

DISTRIBUTED ANALYSIS OF BIG PERSONAL DATA SETS WITH RESPECT TO PRIVACY

ALEKSEY PAKHOMOV, KONSTANTIN ZAYTSEV

National Research Nuclear University MEPhI (Moscow Engineering Physics Institute)
Kashirskoe avenue, 31, Moscow, 115409, Russia

ABSTRACT

Monitoring of cross-boundary financial operations routinely affects the interests of governmental and credit organizations of foreign countries which may not match the objectives of financial monitoring of investigation initiating country. Therefore, cross-boundary information requests should better be done in such a way to eliminate the possibility to identify objects of interest upon requests filed to databases of those organizations. Usually, Zero-knowledge proof technology is used. However, most scientific papers in this field are oriented at abstract study of the efficiency of algorithms (protocols) of interactive proof and assessment of their complexity, without solving concrete applied tasks. The objective of this paper is to seek an approach to anonymous comparisons of personal data and further use of that approach for financial monitoring of large organizations and international scientific projects. To solve the task set up, well-known protocols Fiat-Shamir, Guillou-Quisquater (GQ) and Schnorr were analyzed. As the basic, Fiat-Shamir protocol was chosen; it was modified by widening its use to the extent of checking of personal data storage facts in a database. The solution offered differs from known ones by combining Zero-knowledge proof technology with cryptologic transformation via hash function and is the most exact compared to frequently used approaches in that area and has reasonable productivity. Modified protocol was tested taking as an example identification of international terrorists among the list of financial operations parties.

Keywords: *Zero-knowledge Proof Password, Hash Function, Data Exchange Protocol, Anti-fraud, Money Laundering, Terrorist Financing, Scientific Collaboration*

1. INTRODUCTION

Collaboration of transnational organized crime and terrorist activities intensified in recent years under the influence of globalization processes, for example IGIL case. These trends require the international community to join efforts in fight with such new threats [1]. The integration of the international community can be achieved on the basis of collaboration of the scientists from different countries who demonstrate fundamental scientific methods that provide opportunities to fight crime using their research results. [2]. One of the major directions of joint activities of scientific collaborations is to monitor cross-boundary financial flows for experiments of scientific groups (financial monitoring) in order to develop methods to identify criminal financial transactions, terrorist organizations or their associations. Monitoring of cross-boundary financial transactions affects the interests of academic and banking communities around the world and involves various experts in these sectors of activity. Financial monitoring process involves different aspects, and in accordance to [3], can combine such operations as

detection of cross-boundary cash flows for experiments on mega science level, creation of thematic projects and analytical profiles, analysis of cross-boundary criminal networks. Risk analysis performed by scientific collaborations in the countries around the world will significantly improve financial monitoring process.

Experts that perform financial monitoring of transnational flows could include persons that are subjects of money laundering. That was confirmed by investigations held around the world. There are cases when the money laundering is achieved with the latest technologies, therefore, the same level of technology to be applied to combat money laundering. Otherwise there is a risk of losing the ability to withstand financial crime at the global world market of scientific achievements. The analysis shows that the individual measures to combat highly organized activity, even if they were taken by large financial institutions, are not effective when made through a single institution database [4]. At the same time the introduction of new technologies of information security based on carried out experiments in fundamental science



becomes a prerequisite for detection of fraud schemes.

Thus, there is the problem of control operations, so that it is not possible to determine the objects of interest of requests received in the scientific and financial institutions. This problem is caused by the contradiction between the need to obtain a large amount of factual data and the need to conceal the purpose and content of such data exchange from fraudulent use or privacy violation. One of the most promising approaches to the implementation of secure financial monitoring is an interactive technology Zero-knowledge proof (proof with zero disclosure of information), proposed in 1986, and subsequently modified by Shafi Goldwasser and colleagues [5, 6]. This is a purely academic research direction, but in its practical implementation it allows the relying party to check the accuracy of some mathematical statements without receiving any other information from the auditee (proving) side. This is very important, because it guarantees that the relying party cannot independently prove this to anyone else. Using this approach, analysts can send requests to any scientific and financial institutions, without fear that the purpose of the request will be revealed for the respondent. Technology of Zero-knowledge proof is being actively developing nowadays. However, most research in this area focuses on the fundamental study of the effectiveness of interactive algorithms evidence and assessment of their complexity [7, 8], on the other hand less research works could be found on approaches to the solution of applied problems with the use of heuristics domains. At the same time, we can assume that the analysis of anonymized data from several research institutions can help detection of financial crimes related to distributed attacks.

