
**Law & Digital Technologies 2013-2024**

ISSN 2079-8784

URL - http://ras.jes.su



№ 1 Том 1. 2021

# Система регистрации фактов доступа к защищаемым данным в системах цифровой экономики

**Grusho Alexander**
*Федеральный исследовательский центр «Информатика и управление» Российской академии наук, Москва, Россия*
*Российская Федерация, Москва*

**Piskovsky Victor**
*Федеральный исследовательский центр «Информатика и управление» Российской академии наук, Москва, Россия*
*Российская Федерация, Москва*

**Zabezhailo Mikhail**
*Федеральный исследовательский центр «Информатика и управление» Российской академии наук, Москва, Россия*
*Российская Федерация, Москва*

**Аннотация**

В статье рассматривается подход к построению системы регистрации фактов доступа к защищаемым данным. Информация о факте доступа содержит идентификационные данные субъектов, осуществивших доступ, данные об информационных системах и время доступа. Цель - помочь в расследовании уязвимостей и слабых мест в защите. Кроме того, такая система содержит данные сотрудников, ответственных за возможную утечку информации, со всеми юридическими и судебными последствиями. Идея состоит в применении технологий распределенного реестра (Distributed ledger technology). Система позволяет идентифицировать пользователей, которые пытались получить или получили ценную, защищаемую, информацию. В настоящее время техническая и теоретическая база для таких решений готова. Анализ текущей ситуации в рассматриваемой области показывает, что все ведущие игроки в этом сегменте ИТ-рынка, параллельно с разработкой математических моделей предлагают также использовать методы проблемно-ориентированного интеллектуального анализа данных. Значительное внимание при этом уделяется разработке программных и программно-аппаратных средств обеспечения производительности рассматриваемых решений.

**Ключевые слова:** информационная безопасность, защита ценной информации от компрометации по косвенным признакам, распределенный реестр, интеллектуальный анализ данных, выскокопроизводительные вычисления

**Дата публикации:** 02.07.2021

**Источник финансирования:**

The work is partially supported by Russian Foundation for Basic Research, projects 18-29-16145-mk and 18-07-03124-mk

**Ссылка для цитирования:**

Grusho A. , Piskovsky V. , Zabezhailo M. Система регистрации фактов доступа к защищаемым данным в системах цифровой экономики // Law & Digital Technologies – 2021. – Том 1. – № 1 С. 10-18 [Электронный ресурс]. URL: https://ldt-journal.com/S123456780015725-8-1 (дата обращения: 29.04.2024). DOI: 10.18254/S123456780015725-8

## 1. Introduction

In the coming Industry 4.0 epoch, data is of a particular value. Personal digital identification, medical data, accounts at telecom operators' services, users' behavior profile on the Internet, financial and bank accounts statuses and transactions easily become a trade object (TAdviser 2020). Regardless of commercial purposes, there will always be examples of dishonest use: from the deliberately malicious use of the fake digital identification or blackmail to practically legal methods of determining the credit scoring points for loan agreements, including fraudulent phone calls, e-mail spamming, phishing attacks and other challenges of the early stage of the rapid digital economy growth.

The principal problem of these growth ills is that even after the so-called data anonymization or obfuscation, actual data owners can be relatively easily recovered indirectly. If we are talking about a person, then it can be biographical facts, the profile of his calls, contacts on social networks, mailing lists, recruitment, preferences when visiting Internet resources. Even when a company name is not present, the profile of banking operations did not give a reason for well-informed people to doubt which company they belonged to. With the current development of data aggregation and analysis technologies, it gets simpler and easier to identify the owner of a particular digital profile, and the more data is subjected to the machine analysis, the easier and more accurately the identification problem can be solved. Thus, sensitive information becomes more accessible and cheaper. In other words, even anonymized data have their own specifics, and people can identify an owner according to these data specifics with a high probability. Thus, there is a problem to protect privacy even in case of collecting valuable secondary information. If we cannot resist to collect and analyze data, then we have to make it more transparent who and when accesses it. Such kind of problem was considered in the literature (Grusho et al. 2019). So, if we register who and when retrieves data from various, even closed systems, then we can identify a potential infringer. In this way we could also gather the evidence base of accusation if needed.

## 2. Example

Obviously, for the treatment of most diseases, it is necessary to collect and use the experience gained by other doctors in various clinics. The problem of the combined use of medical databases has been discussed at many conferences and in numerous publications.

