PROTECTION OF VALUABLE INFORMATION IN PUBLIC INFORMATION SPACE

Alexander A. Grusho, Nick A. Grusho, Michael I. Zabezhailo and Elena E. Timonina
Federal Research Center "Computer Science and Control"
of the Russian Academy of Sciences
Vavilova 44-2, 119333, Moscow, Russia
Email: grusho@yandex.ru, info@itake.ru, zabezhailo@yandex.ru, eltimon@yandex.ru

KEYWORDS

Protection of valuable information, public information space, ways of appearance of valuable information in public information space.

ABSTRACT

In some cases a valuable information forcedly appears in public information space. At the same time, as well as in other cases, protection of valuable information is necessary.

In the paper three classes of scenarios of the compelled appearance of valuable information in public information space are considered. Also several ways of protection of valuable information in these cases are suggested. The main problem consists in the fact that in public information space all information can be available to the adversary.

The first class of scenarios of appearance of valuable information in public information space is connected with disclosure of the directions of information search, intended for special researches. Information about researches is valuable and also the directions of its appearance in public information space is valuable information.

The second class is connected with the need of the open publication of a part of valuable information. The third class is connected with creation of ambiguity in attempts of a conclusion of valuable information from the open publication of its usage in results of its application in information technologies.

INTRODUCTION

The problem considered in the paper is formulated as follows. Let's consider the two-level system of classified information. Valuable information is at the level High, and information which is not valuable (further public), is at the level Low where it can be available to an adversary. If the information technology uses valuable and public information, then by the rules of MLS policy (TCSEC, 1985) the results of the information technology should remain at the High level.

However, there is a number of situations when the results of information technology should be present at least partially at the Low level. In these cases for protection of valuable information different methods are used. For example, in many databases of scientific information the summaries are considered as public information, and the results are considered as valuable information since they are sold for money.

Other example is the document (TCSEC, 1985). This standard is the document for public usage, however it is result of in-depth mathematical studies which were put the reasons of security of the classified information. Such reasons in this standard are absent, but its requirements are obligatory for all vendors which produce the products for processing of the classified information.

The listed above examples do not exhaust methods of protection of valuable information in public information space. For deeper analysis of the problem it is necessary to construct models of valuable information and to describe ways of usage of valuable information in public information space. On the basis of these models it is possible to construct methods of protection of valuable information and to estimate its security.

Without applying for completeness, in work three classes of models of valuable information and ways of its usage in public information space are constructed. Each of classes demands various approaches to information security which differ from multilevel security policy (TCSEC, 1985).

The first class of models is connected with formation of input data for information technologies. The second class of models is connected with splitting valuable information on the basis of a compromise between obligatory belonging of information to the High level and a possibility of its partial usage at the Low level.

The third class of models allows to form for protection of valuable information ambiguity in attempts of recovery of valuable information according to results of its usage in information technologies.

The article is structured as follows. Section 2 introduces models of formation of input data for information technologies. Section 3 defines models of splitting of valuable information. In Section 4 we describe model of formation of ambiguity in attempts of recovery of valuable information according to results of its usage in information technologies. In Conclusion we analyze the results.

MODELS OF FORMATION OF INPUT DATA FOR INFORMATION TECHNOLOGIES

Let's consider the problem of protection of the valuable information used for formation of input data in information technologies on the following examples.

Communications of the ECMS, Volume 33, Issue 1,
Proceedings, ©ECMS Mauro Iacono, Francesco Palmieri,
Marco Gribaudo, Massimo Ficco (Editors)
ISBN: 978-3-937436-65-4/978-3-937436-66-1(CD) ISSN 2522-2414

Example 1.

For involving of a large number of specialists doctors for carrying out the analysis of diseases the public (for doctors) database of inspection of patients is created. Research of information from it the database should not open personal data of patients.

Personal data are considered as valuable information. For protection of personal data conditional identifiers of patients (indexes) are used, and in all data personal data are replaced with such indexes. The choice of data for the analysis and results of information technologies of the analysis of these data include indexes, but do not open personal data. Indexes accompany all data of researches of the patient and the description of the course of its treatment.

It is considered that depersonalization is a good protection of personal data. Because except the protected High level database the conformity of indexes and personal data cannot be anywhere disclosed. In this case valuable information is torn off from an information technology of data analysis. And therefore valuable information does not appear at the Low level in an explicit form.

However when forming an input data for an information technology by means of valuable information there can be more complex problems.

Example 2.

Let's consider the High level analytical center and the Low level system of data collection for the High level system in the public information network. The service of the Low level should fill DBMS from which the system of the High level takes information for the analysis. As the tasks interesting for the High level analytical center are valuable information, such DBMS has the High level. In this case a protection of DBMS corresponds to protection of a High level system, and channels from the Low level to the High level are unidirectional (multilevel security policy (TCSEC, 1985).

