Yurii DREIS[1], Iryna LOZOVA[2], Andrii BISKUPSKYI[3],
Lydia KUZMENKO[4], Ali .T. AL-KHWALDEH[5]

Scientific supervisor: Alexander KORCHENKO[6]

A TUPLE MODEL FOR ESTIMATING THE CONSEQUENCES OF PERSONAL DATA LEAKAGE IN AUTOMATED SYSTEMS

Abstract: A tuple model for estimating the consequences of leakage of personal data in automated systems has been developed, which by determining the sets of characteristics of personal data, generally known documents containing personal data, personal data processing environments, personal data processing goals, plurality of personal data protection criteria, identification of personal data security threats, the magnitude of the possible consequences of loss of personal data, risk assessment of personal data protection and risk management to achieve the necessary level of protection of personal data provides the possibility to develop a method for assessing the negative consequences of leakage of personal data during their processing in the state automated systems in order to provide the required level of security at acceptable costs and a given level of limitations.

Keywords: personal data, tuple model, consequences of leakage of personal data, personal data protection.

KROTKOWY MODEL OCENIANIA SKUTKÓW WYCIEKU DANYCH OSOBOWYCH W SYSTEMACH ZAUTOMATYZOWANYCH

Streszczenie: Opracowano krotkowy model oceniania skutków wycieku danych osobowych w systemach zautomatyzowanych. Niektóre systemy polegają na wyznaczeniu mnóstwa charakterystyk danych osobowych na podstawie, powszechnie znanych dokumentów zawierających dane osobowe. Proponowany system chroni i zabezpiecza dane dla różnych

[1] PhD Eng (Information security), Head of Innovative Technologies Professional Education Academic Department, National Aviation University, Dreisyuri@gmail.com
[2] Senior lector of IT-Security Academic Department, National Aviation University illozovaya@gmail.com
[3] Assistant Professor of IT-Security Academic Department, National Aviation University andrii.biskupskyi@gmail.com
[4] Methodologist of II category, Center for advanced planning and monitoring of educational activities, National Academy of Security Service of Ukraine bvl-home@ua.fm
[5] Assistant Professor, Philadelphia University akhwaldeh@philadelphia.edu.jo
[6] Dr Eng (Information security), Professor, Laureate of the State Prize of Ukraine in Science and Technology, Head of IT-Security Academic Department, National Aviation University, Visit-Professor at The University of Bielsko-Biala (Akademia Techniczno-Hu-manistyczna, Bielsko-Biała, Poland), Leading Researcher of the National Academy of SS of Ukraine, icaocentre@nau.edu.ua

środowisk obróbki danych osobowych, różnych celów obróbki danych osobowych. Uwzględnia wiele kryteriów ochrony danych osobowych, identyfikuje zagrożenia bezpieczeństwa danych osobowych, wielkości ewentualnych skutków utraty danych osobowych, ocenia ryzyka ochrony danych osobowych i wspomaga kierowanie ryzykiem dla osiągnięcia koniecznego poziomu ochrony danych osobowych. System daje możliwość opracowania metody oceniania negatywnych skutków wycieku danych osobowych podczas ich obróbki w państwowych automatyzowanych systemach dla zabezpieczenia koniecznego poziomu osłonienia przy dopuszczalnym wydatkowaniu i zadanym poziomie ograniczeń.

Słowa kluczowe: dane osobowe, model krotkowy, skutki wycieku danych osobowych, ochrona danych osobowych

1. Introduction

Distribution of data leakage leads to the fact that countries around the world are engaged in reforming the policy and regulation of personal data (PD) protection. One of the most prominent examples is the GDPR (General data protection regulation) of the European Union, which came into force in May 2018.
The need to protect PD is increasing along with the digital transformation of sectors such as health and financial services. More and more organisations are engaged in PD processing and deal with an increase in their volume. In this regard, the protection of the PD and its enhancement is not only a duty of a State and the subject of a State regulation, but should be considered inseparable in conjunction with the protection of human rights and freedoms, including the protection of the right to respect for private life.