Anonymization is used to protect personal data in the exchange of medical experience. However, as it was mentioned above, overcoming such protection with a malicious intent is quite possible. The method of using functional dependencies in databases has long been known (Su and Özsoyoglu 1987). The method includes multiple, specially created requests to the database. Moreover, in these requests, valuable information of interest to the potential adversary can be only indirectly indicated. At the same time, the summation of the results of queries gives the adversary the chance to unambiguously determine the diagnosis of a particular patient and other valuable data. If you remember the sequence and content of the requests made by the opponent, you can identify her/him and their targets. Note that one can use databases of various clinics under the guise of noble goals.

## 3. Technologies and proposals for technical solutions

Next, we will formulate the requirements for a tool to track attempts to obtain valuable information. We will also consider the existing technologies that allow building such a system (Piskovski et al. 2020). Thus, we'll construct in the whole the design and thereby prove the following statement:

Modern technologies and methods make it possible to develop and implement a publicly available system that tracks who, where and when accessed data, regardless of where this data is processed and stored. The system allows you to reliably identify who is trying to obtain valuable information. At the same time, the system itself is protected from unauthorized access and data collection.

Since this system collecting queries and accesses becomes also an attractive target to attack, we also have to protect this tool and its data. Since the system registers facts of data queries and accesses from different organizations, the most important aspect is to protect the integrity and ordering of such data. Moreover, the adversary is unknown a priori, so it is necessary to remember the requests of all users. Confidentiality is required to protect the analytics of the collected data. Thus, we have to close sensitive information storing in the system, and only authorized personnel could access it. An essential requirement is the amount of information stored and the growth rate of its volume. So, we should integrate distributed databases in order to carry out analysis concurrently for many users.

Summarizing all mentioned above, we could formulate the following principles:

1) Decentralization. This concerns access, storage and accounting. There should be no logical control center.

2) Excluding any falsification.

3) Data are going to be a kind of currency. Therefore, like crypto wallet owners track all their transactions, information owners should be able to see the registered events of a successful access to their protected data, when, how and who performed those actions.

4) Only a board of authorized representatives should have a full access. The minimum number of such representatives is regulated by a public agreement. In other words, only a qualified number of representatives could retrieve all information stored.

5) The system is to store anonymized data, namely identification numbers of the participants, and it should be technically impossible to restore the owners. Because of this, such a system can be opened and combine the logs of several operators of personal data without the risk of direct compromise.

Distributed ledger technology (DLT) perfectly meets 1), 2) and 3) requirements. The requirement 4) implies applying of threshold signature scheme (TSS) (Stathakopoulou and

Cachin 2017). The requirement 5) implies the Diffie-Hellman approach. As practice has shown, the use of mentioned technologies requires a thorough theoretical consideration for not only meeting the requirements, but also for proving the algorithms robustness and the fundamental lack of technical capabilities to compromise the system (Baird et al. 2019).

17 Considering technologies and their roles in the system we start from hash function. A hash function is the primary security tool for DLT. A block in a block chain or a vertex in a directed oriented graph contains, among others, payload, a timestamp, its own hash checksum, and the hash sum of the previous block or vertices. The most frequent exploited property of hash functions is an asymmetry in producing direct check sum and solving an inverse problem. The field nonce contains a random number, which is used in calculating checksum via hash functions. Given this value a direct checksum calculated is obtained easily and almost immediately while an inverse task to find this field is a kind of NP-hard problem. This property guarantees the integrity and content unchangeability for DLT blocks and vertices. In fact, hash function usage had been created to limit spam in email systems. The constraint of zeros in the first 20 bits of checksum is imposed when the hash value itself is as many as 160 bits long. Its analogue in BitCoin has more stringent restrictions, namely the hash function value must be less than a certain constant. The constant depends on the current computation speed to control the rate at which new blocks appear in the distributed ledger.

18 Another technology concerns identification. The Diffie-Hellman approach is used to identify accounts in cryptocurrency systems. Signature forgery also belongs to the category of NP-hard discrete logarithm problems over a finite field. Since keeping the secret part of the key is not recommended on untrusted devices, the signature of the secret part of the account holder is. only allowed on a device fully controlled by the owner, including interfaces with other devices.

19 DLT is a decentralized system (Churyumov 2016). BitTorrent protocol is developed for the distributed storage and exchange of contents (files) in a peer-to-peer network. Various protocols, similar to BitTorrent, are well fitted and used to maintain referential integrity in DLT systems.