Therefore, development of special algorithms to solve the applied tasks of financial monitoring using Zero-knowledge proof technology is important both from the standpoint of using that technology in transfer of various types of requests and from the standpoint of identifying the areas for efficient use of classic protocols under that technology.

This paper is devoted to the development of an algorithm of distributed processing of personal data to protect their security from disclosure and to resist the above threats in the course of financial investigations.

The above features of Zero-knowledge proof technology enabled the authors to move a

hypothesis that the said technology may be put as the cornerstone of cross-boundary data exchange for financial monitoring purposes to ensure solving such urgent tasks of financial monitoring as speed-up and simplification of required information gaining, observance of the regulations on personal data protection, resistance to the above threats for financial monitoring subjects.

2. MATERIALS AND METHOD

2.1. Comparison of protocols

Here are the basic assumptions of the research paper. The disadvantage of simple password protocol when transmitting confidential information while the prover A gives B the password verifier, is that B can threaten security by disclosure of A's password. Zero-knowledge protocol allows to demonstrate knowledge of secret, without disclosing the secret itself. Interactive proof systems are based on the zero-knowledge protocols. Here are some specification of the problem based on financial monitoring case. The main subjects of zero proof protocol are the next. Prover A is a person trying to prove knowledge of the secret. Tester B is a person checking the knowledge of the secret. The first message is sent from A to B on the three round authentication protocol, called the request. The second message is sent from B to A, called the challenge. The final message sent from A to B is called the response.

ZK protocols must satisfy three properties: completeness, soundness and zero-knowledge. The property of completeness means that if the statement is true, then the verifier must be notified in any cases. Justification is that if the statement is false, proving side must be unable to convince the examiner that it was true, except for negligibly small probability. Condition of Zero knowledge proof is that if the statement is true, then the reviewer does not know anything except this fact (the statement is true). Most ZK protocols have three rounds. This means that three messages are transferred between the parties: request, challenge and response. Typically, the protocols ZK metrics are probability of mistake and duration of the interaction. Probability in the request is used to hide secret information. Duration is used to prevent responsible to spend a lot of time on the calculation of the response

We have compared a number of frequently used protocols of ZK: Fiat-Shamir, Guillou-Quisquater (GQ) and Schnorr [9-15]. Each protocol has its advantages and disadvantages.

**Fiat-Shamir protocol**

The purpose of the Fiat-Shamir protocol - to prove knowledge of the secret examiner s B t for iterations of interaction. This probabilistic protocol with probability 2- t for the enemy to deceive the verifier. Fiat-Shamir protocol performs from t = 20 to 40 operations, the probability to deceive the verifier for all t operations is very low. Fiat-Shamir protocol - three round protocol based on the complexity of the factorization problem.

Initial data

A) A trusted center (T) chooses RSA-modulus as $n = pq$, n - public key, p , and q - secrets;

B) A chooses s prime to n , $1 \leq s \leq n - 1$, computes $v = s^2 \bmod n$, and registers T and v , v - the public key, s - secret key.

Protocol

A) A chooses a random challenge r , $1 \leq r \leq n - 1$;

B) A sends B (1): $x = r^2 \bmod n$;

C) B sends A (2): a random e , $e = 0$ or $e = 1$;

D) A sends B (3): $y = r \cdot s^{e \bmod n}$.

Check

A) reject if $y = 0$

B) B receives, if $y^2 = x \cdot v^{e \bmod n}$, otherwise rejects.

Guillou-Quisquater (GQ) protocol

The purpose of the protocol GQ - to provide proof of knowledge in t times interaction. This is the probabilistic protocol with probability $v - t$ to deceive the verifier. Since the range of possible values is, in the range from 1 to v then if v is very large, the probability to deceive the verifier becomes very small. GQ is an extension of the Fiat-Shamir.