2. Development of a model for estimating the negative consequences of personal data leakage in automated systems

Proceeding from the current legislative requirements, a model is proposed for assessing the negative consequences of leakage of PD during their processing in the automated systems (AS).
The model is designed in the form of a tuple:

$$\mathbf{IDF} =< \mathbf{IDF_1}, \mathbf{IDF_2}, ..., \mathbf{IDF_i}, ..., \mathbf{IDF_n} >, \quad (1)$$

where $\mathbf{IDF_i} \subseteq \mathbf{IDF}$ $(i = \overline{1,n})$ – a tuple component which reflects the *i*-th identifier of the object, *n* their number, and for all elements **IDF** order property is characteristic.
For example, for $n = 11$ we define the tuple (1) as:

$$\mathbf{IDF} = \mathbf{< IDF_1, IDF_2, ..., IDF_7, ..., IDF_{11} >} =$$
$$< \mathbf{C, DO, PR, P, A, CR, TH, RE, L, R, MA} >,$$

where $\mathbf{IDF_1} = \mathbf{C}$ (Characteristic) PD (identification of their composition and content); $\mathbf{IDF_2} = \mathbf{DO}$ (generally known (Documents), which contains PD); $\mathbf{IDF_3} = \mathbf{PR}$ (Processing) environment PD); $\mathbf{IDF_4} = \mathbf{P}$ (Purpose) of PD processing); $\mathbf{IDF_5} = \mathbf{A}$

(Audit) of applied security mechanisms); $\mathbf{IDF_6 = CR}$ (set of (Criteria) of the protection of the PD, which are processed in the AS); $\mathbf{IDF_7 = TH}$ (identification of the security (Threats) PD when procesing the DPD in the AS); $\mathbf{IDF_8 = RE}$ (possible occurrence of (Responsibility) of the owners or managers of the databases of the PD); $\mathbf{IDF_9 = L}$ (the magnitude of the possible consequences of the loss of the PD (Losses); $\mathbf{IDF_{10} = R}$ (risk assessment (Risk) of the security of the PD in the AS); $\mathbf{IDF_{11} = MA}$ (risk management (Management) and achievement of the required level of the PD security in the AS).

The first component of the tuple **C** – is the set of identifiers of the characteristics of the PD (identification of their composition and content) is reflected as:

$$\mathbf{C} = \{\bigcup_{i=1}^{n_1} C_i\} = \{C_1, C_2, ..., C_{n_1}\}, \quad (2)$$

where $C_i \subseteq \mathbf{C}$ $(i=\overline{1,n_1})$ – the *i*-th identifier of the characteristics of the PD and n_1 is their number.

It is defined by the Legislation [1] that the PD is information or set of information about an individual that is identified or can be specifically identified. In accordance with [1-4] we will define generally known PDs.

For example, when $n_1 = 25$ $(i = \overline{1,25})$ formula (2) takes the form:

$$\mathbf{C} = \{\bigcup_{i=1}^{25} C_i\} = \{C_1, C_2, ..., C_{24}, C_{25}\},$$

where C_1 = «Names (name, patronymic, surname) of a person»; …; C_{25} = «Other data (issued on the name of the person, etc.)».

The second component **DO** – general known documents, which contain the PD, is determined by the expression:

$$\mathbf{DO} = \{\bigcup_{i=1}^{n_2} DO_i\} = \{DO_1, DO_2, ..., DO_{n_2}\}, \quad (3)$$

where $DO_i \subseteq \mathbf{DO}$ $(i=\overline{1,n_2})$ – *i*-th is a title of the document, which contains the PD and n_2 is their number.

For example, for $n_2 = 18$ $(i = \overline{1,18})$ formula (3) can be represented as:

$$\mathbf{DO} = \{\bigcup_{i=1}^{18} DO_i\} = \{DO_1, DO_2, ..., DO_{17}, DO_{18}\},$$

where DO_1 = «Passport of a citizen of Ukraine»; …; DO_{18} = «Documents on personnel in the organisation (orders regarding personnel, extracts from them, employment contracts, application forms)».

Under the third component of the tuple PR – the processing environment of the PD, it shall be understood software, which carried out any actions associated with the

introduction, modification, destruction of the PD in the database of personal data (DPD). Defined by the expression:

$$\mathbf{PR} = \{\bigcup_{i=1}^{n_3} PR_i\} = \{PR_1, PR_2, ..., PR_{n_3}\}, \quad (4)$$

where $PR_i \subseteq \mathbf{PR}$ $(i = \overline{1, n_3})$ – i-th is the name of the software processing the PD, and n_3 is their number.