However it is possible only in rare cases. It is impossible to fill DBMS with all possible potentially necessary information. Then the service of a High level system should define the directions of information search in public information space and transfer these directions to the Low level. The service of the Low level will organize search in the specified directions, and, thereby, gives to an adversary the area of work and tasks of the High level.

Therefore for protection of the valuable information consisting in the directions of search of input data for analysts, only the method of expansion of field of search is known.

Let's construct the simple model of necessary expansion of the field of search. Let $\overline{x}$ is valuable information and consists of objects $x_1, ..., x_N$. Each object is described by some subset of the set $U = \{a_1, ..., a_m\}$ of characteristics.

For protection of valuable information $\overline{x}$ we will add false objects $y_1, ..., y_t$ for searching which are described by the same values of characteristics. Then elements of sets of objects $\{x_i\}$ and $\{y_i\}$ can be considered as subsets of the set U. The similar model was considered in the book (Anshakov, 2009) in connection with the analysis of search of the empirical reasons in semistructured data.

It is obvious that without set $\{y_i\}$ valuable information is determined unambiguously by observations of search process.

Let set $\{y_i\}$ defines false purposes of search, and they are more large-scale, than the purposes of $\overline{x}$. Then there forms "myth" that the set $\{y_i\}$ defines main objectives of search of analytical center, and the unnecessary purposes from the set $\{x_i\}$ create an artificial hindrance for the purposes from the set $\{y_i\}$. There is an ambiguity which can lead to mistakes of an adversary.

MODELS OF SPLITTING OF VALUABLE INFORMATION

Examples of compromises at placement of a part of valuable information in public information space are briefly given at the beginning of the article. For knowledge of the experts who are looking for results of scientific research there is enough name of article, name of the author and summary. At the same time full results and their substantiation have to be bought. The compromise consists in the fact that the researcher buying results and their substantiation obtains additional information which for it is valuable.

Deeper consideration of a problem of classification of results (further object O) of any information technology leads to rather complex problem.

Let's consider a manual algorithm of classification of output information of information technology at High and Low levels.

1. The list (description) of valuable information, i.e. information at the High level is created.

2. Criteria of reference O to valuable information are established.

3. Representation form of candidate for classification is regulated.

4. The order (algorithm) of classification, i.e. reference of information to the High level is defined. Information which is not referenced to the High level is considered information of the Low level.

In the automatic mode classification gets elements of stochastic character. It is connected with implementation of item 3 above. Really, the data representation form for classification is not determined and can be modelled by some random process. Then the problem of classification is the procedure of statistical check of the hypothesis that an information for classification has the Low level, against the alternative that the provided information contains High level elements. Repeated requirements of classification of the output data of information technologies are possible. We receive model of many small samples.

Applying traditional methods of mathematical statistics, we have possibility of appearance of "false alarm" (i.e. to classify information, as belonging to the High level though it belongs to the Low level) and errors of the random admission of elements of the High level at the Low level. In case of classification of valuable information, it is inadmissible. However, the known effect of many small samples (Axelson, 1999) shows that decrease in probability of a "false alarm" considerably increases the probability of classification of valuable information of the High level as information of the Low level.

Therefore, traditional methods of statistical test of hypotheses are not suitable for this task. In this case it is possible to use the theory of the bans of probability measures on finite probabilistic spaces (Grusho et al., 2013).

Let X_i, $i = 1, 2, ..., n, ...$, be nite sets, $\prod_{i=1}^{n} X_i$ be Cartesian product of X_i, $i = 1, 2, ..., n$, X^∞ be the set of all sequences when i-th element belongs to X_i. Dene $\mathcal{A}$ be a -algebra on X^∞, generated by cylindrical sets.

Define probability measure P on $(X^\infty, \mathcal{A})$. Assume P_n be the projection of measure P on the rst n coordinates of sequences from X^∞. Define

$$X_n^\infty = \prod_{i=n+1}^{\infty} X_i.$$

It is clear that for every

$$B_n \subseteq \prod_{i=1}^{n} X_i$$

the following ratio is carried out

$$P_n(B_n) = P(B_n \times X_n^\infty).$$

Let

$$\overline{x}_k \in \prod_{i=1}^{k} X_i,$$

then $\tilde{x}_{k-1}$ is obtained from $\overline{x}_k$ by dropping the last coordinate.

Definition 1. Ban (Grusho et al., 2013) of measure P_n for each n is a vector

$$\overline{x}_k \in \prod_{i=1}^{k} X_i,\ k \le n,$$

such that

$$P_n(\overline{x}_k \times \prod_{i=k+1}^{n} X_i) = 0.$$

If

$$P_{k-1}(\widetilde{x}_{k-1}) > 0,$$

then $\overline{x}_k$ is the smallest ban (Grusho et al., 2013).