Initial data

A) The trusted center (T) chooses RSA-module as $n = pq$, n - public, p and q - secrets;

B) T selects exponent v , $v > 3$, and the GCD (v , ϕ) = 1, where $\phi = (p-1)(q-1)$, $s = v^{-1} \bmod \phi$, v - public, s - secret.

Parameters for each user

A) Being A is given a unique identifier IA;

B) reserve identity $JA = f(IA)$, where $1 < JA < n$, which satisfies GCD (JA , ϕ) = 1, f - public function;

C) T gives A: $sA = (JA)^{-s} \bmod n$, s - secret.

Protocol

A) A selects a random challenge r , $1 \leq r \leq n - 1$

B) A sends B (1): IA, $x = r^{v \bmod n}$;

C) B sends A (2): a random e , $1 \leq e \leq v$

D) A sends B (3): $y = r \cdot s \cdot A^{e \bmod n}$.

Check

A) B computes $JA = f(IA)$;

B) computes $z = J(A)^e \cdot y^{v \bmod n}$.

C) B rejects if the $z = 0$

D) A proves the knowledge to B, if $z = x$, otherwise A is rejected

Schnorr protocol

The purpose of Protocol Schnorr is to make A prove his identity to B. Schnorr protocol is three round protocol which depends on the complexity of computing discrete logarithms.

Initial data

A) A chooses p such that $p - 1$ is divisible by q ($p = 2 \sim 1024$, $q > = 2 \wedge 160$), p , q - public;

B) β is chosen, $1 \leq \beta \leq p - 1$, bearing the rank q , α - generator of mod p ,

$\beta = \alpha^{((p-1)/q) \bmod p}$, β - public

C) t is chosen, $t > = 40$, 2^t , $t < q$.

Parameters for each user

A) A chooses a secret key a , $0 \leq a \leq q - 1$;

B) A computes $v = \beta^{-a \bmod p}$.

Protocol

A) A selects a random challenge r , $1 \leq r \leq q - 1$;

B) A sends B: $x = \beta^{r \bmod p}$;

C) B sends A: a random e , $1 \leq e \leq 2^t$, $t < q$.

D) A sends B : $y = a \cdot e + r \bmod q$.

Check

A) computes $z = \beta^y \cdot v^{e \bmod p}$

B) B approves, if $z = x$, otherwise rejects.

Comparing protocols

The basic characteristics by which ZK protocols differ are as follows: complexity of connection, complexity of computations, memory, security guarantees and confidence parameters requested by a third party.

Complexity of connection is characterized by the number of messages exchanged between the proving and checking persons, complexity of computations - by the number of operations (modular multiplications) for proving and checking parties accounting for online and offline computations. Security guarantee is the level of resistance against falsification and disclosure of secret information. Finally, confidence parameters vary in various protocols.

The most important among the above characteristics from the practical standpoint in financial monitoring, are complexity of connection and computations.

Complexity of connection

Fiat-Shamir protocol requires 20-40 operations of the three round protocol (60-120 separate messages), while GQ and Schnorr protocols require only 1 operation.

Complexity of computations

Fiat-Shamir protocol has the advantage in complexity of computations – for this protocol they are the most simple. Schnorr protocol has the advantage as it requires only single online computation by the proving party while the computation volume is much higher compared to Fiat-Shamir and GQ protocols.

As the basis for implementation of the algorithm for distributed processing of personal data without disclosing them any of the above protocols fits; to be definite, we chose Fiat-Shamir protocol.

2.2. Description of the modified Protocol

To solve the problem of anonymous comparison it is necessary to clarify the following questions. How to choose a set of data that uniquely identifies the person? How to create a string of personal data independent of the language description? How to form strings for a list of persons presented in a financial intelligence unit watch lists. In addition it is necessary to determine:

1. Compression/ coding / cryptography method for input string that must allow only one-way computation.
2. Fuzzy matching mechanism for substring (description of one person) and rows (a description of all of the people FIU) of encrypted personal data comparison.

3. The first three questions are resolved in the organizational and legal fields, while the following two objectives could be solved by usage of "protocol with zero knowledge proof + hash function".