20 The task to provide transparency and correctness of data initiates a usage of DLT transaction technologies. This kind of technologies essentially depends on DLT models and their consensus protocols.

21 In implementing the considered tracking systems, we have to provide an extremely high performance. Also, the solution is to be well scalable and expandable. For example, the performance of SWIFT network is 50 thousand TPS (transactions per second), Visa - 45-65 thousand TPS (Hauge 2018; VISA 2021). The system under consideration should be many times faster. In order to decide which of DLT model can be applied to build the system we consider 4 DLT types: Blockchain, Directed acyclic graphs (DAG), Hashgraph, and Holochain.

22 *3.1. Blockchain*

23 Blockchain is a linked chain of blocks. In general, this method implements the hash Merkle tree (Wikipedia. Merkle Tree.; Becker 2008). There are more than 20 DLT platforms and about 30 consensus protocols referred to this model with performance from dozens to millions TPS.

24 *3.2. HashGraph*

25 HashGraph (the platform Hedera) is a directed acyclic graph; each node stores its own history of events (analogs of blocks in a block chain) and exchanges information using a specific proprietary gossip protocol. HashGraph is a running consensus algorithm related to

Asynchronous Byzantine Fault Tolerance (ABFT). All this provides high scalability while maintaining reliability with performance about 300 TPS (Baird et al. 2019).

### *3.3. Holochain*

The Holochain model (Harris-Braun et al. 2018), according to its developers, represents a separate branch of the DLT. The Holochain platform consists of a network of agents maintaining a unique chain of sources of their transactions, combined with a common space implemented as a validation, monotone, sharded, distributed hash table (DHT), where each node applies validation rules for this data in the DHT. According to the Holochain developers, the platform maintains system integrity without introducing consensus. Holochain, like Hedera (HashGraph), uses the gossip protocol for nodes to share information about the shared experience of other nodes.

### *3.4. DAG*

DAG is a directed acyclic graph, a highly scalable alternative to blockchains for building DLT. It's worth to note that HashGraph can relate to DAG as well.

DAG is also a standalone DLT method (Directed Acyclic Graphs) with extremely high performance (millions TPS) and scalability. Any node in the peer-to-peer network can both check and create messages containing information registered in the registry. Unlike HashGraph, all nodes in a DAG perform a double function: not only validating, but also representing a verified transaction.

There are two types of DAG platforms: blockDAG and txDAG. In blockDAG, each vertex contains a set of transactions, which is similar to the concept of a block in a block chain. What makes blockDAG different from blockchain is that each block can be hash-targeted to multiple parent blocks.

Unlike blockDAG, in txDAG, each vertex of the graph represents a unique transaction, and the diverging branches of the vertex must contain disjoint transactions. This is convenient for resolving conflicts; therefore, such a network is less demanding on the computations on the nodes, which, in turn, can improve the performance of the registry, which in this case is limited only by the network bandwidth.

Also, the performance of DLT systems depends on applied consensus algorithms (Churyumov 2016; Chen et al. 2018). At present there are about 40 different versions of consensus algorithms related to 4-5 quite different categories. Consensus is required to create a unified access repository database that is maintained by all members.

## **4. Intelligent Data Analysis**

The observed deviations from the formalized norm of behavior of the curves reflecting the frequency characteristics of access to the protected personal data by the subjects of IT systems, including bots and applications operating under the accounts of responsible users, are analyzed for the classification of such deviations along these lines:

− insignificant deviations from the norm; that is, not the result of deliberate (targeted) access attempts;

− representing the result of targeted access attempts or significant deviations from the norm.

Firstly, the tasks of the System in such situations are as follows:

− identification of data access facts;

40 − causal analysis of the structure of the access facts registered in the System, including the sequence, method, subject and object of access, in order to classify the attack;

41 − the choice and optimization of decision-making strategies for attacks classified in this way (together with the implementation of appropriate measures to counter their consequences).

