in criminal matters.

Per the logistics of evidence, when collecting and examining electronic evidence, the investigator builds a system-forming aggregate of information about the crime event and the guilt of the person who committed it. This involves the collection of information through a series of logistic techniques and operations and the formation of a set of evidence-based arguments (levers) that allow the investigator to change the direction (state) of the entire system, offer optimal solutions, and a set of tools and methods to achieve truth in a criminal case. The accumulation of digital information minimizes uncertainty, that is, eliminates information entropy.

The actions of the investigator in a criminal investigations should be based on the principle of logical algorithms, that is, they should be natural, carried out according to a predetermined standard, and should be economical in terms of the use of resources. This might involve the verification of operational electronic information about a crime which has been committed, is being committed, or is being planned; conducting joint investigative and operational measures (electronic surveillance, detention of a suspect, etc.). The effectiveness of the entire system will largely depend on the establishment of information flows and determining priorities in solving problems. The applied digital technologies, which can provide the necessary algorithmization of the proof process, must be improved to ensure quality work of investigators.

Over the past few years, the total number and variety of digital footprints found at crime scenes and during other investigative operations has grown significantly. In addition, there is an increased need for accurate results in a set period of time. The main challenges that coincide with these aforementioned tasks

are to investigate the correct set of evidence and set aside an appropriate time to investigate it (Gupta et al., 2019).

For example, the algorithmization of evidence for bribery can be carried out by filling certain template clusters (models) with digital information and taking into account the most common patterns of evidence obtained. By distributing their significance for this category of cases, further work on the criminal case is carried out. There may be several such models.

This process, sometimes referred to as the chain of custody (CoC), should ensure that evidence is not changed during the investigation, despite the fact that the evidence passed through several organizations, in order to be admissible in the courts. Currently, digital evidence is controlled entirely by CoC, and entities involved in the chain are required to fill out documents accompanying the evidence. The chain of custody can be based on the blockchain, which guarantees the verifiability of the evidence collected and provides the possibility of establishing the owners of evidence. In this case, digital evidence (or electronic evidence) refers to any evidence that is stored in memory or transmitted in digital form, which a party can use in court. Of particular importance is a reliable cryptographic hash function (Bonomi et al., 2020).

Access to evidence (its study and verification) should be available to all parties, depending on the transaction used, taking into account the interests of the investigation (secrecy of the pre-trial investigations). Authorized users must be able to submit (create) a new piece of digital evidence.

Blockchain technologies

Blockchain technologies provide the functionality of the entire digital platform. All transactions are cryptographically signed by the sender (creator), which can easily become known to all participants of the blockchain network. This creates a certain protection of access to the data available in the system[4].

[4] Solidity. For more, see https://solidity.readthedocs.io.

The forensic significance of advancing versions of events is that during the initial stage of investigations, investigators use electronic information that allows them to obtain, with a certain degree of probability, assumptions about the circumstances to be proved in the criminal case. The logistics of advancing forensic versions of events involves adding information included in logical consequences, dynamic planning, and determining the priority of verification of investigative versions of events. Circumstances may change, and this fact is registered in the system through the corresponding information flow, changing the course and direction of the investigation.

The logistics of advancing forensic versions of events at the stage of the investigation of crimes consists of the following logistic operations (algorithms):

1) the entry of primary information into the system in electronic form;
2) determination of possible information flows of information accumulation;
3) distribution of information according to specified criteria (channels);
4) processing and analysis of digital information;
5) formation of assumptions about the individual who committed the crime, their profile;
6) continuous collection of information during the investigation;
7) proposals on areas of actual verification of investigative versions of events in the form of logical consequences and investigative actions to establish their compliance with the version;
8) correction of forensic versions of events taking into account newly received information in the system.

Thus, the logistics of advancing forensic versions of events is a system of the programmed accumulation, processing, analysis, and further use of information about a crime that comes to the investigator to form a reasonable assumption about the person who committed the crime in order to detect and detain them.

Hypothesis testing and elimination of redundant versions of events can be carried out by a system similar to DFP (Sunde and Dror, 2019). Any hypothesis should be based on available information. Based on the results of his study, Rassin

(2018) concluded that using a pen and paper to evaluate evidence for two competing hypotheses can lead to tunnel vision, i.e., opportunities to analyze more information are lost. Successfully this task can be handled by a special computer program.

Large amounts of data may require reliable cloud space as well as a unified algorithm for checking the integrity of streaming data, which allows authorized users to verify the integrity of forensic data and identify and localize any malicious information changes (Liu et al.).

