ISSN 2300-5149

PRZEGLĄD TELEINFORMATYCZNY

TOM 7 (25) **Nr 3-4 (49)**
WARSZAWA **2019**

PRZEGLĄD TELEINFORMATYCZNY
TELEINFORMATICS REVIEW

Dawniej: BIULETYN INSTYTUTU AUTOMATYKI I ROBOTYKI WAT
(ISSN 1427-3578)
Ukazuje się od 1995 r.

UMOWA NR 630/P-DUN/2018 Z MNiSW (z dn. 21.06.2018 r.) na lata 2018-2019:
– wzrost liczby artykułów anglojęzycznych;
– obsługa strony internetowej czasopisma na platformie IC, nadawanie artykułom naukowym numerów DOI i wykorzystanie profesjonalnego panelu edycyjnego.
Wymienione zadania finansowane są w ramach umowy nr 630/P-DUN/2018 ze środków Ministra Nauki i Szkolnictwa Wyższego przeznaczonych na działalność upowszechniającą naukę.

ADRES REDAKCJI
Redakcja Przeglądu Teleinformatycznego
00-908 Warszawa, ul. gen. Sylwestra Kaliskiego 2
tel. 261 83 95 52, fax. 261 83 71 44
e-mail: pt [at] ita.wat.edu.pl
WWW: **http://przeglad.ita.wat.edu.pl/**
https://przegladteleinformatyczny.publisherspanel.com/

Wersją pierwotną czasopisma jest wersja elektroniczna

WYDAWCA: Instytut Teleinformatyki i Cyberbezpieczeństwa WAT

ISSN 2300-5149
ISSN 2353-9836 (on-line)

The regularization method in the classification task according to given examples

Włodzimierz KWIATKOWSKI

Institute of Teleinformatics and Cybersecurity, Faculty of Cybernetics, MUT,
ul. gen. Sylwestra Kaliskiego 2, 00-908 Warsaw, Poland
wlodzimierz.kwiatkowski@wat.edu.pl

ABSTRACT: The article considers the problem of classification based on the given examples of classes. As a feature vector a complete characteristic of object is assumed. The peculiarity of the problem being solved is that the number of examples of the class may be less than the dimension of the feature vector, and most of the coordinates of the feature vector can be correlated. As a consequence, the feature covariance matrix calculated for the cluster of examples may be singular or ill-conditioned. This disenable a direct use of metrics based on this covariance matrix. The article presents a regularization method involving the additional use of statistical properties of the environment.

KEYWORDS: regularization, classification, pattern recognition, exploratory data analysis.

1. Introduction

The methods presented in this article apply to the tasks of classifying objects based on their features in the form of real number vectors. The solution proposed can be used especially when a feature vector is defined as a complete characteristic of objects, rather than previously defined attributes. This usually happens when the classification is based on automatically collected data (for example – measurement results), without selection from the point of view of discriminatory properties. This requires an analysis of vectors of large dimensions and large variety. In this case, mining methods, commonly referred to as exploratory data analysis, show promise for the future.

Methods for determining classification rules examined in this article are based on comparing the distance of clusters composed of the given examples of classes from the feature vector of the analyzed object. The size of an example

cluster is usually small – compared to the dimension of the feature space. This poses a significant problem when the classification is based on the metrics defined separately for individual example clusters.

The classification task on the basis of distances defined separately for each class is presented in paper [7]. This approach refers directly to quadratic discriminant analysis (QDA). Where the feature covariance matrix of example cluster is singular, this approach leads to the concept of using the generalised Mahalanobis distance [12]. This concept is based on the Moore–Penrose pseudo-inverse of covariance matrix. However, solutions based on such approach may turn out to be completely wrong.

The method proposed in this article involves formulating the derivative classification task. This task is formulated for the case when the pattern feature covariance matrices are singular or ill-conditioned (there is a large range between their eigenvalues and their determinants are close to zero). The derivative task is constructed to eliminate the reason for not obtaining an unambiguous solution. This is achieved by supplementing the original task with additional information. In the problem under consideration, this is realized by supplementing the distance function – based on the statistical properties of the example cluster – with a regularization term based on the statistical properties of the environment. The presented approach is interpreted as a method for regularizing the original classification task.

The problem of regularizing classification has been studied in various applications and from various points of view [3], [5], [11] and [14]. The following issues are related to the approach discussed in this article.

Analysis of data from many sources is one of the important problems associated with classification. In most cases, such data cannot be modelled by a common, multidimensional statistical model. Methods based on various models and setting the rules for obtaining a compromise solution are used in this case. Here, we will cite paper [1] as an example, which presents consensus theory-based methods. The use of regularization is one of the conditions for obtaining compromise solutions.