42 From the perspective of the mathematical models features and algorithms necessary for the successful identification and causal analysis of the recorded data, it seems that special attention should be paid to:

43 − learning from use cases, namely descriptions of previously recorded and studied facts of both successfully-achieved goals, and unsuccessful attacks);

44 − operating in a reliable manner, including both very large and small, statistically insignificant samples of accumulated recorded data;

45 − conducting an exhaustive analysis of the reasons for the effectiveness to classify the mentioned access attacks to the protected data in order to form effective means of classification and to develop proactive measures, before the subject of the attack requests an investigation, and counteraction measures;

46 − operating both numerical (for example the frequency of occurrence of certain events, and so on) and non-numeric (that is, the most significant factors of influence and the relationship between them) characteristics describing the analyzed precedents.

47 A certain problem here may be a choice of an adequate algorithmic language to describe accumulated precedents. On the one hand, it should be a tool with sufficiently complete descriptive capabilities, because with its help it is required to introduce all the necessary characteristic features of the studied precedents into the organized causal analysis. However, on the other hand, we would like such descriptive tools to have a minimal complexity, which can provide an effective search for signs that define the appearance of a precedent.

48 Together, the above-mentioned features of the subject area under discussion, namely the tasks of classification and the proactive classifying potentially dangerous precursors, allow us to speak of it as one of the areas to apply data mining methods or Intelligent Data Analysis as a computer analysis assisted by intelligent computer systems allowing to build reasoning and decision-making schemes (Grusho et al. 2015).

## 49 5. How to build Intelligent Data Analysis

50 The problem of classification, expertise, evaluation and properly the definition of the norm, which requires its solution in the framework of monitoring security events (see, for example, ENISA. n.d.), can be reduced to identification and differentiation of two types of situations:

51 − randomness, or fluctuation;

52 − there are signs of targeted malicious activity.

53 The first step in this process is to select the data description method used to capture the accumulated information about the observed access events to the protected data. At the same time, the question of adequate expressive and descriptive capabilities of such a method seems to be of fundamental importance: the use of computer tools for analyzing the data obtained during the recording of events implies the possibility to represent by its means (perhaps in an implicit form) all the essential factors, the interaction of which led to the occurrence of subject

requests to investigate the attack. Indeed, only in this case can we count on the acceptability of the results of such a computer analysis. However, in each particular case, it seems natural to choose an adequate method, since the requirements for the speed of data analysis and their real volumes in solving applied problems can be extremely sensitive to the volume of calculations that arise here.

54 When organizing the discussed classification, examination and evaluation of the results of monitoring the facts recorded by the System from the point of view of the presence of precedents, it seems natural to require that the actual process of calculations, computer analysis of data and the procedures for evaluating its results assessing the sufficiency of grounds for accepting the results of the performed examination were separated and performed as independent procedures of the relevant regulations.

55 The place for an Intelligent Data Analysis in such a process is determined by at least three circumstances. Firstly, an appropriate computer-based data analysis can be used as an intelligent capability booster, which performs the same operations as an expert. This allows the expert to understand and unambiguously interpret the results generated by it, but doing it significantly faster and in significantly larger volumes than the expert using its results.

56 Secondly, the data accumulated during the monitoring process can be used for the examination and evaluation of newly recorded events, if the events described in a uniform way that have already been recorded are used as an appropriate training sample. At the same time, the facts of effective achieved goals and unsuccessful attacks can be used as a set of examples and counterexamples for organizing machine learning based on precedents. Thus, the general scheme of the Intelligent Data Analysis can be described in the following three steps:

57 1. Analysis of accumulated data: recorded facts of access to protected data and recorded requests for investigation of precedents of violations of legal standards and confidentiality agreements.

58 2. Formation of empirical dependencies, decision rules.

59 3. Examination of new incidents conducted with the help of empirical dependencies generated from the training sample.

60 When organizing the process of generating such dependencies, you will have to separate two independent factors according to different principles requiring the use of appropriate analysis tools – mathematical models, methods and algorithms types of monitoring environments. The first is characterized by large samples of frequently repeated events, which allow us to identify stable patterns of behavior. For the second, on the contrary, it is essential to operate with small, statistically insignificant samples of precedents, where the role of each individual event can be critically important.

61 Later in the discussion, we will use the following, in our opinion, convenient concepts and designations:

62 − Events, incidents that can be qualified as random fluctuations or, conversely, as the result of a targeted impact.

63 − Use cases, which are events classified as successful targeted impacts that resulted in a confirmed leak of protected information.

64 − Events, that are described as similar to targeted malicious actions, but are classified as unsuccessful; that is, for which no precedents, that is requests to investigate, have been registered. We will distinguish the following three cases, namely the cases with the results of the examination being undertaken:

(i) yes, a precedent classified as successful;

(ii) no, a precedent classified as unsuccessful;

(iii) we do not know if there is no or there is not enough data for a reasoned classification of a precedent to be classified as successful or unsuccessful.