Problems of test of hypotheses $H_{0,\,n}$ that information for classification has the Low level against alternatives $H_{1,\,n}$, where n means that the analysis is provided in space $\prod_{i=1}^{n} X_i$, are considered.

Bans have that property that always in the specified problems of check of hypotheses the critical set can be chosen so that it consists only of the bans of probability measures. It means that the probability of a "false alarm" is equal to zero at all n. If at the same time it is possible to prove that in data presentation language for classification it is impossible to express at least one ban in two different ways, then it is possible to prove that with probability 1 any ban of the High level can be founded on a finite step of search of High level information in classification procedure (Grusho et al., 2015).

Using the theory of bans, a classification of valuable information needs to be carried out on the basis of the list of bans of the probability measure P, which describes results of information technology.

Check of object O on possession of a ban is connected with representation of this object in the form of expression in language with the finite alphabet in which it is possible to reveal ban existence. I.e. it is necessary to define language of the description of objects O and to construct algorithms of identification of bans if objects O are sent to the input of the algorithm.

The main problem consists in whether it is possible to express in this language valuable information so that the algorithm could not detect presence of valuable information. That is whether a bypass of rules of reference of an object to valuable information is possible.

In the constructed model it is clear what is a compromise. If a ban is not found in object O, but there is no proof of a possibility of bypassing High level information, then object O is classified as Low in supposition that an adversary has also no algorithms of calculation of valuable information at O.

MODEL OF AMBIGUITY FORMATION TO PREVENT DETECTION OF VALUABLE INFORMATION

In this model it is supposed that object O is the result of work of information technology, does not possess bans, but valuable information is used at calculation of O. The danger consists that an adversary can solve the inverse problem of recovery of valuable information under O and information available to him on transformation which is implemented in information technology.

For protection of valuable information in this case it is possible to use the fact that at the solution of the inverse task, as a rule, valuable information cannot be unambiguously restored. Trying to obtain high degree of ambiguity of the solution of the inverse task it is possible to reach the necessary level of security of valuable information.

In a general consideration it is difficult to formalize this approach. Therefore we will illustrate this method of information security on the following example.

Example 3.

Let's assume that valuable information is a binary vector $x = (x_1, ..., x_n)$. Transformation which is implemented by the information technology is known to the adversary and is described by the linear equations system.

$$\begin{cases} a_{11}x_1 + \cdots + a_{1n}x_n = y_1; \\ \dots\dots\dots\dots\dots \\ a_{m1}x_1 + \cdots + a_{mn}x_n = y_m, \end{cases} \qquad (1)$$

where $a_{ij} \in \{0; 1\}$, all calculations are conducted in GF(2), and output result of an information technology $O = \overline{y} = (y_1, \dots y_m)$. Then at $m < n$ and the rank of the transformation matrix equals to m, for every $\overline{y}$ there exist exactly 2^{n-m} solutions of system (1).

Thus, ambiguity of definition of valuable information is exactly calculated. Let's note that if valuable information $\overline{x}$ participates in k independent implementations of the information technology, then we receive k of linear equations systems of form (1) with different matrixes of coefficients. At the same time when $km \geq n$, there is a real opportunity for an adversary to recover valuable information.

In the considered example an ambiguity of recovery of valuable information can be increased, entering a false or dependent variables describing valuable information.

However such need disappears when instead of a linear system of equations (1) there is a nonlinear system of equations of high degree of nonlinearity. Except a number of special cases such systems can be solved only by method of brute force search of possible solutions. The brute force algorithm creates an additional boundary of protection of valuable information in the form of complexity of an algorithm of inverse of results of the information technology.

Example 4.
One of ways of creation of ambiguity of transformation inverse for the purpose of protection of valuable information is a hiding of the transformation.

Let's assume that $F(\overline{x}) = \overline{y}$ is the transformation of valuable information by means of an information technology into information $\overline{y}$ available to the adversary. If it is possible to hide transformation F, then the adversary for recovery of valuable information should recover transformation F. One of ways of hiding of transformation F is placement it in the form of the program on programmable FPGA microprocessor.

Let's note that such problems cannot be solved with the help of the brute force method since there is no criterion of correctness of recovery the program and valuable information. Besides, even the system in Example 3 if it becomes known to the adversary, then the linear system (1) turns into a nonlinear system, and demands brute force methods for its solution.

Example 5.
Method of hiding of valuable information by means of creation of ambiguity is often used in simple tasks.

Let's assume that it is necessary to hide authorization of the user in a banking system. The user uses two-factor authentication (Grusho et al., 2017), namely, the PIN code and the device creating authentication information of considerable complexity. The PIN code is stored in the user memory but an adversary can steal or copy the device. Then authentication in bank on the basis of the device can be compromised, and only the PIN code at its careful usage provides ambiguity of authentication, and can prevent violation of information security due to ambiguity of data of authentication.