The problem of cooperation with the decision-maker (user) to obtain compromise solutions is a separate topic. The method discussed in this article employs the knowledge (experience) of the decision-maker given by indicating examples of patterns [8], [9]. Similar issues occur in the problems of semi-supervised learning algorithms [6], [15], [16], which combine labelled (marked) and unlabelled data. These algorithms are gaining significant interest and are successfully implemented in practical applications for data mining [13], [14]. In these algorithms, the problem of regularization is also significant, and its solution usually involves the idea of penalization [3].

The issue of classification based on the assessment of distance between clusters in the feature space is presented in papers [2], [7]. An example of using such functions is presented in [4].

2. Formulation of the classification problem based on given examples

The given set of objects is numbered 1 to N. A feature vector expressed in real numbers is known for each object. We use the following designation for object number k:

$$\mathbf{a}_k = \left[a_{1,k}, \quad a_{2,k}, \quad \cdots \quad a_{L,k}\right]^T, \qquad \mathbf{a}_k \in R^L \tag{1}$$

Each coordinate $a_{l,k}$ is a real number and parameter L determines the number of vector coordinates. These vectors form a set:

$$\mathrm{A} = \left\{\mathbf{a}_1, \quad \mathbf{a}_2, \quad \ldots \quad \mathbf{a}_N\right\}, \qquad \mathbf{a}_k \in R^L \tag{2}$$

The feature vectors are compiled as the following matrix:

$$\mathbf{A} = \left[\mathbf{a}_1, \quad \mathbf{a}_2, \quad \ldots \quad \mathbf{a}_N\right], \qquad \mathbf{a}_k \in R^L \tag{3}$$

The feature vectors covariance matrix is determined as follows:

$$\mathbf{R} = \tfrac{1}{N-1}\sum_{k=1}^{N}\left(\mathbf{a}_k - \overline{\mathbf{a}}\right)\left(\mathbf{a}_k - \overline{\mathbf{a}}\right)^T \tag{4}$$

where:

$$\overline{\mathbf{a}} = \tfrac{1}{N}\sum_{k=1}^{N}\mathbf{a}_k \tag{5}$$

It is then assumed that

$$\det(\mathbf{R}) \neq 0 \tag{6}$$

The distance between vectors $\mathbf{x}$, $\mathbf{y}$ of feature space R^L is determined in a way that takes into account the dispersion of coordinates and their mutual correlation. This requirement is met by the Mahalanobis distance [10]. It is set by the formula:

$$\mathrm{d}_{\mathrm{e}}(\mathbf{x},\mathbf{y}) = \sqrt{(\mathbf{x}-\mathbf{y})^T\,\mathbf{R}^{-1}(\mathbf{x}-\mathbf{y})}\,, \qquad \mathbf{x},\mathbf{y} \in R^L \tag{7}$$

Examples constituting the class pattern with index $h \in \{1,2,\ldots,H\}$ (where: H – number of classes) are indicated by providing the relevant set of

indexes W_h. Class pattern with index h is therefore represented by the following set of points (cluster) in the feature space:

$$C(W_h) = \{\mathbf{w}_k \in \mathrm{A} : k \in W_h\} \tag{8}$$

The number of elements of such pattern W_h is marked as $N_h = \|C(W_h)\|$.

Inference about the similarity of feature $\mathbf{x}$ to pattern W_h is based on the distance of point $\mathbf{x}$ from cluster $C(W_h)$. For example, the choice of centroid method to determine the distance between clusters results in:

$$\mathrm{D}_\mathrm{e}(\mathbf{x}, C(W_h)) = \mathrm{d}_\mathrm{e}(\mathbf{x}, \overline{\mathbf{w}}_h) = \sqrt{(\mathbf{x} - \overline{\mathbf{w}}_h)^T \mathbf{R}^{-1} (\mathbf{x} - \overline{\mathbf{w}}_h)} \tag{9}$$

where:

$$\overline{\mathbf{w}}_h = \tfrac{1}{N_h} \sum_{j \in W_h} \mathbf{w}_j \tag{10}$$

The classification based on the metric (9) is called environmental.

Due to the method of determining the covariance matrix $\mathbf{R}$, the use of environmental classification is justified when the features of all patterns are uniform in the following sense: the relevant clusters differ only in expected values (and the corresponding covariance matrices are the same). If the pattern covariance matrices differ, it is recommended to diversify the way the distances are measured according to the covariance matrices of respective patterns [7].