The main purpose of our further discussion will be to review the most significant properties or functionality of mathematical models, methods and algorithms:

− Primary events analysis, namely formation of a training sample.

− The formation of rules, namely procedures, methods, decision-making rules, and so on, for the classification of use cases.

− The formation of security policy requirements due to the classification performed on the basis of the generated rules.

At the same time, in the analysis of primary events, it is important to pay special attention to:

− available options for fixing all relevant details;

− the formation of a system for entering recorded data and metadata into the structures, for detailing the description of use cases in the interests of improving the Intelligent Data Analysis;

− the analysis of the similarities and differences in the available case descriptions.

In implementing the classification of secondary events based on the rules, attention should be paid to the correctness of the procedures to extrapolate the empirical dependencies expressed by the rules to the descriptions of secondary events in order to ensure a reasoned sufficiency of grounds for accepting the results of the implemented extrapolation.

## 6. Formation of rules for the classification of use cases, the review of procedural techniques

Among the formal procedures used in this field for the formation and classification of security events, the following two classes of mathematical models and methods deserve the most complete attention:

(a) Search for stable regularities in the accumulated data.

(b) Check of the feasibility of the accepted restrictions of standards and rules for working with protected data.

In turn, the first of these two classes is seemingly appropriate to divide into two subclass treatments, which use significantly different models, methods and algorithms:

(a1) large, that is containing sustained repetitive statistically significant regularity in the behavior of objects;

(a2) small, that is not statistically significant collections of precedents. In the case of (a1), the most widely used mathematical techniques of Intelligent Data Analysis are regression analysis and case clustering, as well as various variants of Bayesian inductive inference technology. Apparently, the most complete version of the implementation of these mathematical data analysis tools can be found in the SAS Enterprise Miner and SAS/STAT software packages offered by the SAS Institute (SAS Institute Inc. 2013; SAS Institute Inc. 2015). Neural network technologies, for example Kohonen self-organizing maps (SAS Institute Inc. 2013), are quite

popular. Tools for analyzing web links, for link analysis in Web log data-see, for example, (SAS Institute Inc. 2013), which are focused on extracting regularities persistently repeated or on the interrelated information about who visited the corresponding web servers, when, and with what requests in stored log files or corresponding databases, are widely used. As a rule, when analyzing security events, such tools help you determine what actually happens and whose actions cause those events.

84 Special attention should be paid to models and tools for analyzing associative relationships. Here, along with the well–studied problems of restoring various types of correlation dependencies from empirical data, pair correlations, correlation trees, and the like (see, for example, SAS Institute Inc. 2013; Pearl 2000), solutions based on the so-called Association Rules, Agrawal (1993; 1994) are increasingly being developed.

85 In the case of (a2), the situation is not so good. Many models and technologies that are successfully applied in the case of (a1), when they use the appropriate algorithmic procedures for analyzing data on small samples of use cases, cease to be mathematically correct. Thus, the question of trust in the results generated, that is whether there are sufficient grounds for their acceptance, becomes a critical factor in deciding whether certain approaches are applicable in specific applications.

86 As early as at the turn of the 1960s and 70s, Plotkin (1970; 1971) showed that, as in the situation with the resolution method, the problem of generating minimal inductive generalizations for sets of first-order predicate logic formulas has exponential characteristics of computational complexity. Attempts to circumvent such difficulties, for example, by using certain discrete procedures for inductive generalization of descriptions of accumulated precedents quite quickly run into the problems of exponentially rapidly growing performance requirements. Finally, the latter results usually into solely academic significance, that is far from real applications.

87 Situations of type (b) are typical, in particular, for an extensive class of problems of checking the correctness of the System construction, and the register of registering the facts of access to protected data.

88 Using mathematical technique of the so-called correct algebras over a set of incorrect or heuristic algorithms was proposed by Zhuravlev (1977, 5–17; 1977, 21–27; 1978) and successfully developed by his school (Zhuravlev et al. 2006; Rudakov 1986). This technique can also provide additional opportunities for organizing Intelligent Data Analysis in the case of (a)-type problems. It can also be useful to use formal models of empirical induction on structural, that is non-numeric descriptions of precedents and methods for optimizing the combinatorial search that arises here, developed by the school of Finn (Finn 2009; 2011; 2012).