For example, for $n_3 = 14$ $(i = \overline{1,14})$ formula (4) can be presented as:

$$\mathbf{PR} = \{\bigcup_{i=1}^{14} PR_i\} = \{PR_1, PR_2, ..., PR_{13}, PR_{14}\},$$

where according to [5] PR_1 = «MS Office»; …; PR_{14} = «EDMS «SX-Government».

The next component **P** – the purpose of processing the PD. The goal of the PD processing is determined depending on where and for which expected final result such PDs will be processed. This parameter is represented as a set of identifiers:

$$\mathbf{P} = \{\bigcup_{i=1}^{n_4} P_i\} = \{P_1, P_2, ..., P_{n_4}\}, \quad (5)$$

where $P_i \subseteq \mathbf{P}$ $(i = \overline{1, n_4})$ – i-th is a subset of the goal of the PD processing, and the n_4 is their number.

For example, taking into account [6] for $n_4 = 13$ $(i = \overline{1,13})$ formula (5) takes the form:

$$\mathbf{P} = \{\bigcup_{i=1}^{13} P_i\} = \{P_1, P_2, ..., P_{12}, P_{13}\},$$

where P_1 = «Ensuring labour relations»; …; P_{13} = «Ensuring other relationships that require processing of personal data».

The next component **A** is an audit of the applied security mechanisms. This component is described by a set of elements and is determined by checking the organisational and legal and program-technical measures and mechanisms of protection of the PDs processed in the AS:

$$\mathbf{A} = \{\bigcup_{i=1}^{n_5} A_i\} = \{A_1, A_2, ..., A_{n_5}\}, \quad (6)$$

where $A_i \subseteq \mathbf{A}$ $(i = \overline{1, n_5})$ – the i-th security mechanism, and n_5 number of mechanisms applied.

For example, taking into account [7] for $n_5 = 6$ $(i = \overline{1,6})$ formula (6) takes form:

$$\mathbf{A} = \{\bigcup_{i=1}^{6} A_i\} = \{A_1, A_2, A_3, A_4, A_5, A_6\},$$

where A_1 = «The consent of the person of the PD for their processing in the AS»; …; A_6 = «Other mechanisms of protection».

The next component **CR** – the set of protection criteria for PD that are processed in the AS is formed as:

$$\mathbf{CR} = \{\bigcup_{i=1}^{n_6} CR_i\} = \{CR_1, CR_2, ..., CR_{n_6}\},$$

where $CR_i \subseteq \mathbf{CR}$ $(i = \overline{1, n_6})$ – *i*-th is a subset of security criteria for PD, and n_6 their number.

For example, for $n_6 = 5$ the subset CR_i takes form [8]:

$$\mathbf{CR} = \{\bigcup_{i=1}^{5} CR_i\} = \{CR_1, CR_2, CR_3, CR_4, CR_5\},$$

where CR_1 = «Privacy Criteria»; …; CR_5 = «Guarantee criteria».

The seventh component is **TH** – the identification of the security threats to the PD when processing the DPD in the AS. The list of possible threats is determined by research [10] and can be supplemented independently. This parameter is presented as a set of elements of known security threats to the PD, namely:

$$\mathbf{TH} = \{\bigcup_{i=1}^{n_7} \mathbf{TH}_i\} = \{\mathbf{TH}_1, \mathbf{TH}_2, ..., \mathbf{TH}_{n_7}\}, \quad (7)$$

where $\mathbf{TH}_i \subseteq \mathbf{TH}$ $(i = \overline{1, n_7})$ – *i*-th is a subset of security threats to PD, and n_7 their number.

For example, for $n_7 = 2$ a subset $\mathbf{TH}_i$ takes form [9]:

$$\mathbf{TH} = \{\bigcup_{i=1}^{2} \mathbf{TH}_i\} = \{\mathbf{TH}_1, \mathbf{TH}_2\},$$

where $\mathbf{TH}_1$ = «The threats of subjective nature»; $\mathbf{TH}_2$ = «The threats of objective nature».