Covariance matrix based on examples of pattern W_h are marked as follows:

$$\mathbf{R}_h = \tfrac{1}{N_h - 1} \sum_{j \in W_h} \left(\mathbf{w}_j - \overline{\mathbf{w}}_h\right)\left(\mathbf{w}_j - \overline{\mathbf{w}}_h\right)^T \tag{11}$$

Distance between vectors $\mathbf{x}$, $\mathbf{y}$ of feature space R^L is matched to the pattern W_h, if it is expressed by the formula [7]:

$$\mathrm{d}_h(\mathbf{x}, \mathbf{y}) = \sqrt{(\mathbf{x} - \mathbf{y})^T \mathbf{R}_h^{-1} (\mathbf{x} - \mathbf{y})}, \qquad \mathbf{x}, \mathbf{y} \in R^L \tag{12}$$

We similarly refer to the distance between feature $\mathbf{x}$ and cluster $C(W_h)$. For example, the distance is specified by the following formula for the centroid method:

$$\mathrm{D}_h(\mathbf{x}, C(W_h)) = \mathrm{d}_h(\mathbf{x}, \overline{\mathbf{w}}_h) = \sqrt{(\mathbf{x} - \overline{\mathbf{w}}_h)^T \mathbf{R}_h^{-1} (\mathbf{x} - \overline{\mathbf{w}}_h)} \tag{13}$$

The classification based on metrics (12) suitably matched to individual patterns is referred to as the classification matched to patterns. The usefulness of such a classification, which means differentiating the method of distance

calculation according to the pattern covariance matrix, is illustrated by the example in Figure 1. The example applies to the division of space R^2 into two classes based on given patterns: W_1 and W_2. Points $C(W_1)$ are shown in the figure as circles, points $C(W_2)$ – as squares. Feature space points closer to points $C(W_1)$ they are marked in a darker colour. In the example, clusters $C(W_1)$ and $C(W_2)$ are not linearly separable and the environmental classification gave poor results. The results of classification matched to the patterns are as expected.

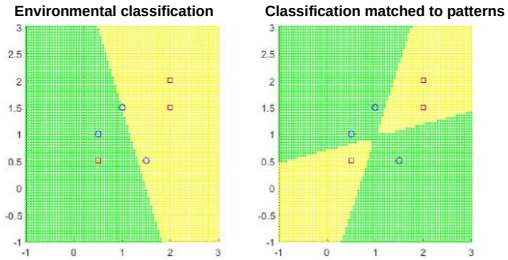


Fig. 1. Example of classification when the pattern covariance matrices are non-singular and the formula (13) is applied. On the left: the distance is determined using a metric based on the covariance matrix calculated for all patterns together, as per the formula (4). On the right: the distance is determined using metrics based on covariance matrices calculated separately for the features of each pattern, as per the formula (11)

3. The method for regularizing the classification task

The problem solved in this article applies to regularizing the task of classification matched to patterns. Regularization is needed when pattern covariance matrices $\mathbf{R}_h$ are singular or ill-conditioned. In the discussed problem, ill-conditioning is understood as a very wide range between the eigenvalues of matrix $\mathbf{R}_h$, causing the matrix determinant to be close to zero.

A routine procedure in the case presented is the application of the generalised Mahalanobis distance, defined as follows [12]:

$$\mathrm{D}_h(\mathbf{x}, C(W_h)) = \mathrm{d}_h(\mathbf{x}, \overline{\mathbf{w}}_h) = \sqrt{(\mathbf{x} - \overline{\mathbf{w}}_h)^T \mathbf{R}_h^+ (\mathbf{x} - \overline{\mathbf{w}}_h)} \tag{14}$$

where: $\mathbf{R}_h^+$ – Moore-Penrose pseudo-inverse of the covariance matrix $\mathbf{R}_h$.
However, in classification tasks based on patterns that are not separable linearly, the application of generalised Mahalanobis distance may lead to false solutions. This is illustrated by the example in Figure 2. As in the example above, space R^2 is divided into two classes based on given patterns: W_1 and W_2. Points

$C(W_1)$ are shown in the figure as circles, points $C(W_2)$ – as squares. Compared to the previous example, both clusters W_1 and W_2 are less numerous: each class is indicated by only two examples. The points of the feature space closer to points $C(W_1)$ are marked in a darker colour. In the example, clusters $C(W_1)$ and $C(W_2)$ are not linearly separable and both classification methods have bad (unexpected) results, and the result of the method using matched metrics is quite the opposite of what was expected.