89 Concluding the brief review of the approaches, methods and models of Intelligent Data Analysis undertaken in this Section, it seems appropriate to draw attention to a number of significant limitations specific to the technologies considered. Hopefully, these circumstances can be an incentive for the development of new, better solutions. So, the most significant bottlenecks are the following:

90 − independence of factors or variables for performing correct statistical analysis within the framework of the Intelligent Data Analysis;

91 − reliability of working with small, namely statistically insignificant, samples;

92 − the completeness of the causal analysis performed within the framework of the Intelligent Data Analysis, namely the completeness of the set of identified factors and of the recorded precedents;

93 − the use of only explanatory variables in regression interpolation and the problem of completeness of the factors that characterize the use cases;

94 − the necessity to use the Intelligent Data Analysis, including processing of structural, non-numeric objects, which are also endowed with numerical characteristics;

95 − the need to assess the sufficiency in order to accept the results, including extrapolation to new objects of empirical dependencies, which are obtained in the process of interpolation of the training sample data;

96 − the need to overcome the curse of computational complexity, which is caused by the presence of difficult-to-solve iterative problems in the subject area under consideration. Those problems, e.g., the so-called problem of controlling and optimizing combinatorial iteration, have a critically significant impact on the performance requirements of the corresponding application software systems.

97 **7. Intelligent Data Analysis and HPC technologies**

98 Analysis of the current situation in the area under consideration shows that all the leading players in this segment of the IT market, in parallel with the development of mathematical models and methods of problem-oriented data mining, pay significant attention to the development of special software and hardware tools to support the performance of such tool solutions.

99 As such, the SAS Institute company, along with the development of the functionality of its products SAS Enterprise Miner, SAS / STAT, SAS/OR, SAS Constraint Programming Solver, and so on, develops related HPC technologies (IBM Cloud Services 2021), whose task is to optimize the computing execution environment and ensure the maximum possible performance of the Company's basic solutions.

100 **8. Conclusion**

101 The paper proposes a new approach to protect valuables. The idea relates to computer auditing, which was outlined in (Grusho et al. 2015), based on the responsibility for violation of information security rules. In the public information space, it is advisable to build a database for tracking user access to a distributed database based on DLT. The paper proves the possibility to build such a system via existing technologies.

102 **Acknowledgments**

103 The work is partially supported by Russian Foundation for Basic Research, projects 18-29-16145-mk and 18-07-03124-mk

---

**Библиография:**

1. Agrawal, Rakesh, Tomasz Imieliński, Arun Swami. 1993. Mining association rules between sets of items in large databases. Proceedings of the 1993 ACM SIGMOD international conference on Management of data, June 1993, N.-Y., 207 – 216. https://doi.org/10.1145/170035.170072

2. Agrawal, Rakesh, and Ramakrishnan Srikant. 1994. Fast Algorithms for Mining Association Rules. Proceedings of the 20th VLDB Conference Santiago, Chile, 1994, 487-499.

3. Baird, Leemon, Mance Harmon, and Paul Madsen. 2019. Hedera: A Public Hashgraph. Network & Governing Council. The trust layer of the internet. https://www.hedera.com/hh-

whitepaper-v1.4-181017.pdf.

4. Becker, Georg. 2008. “Merkle Signature Schemes, Merkle Trees and Their Cryptanalysis.” https://www.emsec.ruhr-uni-bochum.de/media/crypto/attachments/files/2011/04/becker_1.pdf.

5. Chen Tai-Yuan, Wei-Ning Huang, Po-Chun Kuo, Hao Chung, and Tzu-Wei Chao. 2018. A Highly Scalable, Decentralized DAG-Based Consensus Algorithm. https://eprint.iacr.org/2018/1112.pdf.

6. Churyumov, Anton. 2016. “Byteball: A Decentralized System for Storage and Transfer of Value.” https://obyte.org/Byteball.pdf.

7. Cohen, Paul. 2015. Big Mechanism (DARPA Big Mechanism Program). Physical Biology 12(4). https://doi.org/10.1088/1478-3975/12/4/045008

8. Department of Defense Trusted Computer System Evaluation Criteria, DoD. 1985. Accessed April 10, 2021. https://csrc.nist.gov/csrc/media/publications/conference-paper/1998/10/08/proceedings-of-the-21st-nissc-1998/documents/early-cs-papers/dod85.pdf

9. Directed Acyclic Graphs (DAGs). In Version Control by Example. https://ericsink.com/vcbe/html/directed_acyclic_graphs.html.