Let us present results of protocol application. In order to make the basic idea clear, let us present the problem in the example of Alice (A) and Bob (B).

Lets form the main terms of the scenario: "W" - a set of data that is held by Bob, "s" - the secret message of Alice, "w" - an element that forms W dataset.

Alice has secret "s", it must find out whether "s" is the subset of Bob's dataset (database "W").

But she can not disclose personal information especially if Bob hasn't got "s" in his dataset, so

Bob mustn't know neither the "s" nor the result of matching after interaction.

Bob and Alice cannot provide information about the contents of their datasets. This work is based on the protocol of zero-knowledge proofs that was improved using substrings of hash functions, to make the search less time-consuming, sacrificing privacy of data. Rounds of checks will be mixed with the false secret rounds, which make the algorithm less vulnerable to brute-force attack. This is done to make hash function usage less threatening.

Initial data Protocol

The problem can be solved with the use of zero-knowledge protocol (ZKP - Zero-knowledge protocol) for small dataset W, the protocol must be repeated for |W| (where || is used to show the cardinality of set) rounds for the decision. Since protocol is interactive, its working time depends both on complexity of ZKP algorithm computation and t- time of interaction caused by data exchange.

To reduce the time needed for set intersection Bob must form the set "Ws". It is a subset of "W", which must include the element "w" that is equal to secret, if the "w" exists in the whole data set "W". Protocol mustn't allow "w" to be recognized as an object of interest of Alice during interaction.

To insure that "w" is in the original set and was chosen to form "Ws" set we need to form a rule - the predicate "P" that is true for the "s" and a number of elements from "W" that will be chosen to form "Ws". An additional requirement is a condition that the result of "P (s)" should be insufficient to determine any of information about "s" except the information that P(s) is true.

It is proposed to use the complex P as a union of predicates (p1..pn). Only subset of "p" predicates are true for "s" that makes interaction more safe for Alice secret. On the other hand, |Ws| should be much smaller than |W|, as the computational complexity of the algorithm depends on O(|Ws|).

To construct P Alice can use a hash function that has much smaller cardinality than the cardinality of hash("s"). This will lead to collision growth between hash("W") elements. In order to conceal information about the real identity of "s" from Bob, hash collisions reduce the probability of guessing the real Bob "s". Hash function will give more coincidences with increase of |Ws|.

For further improvement of the "s" secrecy Alice may also add false secrets "Sf" set as arguments of

the hash function, so that if Bob receives $|Ws|$ equal to 1, he cannot determine that "Ws" = "s" with confidence greater than $1 / |Sf|$.

As a result, the likelihood that Bob will know hash("s") (if "s" is a member of the "W") is increased by:

- probability that $|Ws| = 1$, if P is to strict; reduced by:
- the number of collisions that can hash function produce from "s" argument;
- the number of false secrets $|Sf|$.

Protocol algorithm

1. Alice builds a result set of hash functions results that are formed by "Sf" arguments, where $Sf = F \cup s$,

2. She mixes the set "S" and sends it to Bob.

3. Bob gets "S". Bob sends a message that he received challenge and he has formed Ws that is not null ($|Ws| > 0$), where "Ws" is set of elements that make predicate P("ws") true (Ws is formed as a subset of "W" - the full dataset of Bob). If $|Ws| = 0$ algorithm returns a false result (Bob don't have any information on Alice's secret). Alice checks hypotheses $ws = secret$ with the ZPF protocol for all of the elements of "Ws". If ZPF has at least one positive response on a "ws", then the algorithm proves that Bob has got information about secret with certain probability, otherwise - it is the fact that he doesn't.

Thus, Alice receives information only about whether Bob knows anything about "s" in the "W" dataset. Bob knows only that the "s" may be an element of "Ws" with the probability equal to $1 / |Ws|$. Furthermore, he knows a hash function result with an argument of "s" with a probability of $1 / |Sf|$.

Algorithm weaknesses:

A non-zero disclosure - if Bob has other rogue Kane, which is a subgroup of "Ws", he could mislead the request by removing information from Ws before ZKP starts.