Ideally, the logistics of decisions made in criminal matters would involve an electronic assistant endowed with artificial intelligence or special software capable of proposing solutions to the investigator. These solutions could change the course of the investigation, and the assistant (or software) would transfer the entire information system to the investigator based on the determination of the forensic significance of newly-received digital information (electronic evidence). Evaluation criteria are set in advance and may vary depending on the model of the crime, the set of factual circumstances, and the emerging investigative practice. The system is configured with consideration of scientifically based expected indicators. The system's proposals provide guidance (recommendations) to investigators and are not required. At the same time, investigators should be ready to explain why they are not acting according to the behavior proposed by the computer. That is, they should be prepared to propose an alternative algorithm of actions or an explanatory model of the evidence available.

Of interest are studies of systems in which decision trees are studied depending on such categories as the value of the selected attribute, confidence level, forecast accuracy, etc. The emphasis is on the use of artificial intelligence (Quinlan, 1986). From the point of view of modern forensics, the elements of the forensic characteristics of crimes can act as branches of these trees.

Currently, digital reporting support and decision support systems (e.g. DERDS) are being actively developed to help professionals assess the reliability of

inferences and assumptions about conclusions regarding any potentially evidence-based results (Horsman, 2019).

However, it should be noted that there is a certain unfoundedness of many studies that focus on the fact that artificial intelligence systems can provide transparency in decision making and support. With the right mathematical modeling, the most successful technological solutions in this area—artificial neural networks and methods based on deep learning—can produce solutions that are effective both in terms of the accuracy of a single solution and the stability of the results, but they are often characterized as "black boxes". This is due to the presence of a hidden layer in the structure of such systems, within which the system is being trained. This factor prevents the transparency of such decisions and therefore the decisions of such systems cannot have probative value.

The most useful implementation of the presented forms is achieved by using a single digital logistics platform, which includes the algorithmization of actions and decisions in a criminal case. The work of this digital platform is based on the circulation of information flows according to predetermined models of crimes committed, taking into account the information available on a specifically identified illegal act.

Such information flows will be:

1. Digital forensic records;
2. Digital tracks;
3. Electronic inquiries and instructions;
4. Digital video library, audio recordings, and photographs;
5. Electronic assistant of the investigator (interrogator);
6. Digital examinations;
7. Electronic forensic recommendations for the investigation of criminal cases, preferably in mobile and offline form.

Digital forensic records are a legally regulated information system that is necessary to concentrate and present information relevant to the investigation of criminal cases and crime prevention to the preliminary investigation bodies and the

court. Accounting is achieved through files containing information about the accounting entity (traces of handprints, etc.).

Digital footprints are the result of the actions of an individual or an automated system, embodied, as a rule, in text or multimedia form, and suitable for transformation into evidence in criminal cases. For example, it has been noted that when investigating the receipt of a bribe, investigators often record electronic traces indicating the preparatory actions of the offender (preliminary agreement on meetings of the briber with the bribe taker, consent to participate as an intermediary in the transfer or receipt of the bribe). Criminal communication between the briber and the bribe taker was not through in-person contact, but occurred through the use of computers or mobile devices via SMS, instant messengers, or emails, and the fact of transfer-receipt of a bribe was also recorded on electronic information media.

Electronic inquiries and instructions are standard templates of the investigator's requests to various organizations and record orders sent to operational officers to obtain the necessary information during the investigation of criminal cases as part of the ongoing electronic document management system.

Digital video, audio recordings, and photographs can accumulate in cloud storage. There are various proposals for the application of software, modeling, and process algorithms to such data (Khan and Varma, 2020).

In addition, the relationship between crimes can be viewed by using markers (electronic tags) which indicate the most significant circumstances. This will greatly simplify the establishment of group crimes.

Electronic assistant

An electronic assistant for investigators can be used to keep records of information relevant to the investigator (legal information consisting of laws and other regulatory legal acts; materials of preliminary reviews; templates of procedural documents) or methodological recommendations.

The reconstruction of criminal acts plays a key role in investigatory procedures. A digital investigator assistant, using mathematical methods, modeling, and automatic verification of information, could be of great help in this (Soltani and Seno, 2019). For example, conducting forensic examinations in relation to digital information carriers, on the basis of which investigators can complete analyses of electronic document management, research of digital images, or research of material objects using their digitalized images. This would help automate investigatory experts' desks.

Electronic forensic recommendations for investigating criminal cases contain elements of private forensic techniques for investigating certain types of crimes and recommendations for organizing and conducting investigations in criminal cases (forensic characteristics, typical investigative situations, typical investigative versions of events, algorithm for investigating criminal cases).

Conclusions

Digital forensic logistics acts as a system of forensic and computerized methods used in solving and investigating crimes, providing for an interdisciplinary synthesis of scientific achievements and the joint functioning of knowledge of the law, natural science, and technology.

The actions of investigators in criminal investigations should be based on the principle of logical algorithms, that is, they should be natural, fit a predetermined standard, and should be resource-efficient.