10. ENISA. n.d. “ISO/IEC Standard 15408 - Information technology -- Security techniques -- Evaluation criteria for IT security.” Accessed April 10, 2021. https://www.enisa.europa.eu/topics/threat-risk-management/risk-management/current-risk/laws-regulation/rm-ra-standards/iso-iec-standard-15408

11. Finn, Victor et al. 2009. Automatic Generation of Hypotheses in Intelligent Systems, edited by Victor Finn. 2009. Moscow: Librokom.

12. Finn, Victor. 2011. J.S. Mill’s inductive methods in artificial intelligence systems. Scientific and Technical Information Processing 38(6): 385–402.

13. Finn, Victor. 2012. J.S. Mill’s inductive methods in artificial intelligence systems. Scientific and Technical Information Processing 39(5): 241–260.

14. Grusho, Alexander, Mikhail Zabezhailo, Alexander Zatsarinnyi, Victor Piskovskii, Sergey Borokhov. 2015. On the potential applications of data mining for information security provision of cloud-based environments. Automatic Documentation and Mathematical Linguistics 49(6): 193-201. https://doi.org/10.3103/S0005105515060023

15. Grusho, Alexander, Nikolay Grusho, Mikhail Zabezhailo and Elena Timonina. 2019. “Protection of valuable information in public information space” Communications of the ECMS. Proceedings of the 33th European Conference on Modelling and Simulation 33(1): 451–455.

16. Harris-Braun, Eric, Nicolas Luck, and Arthur Brock. 2018. Holochain. Scalable agent-centric distributed computing,” DRAFT (ALPHA 1) - 2/15/2018. https://whitepaperdatabase.com/holo-chain-hot-whitepaper/

17. Hauge, Bjorn. 2018. SWIFTNet, VisaNet and Blockchain: The Future of Clearing. https://medium.com/datadriveninvestor/swiftnet-visanet-and-blockchain-the-future-of-clearing-f42de3ced34c.

18. IBM Cloud Services. n.d. Accessed April 10, 2021. http://www-935.ibm.com/services/us/en/it-services/cloud-services/

19. Lally, Adam, Sugato Bachi, Michael Barborak, David Buchanan, Jennifer Chu-Carroll, David Ferrucci, Michael Glass, Aditya Kalyanpur, Erik Mueller, William Murdock et al. 2014. WatsonPaths: Scenario-based Question Answering and Inference over Unstructured Information (IBM Research Report RC25489). www.patwardhans.net/papers/LallyEtAl14.pdf

20. Leemon, Baird. 2016. "The Swirlds Hashgraph Consensus Algorithm: Fair, Fast, Byzantine Fault Tolerance." https://www.swirlds.com/downloads/SWIRLDS-TR-2016-01.pdf.

21. Pearl, Judea. 2000. Causality: Models, Reasoning, and Inference. Cambridge: Cambridge University Press.

22. Piskovski, Victor, Alexander Grusho, Mikhail Zabezhailo, Andrey Nikolaev, Vladimir Senchilo, and Elena Timonina. 2020. Security Architectures in Digital Economy Systems, International. Journal of Open Information Technologies 8(9): 48-52.

23. Plotkin, Gordon. 1970. "A Note on Inductive Generalization." In Machine Intelligence 5. 153-164. Edinburgh: Edinburgh University Press.

24. Plotkin, Gordon. 1971. "A Further Note on Inductive Generalization." In Machine Intelligence 6. 101-124. Edinburgh: Edinburgh University Press.

25. Rudakov, Konstantin. 1986. Some universal restrictions for classification algorithms. Zh. Vychisl. Mat. Mat. Fiz. 26(11): 1719–1730.