Security problems / speed - if the number of collisions that was produced by hash function in W from secret is equal to 1 and $|Sf| = 2$ the security of secret from Bob is not guaranteed as he knows exactly one person that is the person of interest, otherwise the computational complexity of $O(|Ws|)$ algorithm became inefficient for real application without that hash optimization. The problem can be solved if the algorithm of "S" formation and a hash

function could be found. The basic requirements for such "S" and a hash function is to minimize the ($|Ws|$) and probability of a situation when $|Ws| = 1$.

Fuzzy matching - since the result of the algorithm is a Boolean value, based on the cryptosafe one-way hash function, the slightest changes in the "S" and "W" that lead to the errors in writing or poor quality of the data. It means that the results of exchange are right with probability equal to probability that a name is written correctly at secret or Alice or Bob must try completely all spellings of secret information. String standards must be determined before the interaction starts between Alice and Bob. They must ensure that they use the same character set and a set of attributes for the "S" and "W", it could be done for example by alphabet exchange.

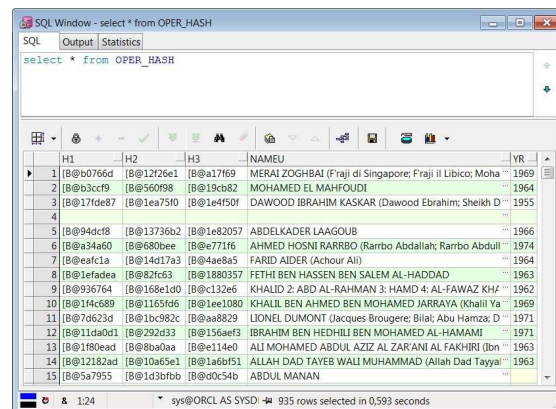
Selection of zero-knowledge algorithm involves additional preparation of input data that is necessary to apply the selected protocol implementation of ZKP for matching strings, that may lead to additional security problems.

3. RESULTS

3.1 The data for testing

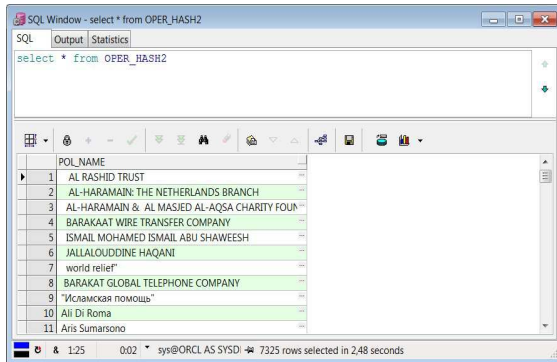
The Data for testing are two tables. The first table (Table 1) contains 935 records. It is publicly available information from international lists of terrorists, as well as hashed entries by hash functions. The second table (Table 2) contains 7325 entries. It is a list of test participants in financial transactions.

Table 1. List Of Persons And Images Of Hash Entries



H1	H2	H3	NAMEU	YR
[B@b0766d]	[B@12f26e1]	[B@a17f69]	MERAI ZOGHBAI (F'raji di Singapore; F'raji il Libico; Moha	1969
[B@b3ccf9]	[B@560f98]	[B@19cb82]	MOHAMED EL MAHFOUDI	1964
[B@17fde87]	[B@1ea75f0]	[B@1e4f50f]	DAWOOD IBRAHIM KASKAR (Dawood Ebrahim; Sheikh D	1955
[B@94dcf8]	[B@13736b2]	[B@1e82057]	ABDELKADER LAAGOUR	1966
[B@a34a60]	[B@680bee]	[B@e771f6]	AHMED HOSNI RARRBO (Rarrbo Abdallah; Rarrbo Abdull	1974
[B@eaf1a]	[B@14d17a3]	[B@4ae8a5]	FARID AIDER (Achorou Aii)	1964
[B@1efafdea]	[B@93fcd3]	[B@1880357]	FETHI BEN HASSEN BEN SALEM AL-HADDAD	1963
[B@936764]	[B@168e1d0]	[B@c132e6]	KHALID 2. ABD AL-RAHMAN 3. HAMD 4. AL-FAWAZ KHf	1962
[B@14c689]	[B@1165fd6]	[B@1ee1080]	KHALIL BEN AHMED BEN MOHAMED JARRAYA (Khalil Ya	1969
[B@7d623d]	[B@1bc982c]	[B@aa8829]	LIONEL DUMONT (Jacques Brougere; Bilal; Abu Hamza; D	1971
[B@11da0d1]	[B@292d33]	[B@156aef3]	IBRAHIM BEN HEDHILI BEN MOHAMED AL-HAMAMI	1971
[B@1f80ead]	[B@8ba0aa]	[B@e114e0]	ALI MOHAMED ABDUL AZIZ AL ZARANI AL FAKHRI (Ibn	1963
[B@12182ad]	[B@10a65e1]	[B@1a6bf51]	ALLAH DAD TAYEB WALI MUHAMMAD (Allah Dad Tayyal	1963
[B@5a7955]	[B@1d3bfbf]	[B@d0c54b]	ABDUL MANAN	