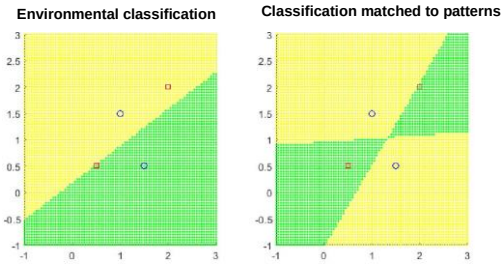


Fig. 2. Example of classification when the pattern covariance matrices are singular and the formula (14) is applied. On the left: the distance is determined using the metrics based on the covariance matrix calculated for all patterns together, as per the formula (4). On the right: the distance is determined using metrics based on covariance matrices calculated separately for each pattern, as per the formula (11)

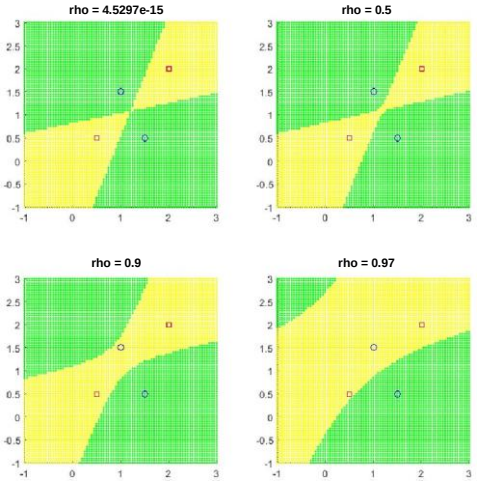


Fig. 3. Illustration of the results of matched classification using regularization

We base the proposed method of regularization on the introduction of a distance function whose values are defined as follows:

$$\mathrm{d}_h^r(\mathbf{x},\mathbf{y}) = \sqrt{(\mathbf{x}-\mathbf{y})^T[(1-\rho)\mathbf{R}_h + \rho\mathbf{R}]^{-1}(\mathbf{x}-\mathbf{y})}\,, \quad \mathbf{x},\mathbf{y} \in R^L \tag{15}$$

where: $\rho \in [0,1]$ – regularization parameter.

Regularization consists in replacing the covariance matrix $\mathbf{R}_h$ with a convex combination of matrix $\mathbf{R}_h$ and matrix $\mathbf{R}$. Value $\rho = 0$ means no regularization and matching classification, while value $\rho = 1$ means transition to the environmental classification.

The results are illustrated for the data as in Figure 2. Figure 3 shows the results of matched classification using regularization. Correct classification results have already been observed for the value approx. 10^{-15} of the regularization parameter. The maximum value of this parameter was approx. 0.5 (a further increase in the parameter causes a smooth transition to the results of the environmental classification).

Figure 4 presents similar results for the task of dividing the feature space into three classes. The comparison included the results of the classification based on Euclidean metric (in the figure marked as Euclidean classification), the classification based on the Mahalanobis metric (in the figure marked as environmental classification), the pattern-matched classification, based on the generalised Mahalanobis metric (marked as matched classification) and pattern-matched classification using regularization with parameter $\rho = 0{,}01$ (marked as regularized classification). We can see that only the results of the last classification gave satisfactory results.

To illustrate the impact of the regularization parameter on the quality of classification, we present the result of a computational experiment consisting in dividing a set of objects into two classes. Features of N first class objects and the same number of second class objects have been randomised in the experiment. Of these, N_1 examples of first class objects and N_2 examples for second-class objects have been indicated at random. For both classes of objects, the features are points on the plane, randomised according to properly selected normal distributions. The following are assumed in the example in Figure 5: $N = 20$, $N_1 = 4$, $N_2 = 2$. It can be seen that the subspace generated by two object indications is a straight line and the matched classification task is ill-conditioned. Regularization of this task at parameter $\rho = 0{,}01$ has allowed us to obtain correct classification results.

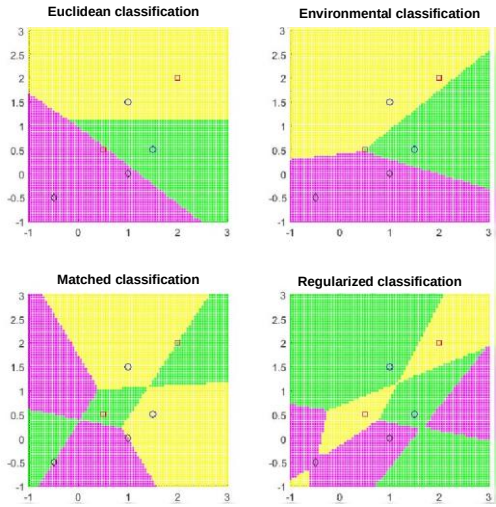


Fig. 4. Comparison of regularized classification results with classification results obtained through other methods

Collective results of the computational experiment under discussion are presented in Figure 6. The abscissa axis indicates the rate of misclassification to class 2, and the ordinate axis – the rate of misclassification to class 1. These rates were determined based on 1000 tests. The green colour indicates the classification results using the matched regularized method for various values of regularization parameter $\rho \in (0,1)$. The end point for the value $\rho = 1$ (marked in blue) corresponds to the quality of the environmental classification. The end point for the value $\rho = 0$ (marked in red) corresponds to the quality of the matched classification without regularization. There is a noticeable lack of continuity of features when transiting from zero value of the regularization parameter ($\rho = 0$) to a positive value. The leap observed in the experiment occurred at the value $\rho \approx 10^{-14}$. This is the minimum value of the regularization parameter for the computing environment being used. The obtained quality curves for regularized classification tend to form a curve illustrating the situation when regularization is not necessary ($N_1 = 4$, $N_2 = 4$). However, also in this case, it is possible to slightly improve the classification quality through regularization. In the experiment discussed, acceptable classification results have been obtained for the regularization parameter value of 0.01 to 0.1.