26. SAS Institute. n.d. FORTUNE: 100 Best Companies to Work for. Accessed April 10, 2021. http://fortune.com/best-companies/sas-institute-4/

27. SAS Institute Inc. n.d. Patent applications. Accessed April 10, 2021. http://www.faqs.org/patents/assignee/sas-institute-inc/

28. SAS Institute Inc. 2013. Data Mining Using SAS Enterprise MinerTM: A Case Study Approach. http://support.sas.com/documentation/cdl/en/emcs/66392/PDF/default/emcs.pdf

29. SAS Institute Inc. 2015. SAS/STAT 14.1 User's Guide: High-Performance Procedures. http://support.sas.com/documentation/cdl/en/stathpug/68163/PDF/default/stathpug.pdf

30. SAS Institute Inc. 2015. Base SAS. High-Performance Procedures. Accessed April 10, 2021. http://support.sas.com/documentation/cdl/en/prochp/68141/PDF/default/prochp.pdf

31. Sobti, Rajeev, and Ganesan Geetha. 2012. Cryptographic Hash Functions: A Review. International Journal of Computer Science Issues 9(2): 461-479. https://www.researchgate.net/publication/267422045_Cryptographic_Hash_Functions_A_Review.

32. Stathakopoulou,

33. Su, Tzong-An, and Gultekin Özsoyoglu. 1987. "Data Dependencies and Inference Control in Multilevel Relational Database Systems." In Proceedings of the IEEE Symposium on Security and Privacy: 202-202.

34. TAdviser. 2020. Systems of BI and Big Data in Russia. Accessed April 10, 2021. https://www.tadviser.ru/index.php/BI.

35. VISA. Accessed April 10, 2021. https://usa.visa.com/partner-with-us/payment-technology/visa-b2b-connect.html.

36. Wikipedia. Merkle Tree. Accessed April 10, 2021. https://en.wikipedia.org/wiki/Merkle_tree.

37. Zabezhailo, Mikhail. 2014. Some capabilities of enumeration control in the DSM method. Scientific and Technical Information Processing 41(6): 335–361.

38. Zhu, Wei-Dong, Bob Foyle, Daniel Gagné, Vijay Gupta, Josemina Magdalen, Amarjeet Mundi, Tetsuya Nasukawa, Mark Paulis, Jane Singer and Martin Triska. 2014. IBM Watson Content Analytics: Discovering Actionable Insight from Your Content IBM Redbooks: IBM Corp. http://www.redbooks.ibm.com/abstracts/sg247877.html?Open

39. Zhuravlev, Yuri. 1977. Correct algebras on sets of incorrect (heuristic) algorithms. Kibernetika 4: 5–17.

40. Zhuravlev, Yuri. 1977. Correct algebras on sets of incorrect (heuristic) algorithms. Kibernetika 6: 21–27.;

41. Zhuravlev, Yuri. 1978. Correct algebras on sets of incorrect (heuristic) algorithms. Kibernetika 2: 35–43.

42. Zhuravlev, Yuri, Vladimir Ryazanov, and Oleg Sen'ko. 2006. "Recognition." Mathematical Methods. Software System. Practical Applications. Moscow: Fazis.

# System to Track Access in Digital Economy Systems

**Alexander Grusho**
*Federal Research Center Computer Science and Control Russian Academy of Sciences Russian Federation, Moscow*
**Victor Piskovsky**
*Federal Research Center Computer Science and Control Russian Academy of Sciences Russian Federation, Moscow*
**Mikhail Zabezhailo**
*Federal Research Center Computer Science and Control, Russian Academy of Sciences Russian Federation, Moscow*

**Abstract**

Leakage of protected data became an acute topic. The system tracking who, where and when has accessed a certain record of such a type of data could be a help in investigating vulnerabilities and weaknesses in defense. Also, it could name and point out responsible staff for the leakage with all legal and justice consequences. The paper considers an approach to build a system to register such kind of facts. The essence is to apply the distributed ledger technology, which is an open data storage. The system allows you to identify users who are trying to retrieve valuable information. At present, a technical and theoretical basis is ready for such solutions. Analysis of the current situation in the area under consideration shows that all the leading players in this segment of the IT market, in parallel with the development of mathematical models and methods of problem-oriented data mining, pay significant attention to the development of special software and hardware tools to support the performance of such tool solutions.

**Keywords:** information security, protection of valuable information from compromise on indirect grounds, distributed ledger, Intelligent Data Analysis, HPC

**Publication date:** 02.07.2021

**Citation link:**

Grusho A., Piskovsky V., Zabezhailo M. System to Track Access in Digital Economy Systems // Law & Digital Technologies – 2021. – V. 1. – № 1 C. 10-18 [Electronic resource]. URL: https://ldt-journal.com/S123456780015725-8-1 (circulation date: 29.04.2024). DOI: 10.18254/S123456780015725-8

---