Table 2. Test The List Of Participants Of Financial Transactions



id	POL_NAME
1	AL RASHID TRUST
2	AL-HARAMAIN THE NETHERLANDS BRANCH
3	AL-HARAMAIN & AL MASIED AL-AQSA CHARITY FOUR
4	BARAKAT WIRE TRANSFER COMPANY
5	ISMAIL MOHAMED ISMAIL ABU SHAWEEH
6	JALLALOUDDINE HAQANI
7	world relief
8	BARAKAT GLOBAL TELEPHONE COMPANY
9	"Асская помощь"
10	Ali Di Roma
11	Aris Sumarsono

Table 3. The Results Of The Algorithm Using A Hash Function

The number of characters of the hash time	Time, s	With positive	False positives / alarms
no	2210.728	14406	0.9857
1	109.376	598	0.6555
2	7.527	236	0.1271
3	0.832	208	0.0096
4	0.468	206	0
5	0.376	206	0

3.2 Implementation of the Protocol

To test the proposed algorithm for hashing, we have used common functions (MD4, md5, sha1). To test the hypothesis that ZKP authentication protocol with zero knowledge of Fiat-Shamir (which is presented in section 1.2) was applied for solving of problem. The protocol implementation revealed that comparison of strings is computationally difficult as it has $O(m+n)$ computational complexity where m and n are the cardinalities of Bob and Alice databases. Protocol requires the use of numeric secrets, and the input data is strings. Therefore, we use a hash function that allows to get all the input array to the fixed-length string. After that the process of comparison becomes much easier. For the efficiency of the algorithm Alice and Bob must define such hash function and the number of false secrets that will be enough for secrecy and that will reduce the computational complexity of verification

3.3 Testing the speed and efficiency

The algorithm was applied to compare the database of real operations (transactions), and real global list of terrorists that can launder money through the research projects. The main metrics are the number of real test (negative) and false (positive) responses (Table 3). To improve the timing hash optimization has been applied.

The time required to perform the checks was measured on a test computer stand with the following features: i3 CPU, 8GB RAM, database Oracle 11g r2. Results of failed rounds are shown in the diagram (Fig. 1). The theoretical distribution curve has the form $1 / (2^n)$, which is proven by the test.

3.4 Results

The algorithm has shown good results, but should be tested using a variety of ZKP and hash function for finding optimal solutions to identify metrics which are speed, time, security, accuracy, false positives. If it proves to be effective, it can be used for a large scale tasks connected with exchange of information between different parts of the institutions supporting the implementation of mega science projects in various areas, FATF style regional bodies and other international regulatory organizations. This can significantly improve the safety of experiments conducted by the international collaboration of scientists, to strengthen cooperation between the governments and organizations involved in safety by reducing the time required to fulfill the requests.

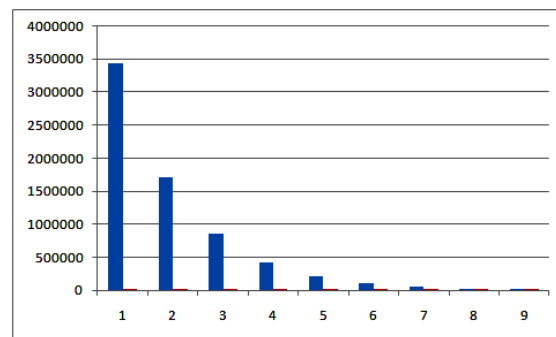


Figure1. Diagram Of Unsuccessful Rounds Of The Algorithm.