The effectiveness of the entire system will largely depend on the establishment of information flows and determining priorities in solving problems. The applied digital technologies must be improved to achieve a higher quality of investigation. These technologies can provide the necessary algorithmization of the proof process. The use of special software, including the application of artificial intelligence systems, is becoming increasingly relevant.

Acknowledgements

The reported study was funded by RFBR according to the research project № 18-29-16001.

References

Alesinskaya T.V. 2010. Fundamentals of Logistics. Functional areas of logistics management. Part 3. Taganrog: Publishing house of TTI SFU, 2010.116 p.

Islam J., Anslow C., Xu K. and William Wong B.L. 2017. Analytical Provenance for Criminal Intelligence Analysis. Valcri white paper series. VALCRI–WP–2017–009 / ed. by B.L. William Wong. http://valcri.org/our-content/uploads/2017/02/VALCRI-WP-2017-009-Provenance3.pdf.

Bonomi S., Casini M., and Ciccotelli C. 2019. B-CoC: A Blockchain-Based Chain of Custody for Evidences Management in Digital Forensics. International Conference on Blockchain Economics, Security and Protocols, Tokenomics 2019, May 6–7, 2019. Paris, France. Article. No. 12, pp. 12:1–12:15. https://drops.dagstuhl.de/opus/volltexte/2020/11964/pdf/OASIcs-Tokenomics-2019-0.pdf.

Gupta J.N.D., Kalaimannan E., Yoo SM. 2019. A Sequential Investigation Model for Solving Time Critical Digital Forensic Cases Involving a Single Investigator. National Cyber Summit (NCS) Research Track. NCS 2019. Advances in Intelligent Systems and Computing. Vol 1055. Springer, Cham, pp. 202-219. https://www.springer.com/gp/book/9783030312381.

Hilbe J.M. 2009. Logistic Regression Models, Chapman & Hall/CRC Press, 2009. https://books.google.ru/books/about/Logistic_Regression_Models.html?id=tmHMBQAAQBAJ&redir_esc=y. (Accessed 11 мая 2009).

Hoanrsm G. 2019. Formalising investigative decision making in digital forensics: Proposing the Digital Evidence Reporting and Decision Support

(DERDS) framework. Digital Investigation. 2019. Vol. 28, pp. 146-151. https://research.tees.ac.uk/en/publications/formalising-investigative-decision-making-in-digital-forensics-pr. (Accessed 25 January 2019) published 1 March 2019.

Khan Y., Varma S. 2020. Development and Design Strategies of Evidence Collection Framework in Cloud Environment. Social Networking and Computational Intelligence. Lecture Notes in Networks and Systems. Vol 100. Springer, pp. 27-37. https://www.springer.com/gp/book/9789811520709.

Liu, A., Liu, J., Uehara, T. 2014. Secure streaming forensic data transmission for trusted cloud SFCS 2014. Proceedings of the 2nd International Workshop on Security and Forensics in Communication Systems, pp. 3-10. https://www.scimagojr.com/journalsearch.php?q=21100320410&tip=sid&clean=0.

Rassin E. 2018. Reducing tunnel vision with a pen-and-paper tool for the weighting of criminal evidence // Investigative Psychol. Vol. 15 (2) (2018), pp. 227-233. https://onlinelibrary.wiley.com/doi/abs/10.1002/jip.1504. published 14 April 2018.

Soltani, S., Seno, S.A.H. 2019. A formal model for event reconstruction in digital forensic investigation Digital Investigation // Digital Investigation. 2019. 30, c. 148-160. https://www.sciencedirect.com/science/article/pii/S1742287619301185. (Accessed 13 August 2019).

Sunde, N., Dror, I.E. 2019. Cognitive and human factors in digital forensics: Problems, challenges, and the way forward // Digital Investigation. 2019. 29, c. 101-108. https://www.sciencedirect.com/science/article/pii/S1742287619300441.

Taha K.,Yoo P. D. 2018. A Forensic System for Identifying the Suspects of a Crime with No Solid Material Evidences. 2018 IEEE 16th Intl Conf on Dependable, Autonomic and Secure Computing; 16th Intl Conf on Pervasive Intelligence and Computing; 4th Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech), Athens, 2018. pp. 576-583. https://ieeexplore.ieee.org/document/8511950. published 29 October 2018.

Quinlan, J.R. 1986. Induction of decision trees. Mach Learn 1, pp. 81–106 (1986). https://link.springer.com/article/10.1007/BF00116251. published March 1986.

Walder H., Hansjakob T. 2016. Kriminalistisches Denken. Heidelberg, 2016. pp. 7–8. https://www.schulthess.com/buchshop/detail/ISBN-9783783200430/ Hansjakob-Thomas-Hrsg.-Walder-Hans/Kriminalistisches-Denken.