A subset $\mathbf{TH_i}$ define as:

$$\mathbf{TH_i} = \{\bigcup_{j=1}^{n_{7i}} TH_{ij}\} = \{TH_{i1}, TH_{i2}, ..., TH_{in_{7i}}\}, \quad (8)$$

where $TH_{ij} \subseteq \mathbf{TH_i}$ $(j = \overline{1, n_{7i}})$ – the *j*-th subset of groups of threats related to a certain topic or close to certain characteristics within the boundaries of *i*-th subset, and n_{7i} number of groups of the *i*-th subset.

Taking into account (8) the expression (7) can be represented in the following form:

$$\begin{aligned}\mathbf{TH} = \{\bigcup_{i=1}^{n_7} \mathbf{TH_i}\} = \{\bigcup_{i=1}^{n_7}\{\bigcup_{j=1}^{n_{7i}} TH_{ij}\}\} = \{\{TH_{11}, TH_{12}, ..., TH_{1n_{71}}\},\\ \{TH_{21}, TH_{22}, TH_{23}, ..., TH_{2n_{72}}\}, ...,\\ \{TH_{n_7 1}, TH_{n_7 2}, TH_{n_7 3}, ..., TH_{n_7 n_{7n_7}}\},\end{aligned} \tag{9}$$

For example, for $n_7 = 2$ $(i = \overline{1,2})$, $n_{71} = 42$ $(i = \overline{1,42})$, $n_{72} = 5$ $(i = \overline{1,5})$, taking into account [9] formula (9) takes form:

$$\mathbf{TH} = \{\bigcup_{i=1}^{2}\{\bigcup_{j=1}^{n_{7i}} TH_{ij}\}\} = \{\{TH_{11}, TH_{12}, ..., TH_{142}\}, \{TH_{21}, TH_{22}, ..., TH_{25}\}\},$$

where TH_{11} = «Distribution of distorted, inaccurate and biased information in the information space»; …; TH_{25} = «Software halts and failures».

The eighth parameter tuple **RE** – possible occurrence of responsibility of owners or managers of the database of personal data. This parameter is represented as a set of identifiers:

$$\mathbf{RE} = \{\bigcup_{i=1}^{n_8} \mathbf{RE}_i\} = \{\mathbf{RE}_1, \mathbf{RE}_2, ..., \mathbf{RE}_{n_8}\}, \tag{10}$$

where $\mathbf{RE}_i \subseteq \mathbf{RE}$ $(i=\overline{1,n_8})$ – a subset of responsibility criteria, and n_8 their number.

For example, for $n_8 = 10$ a subset $\mathbf{RE}_i$ takes form [10]:

$$\mathbf{RE} = \{\bigcup_{i=1}^{10} \mathbf{RE}_i\} = \{\mathbf{RE}_1, \mathbf{RE}_2, ..., \mathbf{RE}_9, \mathbf{RE}_{10}\},$$

where $\mathbf{RE}_1$ = «The commission of actions related to personal data without the consent of the subject of personal data»; …; $\mathbf{RE}_{10}$ = «Responsibility for breaching the requirements for the dissemination (distribution, sale, transfer) of information about an individual, taking into account that the implementation of the requirements of the established mode of protection of personal data is provided by the party that disseminates this data».

A subset $\mathbf{RE}_i$ define as:

$$\mathbf{RE}_i = \{\bigcup_{j=1}^{n_{8i}} RE_{ij}\} = \{RE_{i1}, RE_{i2}, ..., RE_{in_{8i}}\}, \tag{11}$$

where $RE_{ij} \subseteq \mathbf{RE}_i$ $(j = \overline{1, n_{8i}})$ – the j-th subset of groups of attributes of responsibility related to a certain topic or close to certain characteristics within the boundaries of i-th subset, and n_{8i} number of groups of the i-th subset.