Practical adaptation and use of the most of ZKP algorithms is the main result of this work and opens the possibilities for the secure fight exchange of data between organizations in the fight against fraud in



large scientific projects, without violating of the rights of members of the private collaborations.

4. DISCUSSION

Using of Zero-knowledge proof technology in a systems that create privacy preserving requests for information exchange enhances overall economical security of projects while compared with watch lists of terrorist and suspicious persons. The presence of specific persons or organizations in the databases of the major collaborations of scientists, financial, credit and supervisory authorities is critical to get data from those databases. We should solve various technical issues connected with accurate identification of personalities that is complicated by variability in fillings of various documents. Another direction is efficiency of the algorithms of identification and hash functions. There are additional questions that are connected with non-technical aspects of exchange including problems related to the political, cultural and infrastructural features of individual countries and territories that are widely debated by researchers of globalization.

Thus, in the pioneering papers [5, 6] it was actually introduced a new branch of mathematics, aimed at interactive probability theorem proving without possible confidential information leak. It gives an opportunity to prove without opening the full theorem. That was the basic philosophical question for cryptology research on the one hand [15-16] and epistemological research of mathematics and theoretical computer science [17] - on the other. Another direction of research include such work [18] that shows that there does not exist two-round zero-knowledge proof system with perfect completeness for an NP-complete language; there does not exist a constant-round zero-knowledge strong proof or argument of knowledge and there does not exist a constant-round public-coin proof system for a non-trivial language that is resettable zero knowledge.

The paper [6] proposed procedure with some restrictions on the number of iterations that has better efficiency. This is an attempt to prove the existence of such restrictions iterations for all languages considered in the article.

It describes the building such procedure with a limited number of steps on the basis of modeling paradigm. The article [8] provides an overview of the most important works in the field of Zero-knowledge proof that gives the direction of development. Also it describes the research tools for the design of reliable cryptographic systems.

Another work [12] introduces the concept of hybrid trapdoor and on the basis of this concept is considering hybrid nature obligations under various schemes of distribution of obligation under evidence. Here and unconditional and miscellaneous liabilities and so-called hybrid commitments and based on them.

There were works [19], which described some of the computer systems that are able to solve applied problems with the use of other technologies, but unfortunately algorithms are not clear from the descriptions, respectively, it is not possible to assess the effectiveness of these "closed" systems. Thus, the proposed approach to the anonymous comparison of personal data based on the protocol with zero proof of disclosure of information is extremely popular for a variety of practical problems in the field of financial monitoring.

5. CONCLUSION

The use of the approach offered in this paper to anonymous comparison of personal data based on modified Fiat-Shamir protocol is rather in demand for many practical tasks in financial monitoring for international governmental and credit organizations, large scientific projects as it ensures the effect of anonymizing data while disclosing the fact of their existence. The algorithm described here allows to conclude on the following.

Anonymous comparison of personal data is based on Zero-knowledge proof technology, therefore the three most frequently used protocols Fiat-Shamir, Guillou-Quisquater (GQ) and Schnorr should be analyzed by the following parameters – complexity of connection, complexity of computations, memory used, security guarantee as well as confidence parameters requested by a third party, to further choose the best of them as the basic one. As such, Fiat-Shamir protocol was chosen.

To use Fiat-Shamir protocol for anonymous comparison of personal data it is required to additionally solve the tasks on identifying a set of personal data uniquely identifying the investigation object, method of description in the string of multitude of investigation objects independent or dependent from the language of description; identifying the way of compression/coding/cryptography of strings preventing reverse reproduction of the data transferred and viewing mechanism for input by matching or close similarity of the transferred



substring in database string for encrypted personal data.

The solution of the above tasks within Zero-knowledge proof technology modifies Fiat-Shamir protocol, widening it in connection with checking facts of personal data storage in databases of international organizations and combining Zero-knowledge proof and hash functions for optimization of requests.