The listed above three classes of problems of protection of valuable information in public information space do not exhaust a set of problems in this area. For example, in the paper (Abbas et al., 2018) the problem of ensuring deductive security of requests to databases is considered.

CONCLUSION

In the paper the problem of protection of the valuable information which is in public access of information space is considered. The need for such protection can arise when forming input data for analysts out of public information resources.

In case the valuable information is sufficient well separated from the analyzed data, the specified formation of input data can be made sufficient protected as it is described in Example 1.

If the directions of formation of data for analysts are a valuable information, then reliability of the method in Example 2 raises doubts.

The method of introduction of additional false information demands further development. For example, the perspective way is an addition of various false purposes in various periods of time. At the same time the true purposes need to be included in a set of false purposes. However estimates of security of valuable information at such approach demand creation of special mathematical models.

At the splitting of valuable information and the conclusion of its part in public information space there arises a complex mathematical task of assessment of security of such method. Namely, reliability of information security is defined by a possibility to define at least of one ban in data presentation language for classification of information. Though this approach is represented as a complex problem generally, at rather poor languages of data presentation for classification this problem can effectively be solved.

In daily usage of this method data presentation language for classification is not constructed at all and it is possible to consider this fact as a compromise in the problem of splitting of valuable information.

The method of introduction an ambiguity of the problem of definition of valuable information allows to estimate sometimes sufficient strictly a security by means of estimation of ambiguity. However, in a general view of formalization of this approach can be based only in specific conditions. It does not allow to build universal estimates of security when using this method.

Acknowledgements
This work was partially supported by the Russian Foundation for Basic Research (grant No. 18-29-03081).

REFERENCES

TCSEC. Department of Defense Trusted Computer System Evaluation Criteria. 1985, DoD.

Anshakov, O.M. (Eds). 2009. *JSM-method of automatic hypothesis generation: logical and epistemological*, Moscow: KD LIBROKOM, 432 p.

Axelsson, S. 1999. The Base-Rate Fallacy and its Implications for the Diculty of Intrusion Detection. *Proc. of the 6th Conference on Computer and Communications Security*, 1–7.

Grusho, A., N. Grusho, and E. Timonina. 2013. "Consistent sequences of tests dened by bans". *Springer Proceedings in Mathematics and Statistics, Optimization Theory, Decision Making, and Operation Research Applications*, 281–291.

Grusho, A., N. Grusho, and E. Timonina. 2015. "Power functions of statistical criteria defied by bans". *SProceedings of 29th European Conference on Modelling and Simulation*, (Varna, Bulgaria, May 26-30). Digitaldruck Pirrot GmbHP, 617–621.

Grusho, A.A., N.A. Grusho, M.I. Zabezhailo, D.V. Smirnov, and E.E. Timonina. 2017. "About complex authentication". *Systems and Means of Informatics*, 27, No. 3, 3–10.

Abbas, M.M., N.P. Varnovskij, V.A. Zakharov and A.V. Shokurov. 2018. "On the deductive security of queries to databases with multi-bit entries". *Bulletin of the Moscow University. Series 15: Calculus Mathematics and Cybernetics*, 1, 40–45.

AUTHOR BIOGRAPHIES

ALEXANDER A. GRUSHO,Professor (1993), Doctor of Science in physics and mathematics (1990). He is principal scientist at Institute of Informatics Problems, Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences and Professor of Moscow State University.

Research interests: probability theory and mathematical statistics, information security, discrete mathematics, computer sciences.

His email is grusho@yandex.ru.

NICK A. GRUSHO has graduated from the Moscow Technical University. He is Candidate of Science (PhD) in physics and mathematics. At present he works as senior scientist at Institute of Informatics Problems, Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences (FRC CSC RAS).

Research interests: probability theory and mathematical statistics, information security, simulation theory and practice, computer sciences.

His email is info@itake.ru.

MICHAEL I. ZABEZHAILO has graduated from the Institute of Physics and Technology and gained the Candidate degree (PhD) in theoretical computer science (1983). He is Doctor of Science in physics and mathematics (2016). Now he works as Head of laboratory in Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences.

Research interests: mathematical foundations of artificial intelligence, reasoning modeling, information security, theoretical computer sciences.

His email is: zabezhailo@yandex.ru.

ELENA E. TIMONINA has graduated from the Moscow Institute of Electronics and Mathematics and obtained the Candidate degree (PhD) in physics and mathematics (1974). She is Doctor in Technical Science (2005), Professor (2007). Now she works as leading scientist in Institute of Informatics Problems, Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences (FRC CSC RAS).

Research interests: probability theory and mathematical statistics, information security, cryptography, computer sciences.

Her email is eltimon@yandex.ru.