Taking into account (11) the expression (10) can be represented in the following form:

$$\mathbf{RE} = \{\bigcup_{i=1}^{n_8} \mathbf{RE}_i\} = \{\bigcup_{i=1}^{n_8} \{\bigcup_{j=1}^{n_{8i}} RE_{ij}\}\} = \{\{RE_{11}, RE_{12}, ..., RE_{1n_{81}}\},$$
$$\{RE_{21}, RE_{22}, RE_{23}, ..., RE_{2n_{82}}\}, ..., \{RE_{n_8 1}, RE_{n_8 2}, RE_{n_8 3}, ..., RE_{n_8 n_{8n_8}}\}, \quad (12)$$

For example, for $n_8 = 10$ $(i = \overline{1,10})$, $n_{81} = 5$ $(i = \overline{1,5})$, $n_{82} = 6$ $(i = \overline{1,6})$, $n_{83} = 2$ $(i = \overline{1,2})$, $n_{84} = 4$ $(i = \overline{1,4})$, $n_{85} = 3$ $(i = \overline{1,3})$, $n_{86} = 2$ $(i = \overline{1,2})$, $n_{87} = 4$ $(i = \overline{1,4})$, $n_{88} = 9$ $(i = \overline{1,9})$, $n_{89} = 5$ $(i = \overline{1,5})$, $n_{810} = 6$ $(i = \overline{1,6})$, taking into account [10] formula (12) takes form:

$$\mathbf{RE} = \{\bigcup_{i=1}^{10} \{\bigcup_{j=1}^{n_{8i}} RE_{ij}\}\} = \{\{RE_{11}, RE_{12}, ..., RE_{15}\}, \{RE_{21}, RE_{22}, ..., RE_{26}\},$$
$$\{RE_{21}, RE_{22}, ..., RE_{25}\}, \{RE_{31}, RE_{32}\}, \{RE_{41}, RE_{42}, ..., RE_{44}\},$$
$$\{RE_{51}, RE_{52}, RE_{53}\}, \{RE_{61}, RE_{62}\}, \{RE_{71}, RE_{72}, ..., RE_{74}\},$$
$$\{RE_{81}, RE_{82}, ..., RE_{89}\}, \{RE_{91}, RE_{92}, ..., RE_{95}\}, \{RE_{101}, RE_{102}, ..., RE_{106}\}\},$$

where RE_{11} = «Processing of personal data without specific and legitimate purposes, determined by agreement of the subject of personal data»; …; RE_{106} = «Use the information about the private life of an individual as a factor that confirms or denies his business qualities».

The ninth parameter of the tuple **L** - the magnitude of the possible consequences of loss of PD, can be defined both quantitative and qualitative indicators. The assessment of the consequences and losses is carried out on the scale of known methods [11], characterizing material or moral damage, damage from violation of human rights and freedoms, etc.

Component **L** is formed as:

$$\mathbf{L} = \{\bigcup_{i=1}^{n_9} L_i\} = \{L_1, L_2, ..., L_{n_9}\},$$

where $L_i \subseteq \mathbf{L}$ $(i = \overline{1, n_9})$ – a subset of the criteria of the consequences, and n_9 their number.

For example, for $n_9 = 8$ a subset L_1 takes form [11]:

$$\mathbf{L} = \{\bigcup_{i=1}^{8} L_i\} = \{L_1, L_2, ..., L_7, L_8\},$$

where L_1 = «Losses of reputation of the organisation»; …; L_8 = «Disorganisation of activity».

The tenth parameter of the tuple **R** – security risk assessment of PD in AS. Under the security risk for the PD during their processing in the AS is deemed the function of the probability of realising the threat to the type and magnitude of the damage caused by the possible loss of the PD in the presence of vulnerabilities and the degree of their acceptability for the operation of the AS.

This parameter is represented as a set of identifiers:

$$\mathbf{R} = \{\bigcup_{i=1}^{n_{10}} R_i\} = \{R_1, R_2, ..., R_{n_{10}}\},$$

where $R_i \subseteq \mathbf{R}$ $(i=\overline{1,n_{10}})$ – the *i*-th subset of the security risk to PD, and n_{10} their number.

For example, for $n_{10} = 5$ [11] a subset R_i takes form:

$$\mathbf{R} = \{\bigcup_{i=1}^{5} R_i\} = \{R_1, R_2, R_3, R_4, R_5\},$$

where R_1 = «Very low»; …; R_5 = «Very high».