The test of the offered modified Fiat-Shamir protocol taking as an example anonymous comparison of personal data within financial monitoring of verifying whether there are any international terrorists in the list of financial operations parties showed its efficiency – it is the most precise compared to other frequently used algorithms and reasonable productivity.

REFERENCES:

- [1] Zubkov, V., & Osipov, S., (2006) Russian Federation in the international system of combating legalization (laundering) of criminal income and financing of terrorism. Moscow, Publishing house "Gorodets", 2006, 752 p.
- [2] L.M. Hinkey, K.T. Ellenberg, B. Kessler (2005), Strategies for Engaging Scientists in Collaborative Processes. - Extension Journal, Volume 43, Number 1, February 2005. e, joeed@joe.org.
- [3] Ma³tch (2013) <https://www.fiu.net/fiunet-unlimited/match/match3>
- [4] D. A. Flores, Olga Angelopoulou, R. J. Self, An Anti-Money Laundering Methodology: Financial Regulations, Information Security and Digital Forensics Working Together <http://isyu.info/jisis/vol3/no12/jisis-2013-vol3-no12-07.pdf>
- [5] Goldwasser, S., Micali, S. & Rackoff, C. (1989), The knowledge complexity of interactive proof systems, *SIAM Journal on Computing (Philadelphia: Society for Industrial and Applied Mathematics)* . — *T. 18 (1): 186–208*, ISSN1095-7111, doi:10.1137/0218012.
- [6] Goldwasser, S., & Micali R (2013) ACM Turing Award for Advances in Cryptography. ACM. Turing Award for Advances in Cryptography. ACM. Retrieved 13 March.
- [7] Zhang, Z., Cao, Z., & Zhu, H. (2014). Constant-round adaptive zero-knowledge proofs for NP Information Sciences, Volume 261, 10 March, 219-236.
- [8] Li, F., & McMillin B. (2014). Chapter Two - A Survey on Zero-Knowledge Proofs. *Advances in Computers*, Volume 94, 25-69.
- [9] Feige U., Fiat A., Shamir A. Zero-knowledge proofs of identity, *J.Cryptology* 1, 1988
- [10] Cheng-Fen Lu, Shiuhyng Shieh, Efficient Key-Evolving Protocol for the GQ Signature (2004). - *Journal of information science and engineering* 20, 763-769.
- [11] Varnovsky N. P., *Cryptographic protocols // Introduction to cryptography / edited by V. V. Yaschenko. — Peter, 2001. — 288 p. — ISBN 5-318-00443-1.*
- [12] Catalano, D. & Visconti, I. (2007). Hybrid commitments and their applications to zero-knowledge proof systems. *Theoretical Computer Science*, Volume 374, Issues 1–3, 20 April, 229-260.
- [13] Guillou, LC, & Quisquater, JJ (1990), A "paradoxical" identity-based signature scheme resulting from zero-knowledge proof idea. In *Proceedings on Advances in cryptology* (pp. 216-231). Springer-Verlag New York, Inc.).
- [14] Schnorr CP, Efficient Identification and Signatures for Smart Cards. - *Advances in Cryptology - CRYPTO'89. Lecture Notes in Computer Science* 435. - 1990. - S. 239 - 252.
- [15] Yao, Andrew C.C., Yao, Frances F. & Zhao, Yunlei (2009). *Theoretical Computer Science*, Volume 410, Issue 11, 6 March, 1099-1108.
- [16] Lin X. J., Sun L., & Qu H. (2015). Insecurity of an anonymous authentication for privacy-preserving IoT target-driven applications. *Computers & Security*, Volume 48, February, 142-149.
- [17] Bledin J. (2008). Challenging epistemology: Interactive proofs and zero knowledge. *Journal of Applied Logic*, Volume 6, Issue 4, December, 490-501.
- [18] Barak, B., Lindell, Y., & Vadhan, S. (2006). *Journal of Computer and System Sciences*, Volume 72, Issue 2, March, 321-391.
- [19] Bogdan Carbunar, Mahmudur Rahman, Jaime Ballesteros, Naphtali Rishe, Eat the Cake and Have It Too: Privacy Preserving Location Aggregates in Geosocial Networks - <http://arxiv.org/pdf/1304.3513.pdf>.