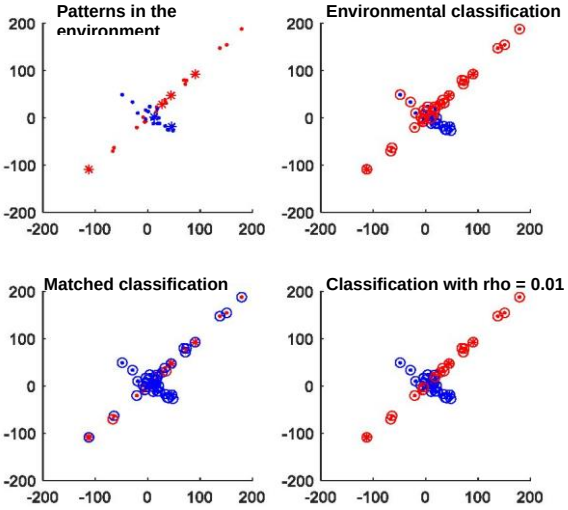


Fig. 5. Example of a computational experiment. Class 1 objects and class 2 objects are marked with red and blue points, respectively. The examples are marked with stars of the relevant colour. Classification results are marked with circles of the relevant colour

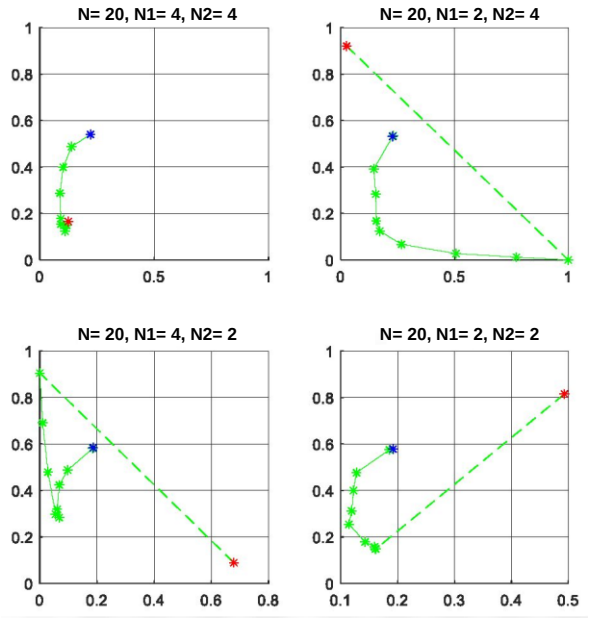


Fig. 6. Illustration of the impact of regularization parameter on the classification quality. The abscissa axis shows the rate of misclassification to class 2, and the ordinate axis – the rate of misclassification to class 1

4. Conclusions

1) The proposed method for regularization can be applied wherever the features of classified objects can be presented as vectors of real numbers. If the covariance matrix of all examined objects is singular (or ill-conditioned), pre-processing should be carried out to select the features that will ensure the non-singularity of their covariance matrix.
2) The interpretation of the derivative task of classification is clear. The proposed approach consists in supplementing the pattern data with data on statistical properties of the environment.
3) The calculation algorithm is attractive because of its simplicity. It allows the use of more complex methods for determining the distance between clusters than method applied in the examples shown in the article. It is also possible to obtain classifications based on different rules for assessing cluster similarity.
4) Classification resulting in trivial, multiple or too extensive classes usually means inconsistencies in the indicated class patterns.

Literature

[1] BENEDIKTSSON J.A., BENEDIKTSSON K., *Hybrid consensus theoretic classification with pruning and regularization*. IEEE 1999 International Geoscience and Remote Sensing Symposium, 1999, Volume 5, pp. 2486-2488.

[2] FUKUNAGA K., *Feature Extraction Algorithm Using Distance Transformation.* IEEE Transactions on Computers C-21(1), February 1972, pp. 56-63.

[3] HSUN-HSIEN CHANG, MOURA JOSE M.F., *Classification by Cheeger Constant Regularization*. 2007 IEEE International Conference on Image Processing, 2007, Volume 2, pp. II-209-II-212.