The final step is to determine the last parameter of the tuple **MA** – risk management and achievement the required level the PD security in the management system. At this stage, using the obtained values of the parameters of the probability of realisation of the threat and the magnitude of the losses incurred, the required level of PD security in the AS for the application of the recommended security policy is determined.

$$\mathbf{MA} = \{\bigcup_{i=1}^{n_{11}} MA_i\} = \{MA_1, MA_2, ..., MA_{n_{11}}\},$$

where $MA_i \subseteq \mathbf{MA}$ $(i=\overline{1,n_{11}})$ – the *i*-th subset of the recommended security policy, and n_{11} their number.

For example, for $n_{11} = 5$ [12] a subset MA_i takes form:

$$\mathbf{MA} = \{\bigcup_{i=1}^{5} MA_i\} = \{MA_1, MA_2, MA_3, MA_4, MA_5\},$$

where MA_1 = «Implementation of organisational and legal measures for the protection of information»; …; MA_5 = «Application of integrated information security system».

The recommended security policy is based on the analysis of the requirements of Ukrainian legislation in the field of protection of PD. It is envisaged the revision of the set of measures for risk assessment, selection, realisation and implementation of measures (mechanisms) for protecting the PD in the DPD in the AS, which conducting during the entire life cycle of the AS and aimed at achieving an acceptable level of residual risk. At the heart of the recommended security policy, it is envisaged to apply the necessary measures and means through their implementation in the AS to increase the level of protection of the PD.

CONCLUSIONS

A tuple model for estimating the negative consequences of the PD leak in the AS was developed in the research, based on the provisions of international standards and the current requirements of the Ukrainian legislation, which will enable the possibility to develop a method for assessing the negative consequences of leakage of PD during their processing in the state AS to provide the required level of security at acceptable costs and a given level of constraints.

REFERENCE

1. Verkhovna Rada of Ukraine: On personal data protection. Law of 01.06.2010 № 2297-VI: *http://zakon.rada.gov.ua/laws/show/2297-17*
2. Verkhovna Rada of Ukraine: On access to court decisions. Law of 22.12.2005 № 3262-IV: *http://zakon5.rada.gov.ua/laws/show/3262-15*
3. Verkhovna Rada of Ukraine: Civil Code of Ukraine. Law of 16.01.2003 № 435-IV: *http://zakon.rada.gov.ua/laws/show/435-15/page21*
4. Cabinet of Ministers of Ukraine: On Approval of the Regulation on the State Register of personal data bases and the procedure for its administration. Law of 25.05.2011 № 616: http://zakon5.rada.gov.ua/laws/show/616-2011-%D0%BF
5. State Service for Special Communications and Information Security of Ukraine: Information on the means of technical information security that have certificates of conformity and expert conclusions: *http://www.dsszzi.gov.ua/dsszzi/control/uk/publish/article?art_id=234241&cat_id=39181*
6. Recommendations on the procedure for processing personal data in the databases of personal data: *https://www.kadrovik.ua/content/rekomendats-shchodo-poryadku-obrobki-personalnikh-danikh-u-bazakh-personalnikh-danikh*
7. KORCHENKO O., DREIS Y., LOZOVA I. Model and method of risk assessment of personal data security during their processing in automated systems. 2016, 39 - 48.
8. Department of Special Telecommunication Systems and Information Security of the Security Service of Ukraine: Criteria for assessing the security of information in computer systems from unauthorised access. 28.04.1999: *http://www.archives.gov.ua/Archives/Info/ND_TZI_2.5-004-99.pdf*
9. YUDIN O., BUCHIK S. State information resources. Methodology of constructing a threat classifier: a monograph. Kyiv, 2015, p. 214 .
10. MERVINSKYI O., NIKOLAEV A.: Responsibility for violation of legislation requirements in the field of protection of personal data. Kyiv, 2011, 5 – 10.
11. KORCHENKO O., KAZMIRCHUK S., AHMETOV B. Application systems for assessing information security risks: monograph. Kyiv, 2017, p. 435.

12. KORCHENKO O., ARHIPOV O., DREIS Y. Оцінювання шкоди національній безпеці України у разі витоку державної таємниці: монографія. Kyiv, 2014. p. 332.