[4] JIANGTAO PENG, LEFEI ZHANG, LUOQING LI, *Regularized set-to-set distance metric learning for hyperspectral image classification*. Pattern Recognition Letters, Volume 83, Part 2, 1 November 2016, pp. 143-151.

[5] JIM JING-YAN WANG, YI WANG, SHIGUANG ZHAO, XIN GAO, *Maximum mutual information regularized classification*. Engineering Applications of Artificial Intelligence, Volume 37, January 2015, pp. 1-8.

[6] JUN WANG, GUANGJUN YAO, GUOXIAN YU, *Semi-supervised classification by discriminative regularization*. Applied Soft Computing, Volume 58, September 2017, pp. 245-255.

[7] KWIATKOWSKI W., *Metody automatycznego rozpoznawania wzorców*. BEL Studio, Warszawa, 2010.

[8] KWIATKOWSKI W., *Wykrywanie anomalii bazujące na wskazanych przykładach*. Przegląd Teleinformatyczny, nr 1-2, 2018, s. 3-21.

[9] KWIATKOWSKI W., *Recommendations as a result of decision evaluations based on reference examples.* Teleinformatics Review, No. 1-2, 2019, pp. 3-23.

[10] MAHALANOBIS P.C., *On the generalized distance in statistics*. Proceedings of National Institute of Sciences (India), Vol. 2, No. 1, 1936, pp. 49-55.

[11] MAJUMDAR A., WARD R.K., *Classification via group sparsity promoting regularization*. 2009 IEEE International Conference on Acoustics, Speech and Signal Processing, 2009, pp. 861-864.

[12] WARMUS M., *Uogólnienie odległości Mahalanobisa*. Listy Biometryczne, Nr 30-33, 1971, s. 3-7.

[13] TAO ZHANG, CHEN GONG, WENJING JIA, XIAONING SONG, JUN SUN, XIAOJUN WU, *Supervised Image Classification with Self-Paced Regularization*. 2018 IEEE International Conference on Data Mining Workshops (ICDMW), 2018, pp. 411-414.

[14] YANG LI, DAPENG TAO, WEIFENG LIU, YANJIANG WANG, *Co-regularization for classification*. 2014 IEEE International Conference on Security, Pattern Analysis, and Cybernetics (SPAC), 2014, pp. 218-222.

[15] YATING SHEN, YUNYUN WANG, ZHIGUO MAARMUS, *Label-expanded manifold regularization for semi-supervised classification*. 12th International Conference on Intelligent Systems and Knowledge Engineering (ISKE), 2017, pp. 1-4.

[16] ZHILEI CHAI, WEI SONG, HUILING WANG, FEI LIU, *A semi-supervised auto-encoder using label and sparse regularizations for classification*. Applied Soft Computing, Volume 77, April 2019, pp. 205-217.

Metoda regularyzacji w zadaniu klasyfikacji według zadanych przykładów

STRESZCZENIE: W artykule rozpatrywany jest problem klasyfikacji na podstawie wskazanych przykładów klas. Jako wektor cech przyjmuje się kompletną charakterystykę obiektów. Osobliwość rozwiązywanego zadania wynika z tego, że liczba przykładów klasy może być mniejsza od wymiaru wektora cech, a także wektor cech może zawierać współrzędne skorelowane. W konsekwencji macierz kowariancji cech obliczana dla klastra przykładów może być osobliwa albo źle uwarunkowana. Uniemożliwia to bezpośrednie stosowanie metryk bazujących na tej macierzy kowariancji. W artykule została przedstawiona metoda regularyzacji polegająca na dodatkowym wykorzystaniu statystycznych właściwości środowiska.

SŁOWA KLUCZOWE: regularyzacja, klasyfikacja, rozpoznawanie wzorców, eksploracja danych

Received by the editorial staff on: 29.04.2019

PRZEGLĄD TELEINFORMATYCZNY
NR 3-4, 2019

Metoda regularyzacji w zadaniu klasyfikacji według zadanych przykładów

Włodzimierz KWIATKOWSKI

Instytut Teleinformatyki i Cyberbezpieczeństwa, Wydział Cybernetyki, WAT,
ul. gen. Sylwestra Kaliskiego 2, 00-908 Warszawa
wlodzimierz.kwiatkowski@wat.edu.pl

STRESZCZENIE: W artykule rozpatrywany jest problem klasyfikacji na podstawie zadanych przykładów klas. Jako wektor cech przyjmuje się kompletną charakterystykę obiektów. Osobliwość rozwiązywanego zadania wynika z tego, że liczba przykładów klasy może być mniejsza od wymiaru wektora cech, a także większość współrzędnych wektora cech może być skorelowana. W konsekwencji macierz kowariancji cech obliczona dla klastra przykładów może być osobliwa albo źle uwarunkowana. Uniemożliwia to bezpośrednie stosowanie metryk bazujących na tej macierzy kowariancji. W artykule została przedstawiona metoda regularyzacji polegająca na dodatkowym wykorzystaniu statystycznych właściwości środowiska.

SŁOWA KLUCZOWE: regularyzacja, klasyfikacja, rozpoznawanie wzorców, eksploracja danych

1. Wprowadzenie

Opisane w artykule metody dotyczą zadań klasyfikacji obiektów na podstawie ich cech przedstawionych w postaci wektorów liczb rzeczywistych. Proponowane rozwiązanie znajduje zastosowanie zwłaszcza wtedy, gdy jako wektor cech definiuje się kompletną charakterystykę obiektów, a nie wcześniej zdefiniowane atrybuty. Sytuacja taka zwykle występuje wtedy, gdy klasyfikacja jest przeprowadzana na podstawie danych (wyników pomiarów) zbieranych automatycznie, bez ich selekcji z punktu widzenia właściwości dyskryminacyjnych. Powoduje to konieczność analizy wektorów o dużym wymiarze i o dużej rozmaitości. W tym przypadku nadzieje wiąże się z zastosowaniem metod o charakterze wydobywczym, powszechnie określanych jako eksploracja danych.

Rozpatrywane w tym artykule metody wyznaczania reguł klasyfikacji bazują na porównywaniu odległości wektora cech klasyfikowanego obiektu od klastrów złożonych ze wskazanych przykładów klas. Zwykle liczebności przykładowych klastrów są niewielkie – w porównaniu z wymiarem przestrzeni cech. Jest to istotny problem w przypadkach, gdy klasyfikacja oparta jest na wykorzystywaniu metryk definiowanych oddzielnie dla poszczególnych, przykładowych klastrów.

Zadanie klasyfikacji na podstawie odległości definiowanych osobno dla każdej klasy zostało przedstawione w pracy [7]. Podejście to ma bezpośrednie odniesienie do kwadratowej analizy dyskryminacyjnej (QDA, *quadratic discriminant analysis*). Zastosowanie tego podejścia w przypadku osobliwych macierzy kowariancji wzorców prowadzi do koncepcji wykorzystywania uogólnionej odległości Mahalanobisa [12]. Koncepcja ta oparta jest na wykorzystywaniu pseudoinwersji Moore'a–Penrose'a macierzy kowariancji cech. Rozwiązania bazujące na tym podejściu mogą okazać się jednak całkowicie błędne.

Proponowana w niniejszym artykule metoda polega na sformułowaniu pochodnego zadania klasyfikacji. Zadanie to jest formułowane dla przypadku, gdy macierze kowariancji cech wzorców są osobliwe lub źle uwarunkowane (występuje duża rozpiętość między ich wartościami własnymi, a ich wyznaczniki są bliskie zeru). Zadanie pochodne jest konstruowane w celu likwidacji przyczyny nieuzyskiwania jednoznacznego rozwiązania. Uzyskuje się to poprzez uzupełnienie zadania pierwotnego dodatkową informacją. W rozpatrywanym problemie jest to osiągane drogą uzupełniania funkcjonału jakości – bazującego na właściwościach statystycznych klastra przykładów – członem regularyzacyjnym bazującym na statystycznych właściwościach środowiska. Przedstawiane podejście jest interpretowane jako metoda regularyzacji pierwotnego zadania klasyfikacji.

Problem regularyzacji klasyfikacji był badany w różnorodnych zastosowaniach i z różnych punktów widzenia [3, 5, 11, 14]. Następujące zagadnienia związane są z podejściem przedstawianym w niniejszym artykule.

Analiza danych z wielu źródeł należy do ważnych problemów klasyfikacji. Takie dane w większości przypadków nie mogą być modelowane przez wspólny, wielowymiarowy model statystyczny. W tym przypadku stosowane są metody bazujące na różnych modelach i ustalające reguły uzyskiwania rozwiązania kompromisowego. Przytoczymy tu przykładowo pracę [1], w której przedstawiono metody oparte na teorii konsensusu. Zastosowanie regularyzacji jest tu jednym z warunków uzyskania kompromisowych rozwiązań.

Problem współpracy z decydentem (użytkownikiem) w celu uzyskania kompromisowych rozwiązań jest oddzielnym tematem. Proponowana w niniejszym artykule metoda wykorzystuje wiedzę (doświadczenia) decydenta

(użytkownika), bazując na wskazywanych przez niego wzorcowych przykładach [8, 9]. Podobna problematyka występuje w algorytmach uczenia się częściowo nadzorowanych [6, 15, 16], które łączą dane oznaczone (etykietowane) i nieznakowane. Algorytmy te zyskują duże zainteresowanie i są z powodzeniem wdrażane w praktycznych aplikacjach do eksploracji danych [13, 14]. Również w tych algorytmach problem regularyzacji jest istotny, a jego rozwiązanie zwykle bazuje na wykorzystaniu idei penalizacji [3].

Problematyka klasyfikacji bazującej na ocenie odległości między klastrami w przestrzeni cech jest przedstawiona w pracach [2, 7]. Przykład zastosowania takich funkcji odległości zaprezentowano w artykule [4].

2. Sformułowanie problemu klasyfikacji na podstawie zadanych przykładów

Dany jest zbiór obiektów ponumerowany od 1 do N. Dla każdego obiektu znany jest jego wektor cech wyrażonych liczbami rzeczywistymi. Dla obiektu o numerze k stosować będziemy oznaczenie:

$$\mathbf{a}_k = \left[a_{1,k},\quad a_{2,k},\quad \cdots\quad a_{L,k}\right]^T, \qquad \mathbf{a}_k \in R^L \tag{1}$$

Każda współrzędna $a_{l,k}$ jest liczbą rzeczywistą, a parametr L określa liczbę współrzędnych wektora. Wektory te tworzą zbiór:

$$\mathrm{A} = \left\{\mathbf{a}_1,\quad \mathbf{a}_2,\quad \ldots\quad \mathbf{a}_N\right\}, \qquad \mathbf{a}_k \in R^L \tag{2}$$

Wektory cech zestawiamy w postaci następującej macierzy:

$$\mathbf{A} = \left[\mathbf{a}_1,\quad \mathbf{a}_2,\quad \ldots\quad \mathbf{a}_N\right], \qquad \mathbf{a}_k \in R^L \tag{3}$$

Macierz kowariancji wektorów cech wyznaczana jest następująco:

$$\mathbf{R} = \tfrac{1}{N-1}\sum_{k=1}^{N}\left(\mathbf{a}_k - \overline{\mathbf{a}}\right)\left(\mathbf{a}_k - \overline{\mathbf{a}}\right)^T \tag{4}$$

gdzie:

$$\overline{\mathbf{a}} = \tfrac{1}{N}\sum_{k=1}^{N}\mathbf{a}_k \tag{5}$$

Przyjmiemy dalej, że

$$\det(\mathbf{R}) \neq 0 \tag{6}$$

Odległość pomiędzy wektorami $\mathbf{x}$, $\mathbf{y}$ przestrzeni cech R^L będziemy wyznaczać w sposób uwzględniający wielkość rozrzutu (rozproszenia) współrzędnych oraz

ich wzajemną korelację. Wymaganie to spełnia odległość Mahalanobisa [10]. Jest ona określona wzorem:

$$\mathrm{d}_{\mathrm{e}}(\mathbf{x},\mathbf{y}) = \sqrt{(\mathbf{x}-\mathbf{y})^T \mathbf{R}^{-1}(\mathbf{x}-\mathbf{y})}\,, \qquad \mathbf{x},\mathbf{y} \in R^L \tag{7}$$

Wskazania przykładów stanowiących wzorzec klasy o indeksie $h \in \{1,2,\ldots,H\}$ (gdzie: H – liczba klas) będziemy dokonywać przez podanie odpowiedniego zbioru indeksów W_h. Wzorzec klasy o indeksie h jest więc reprezentowany przez następujący zbiór punktów (klaster) w przestrzeni cech:

$$C(W_h) = \{\mathbf{w}_k \in \mathrm{A} : k \in W_h\} \tag{8}$$

Liczbę elementów tak rozumianego wzorca W_h oznaczymy jako $N_h = \|C(W_h)\|$.

Wnioskowanie o podobieństwie cechy $\mathbf{x}$ do wzorca W_h bazuje na określeniu odległości punktu $\mathbf{x}$ od klastra $C(W_h)$. Przykładowo, wybierając metodę centroidalną wyznaczania odległości między klastrami, otrzymujemy zależność:

$$\mathrm{D}_{\mathrm{e}}(\mathbf{x}, C(W_h)) = \mathrm{d}_{\mathrm{e}}(\mathbf{x}, \overline{\mathbf{w}}_h) = \sqrt{(\mathbf{x}-\overline{\mathbf{w}}_h)^T \mathbf{R}^{-1}(\mathbf{x}-\overline{\mathbf{w}}_h)} \tag{9}$$

gdzie:

$$\overline{\mathbf{w}}_h = \tfrac{1}{N_h} \sum_{j \in W_h} \mathbf{w}_j \tag{10}$$

Klasyfikację opartą na wykorzystywaniu metryki (9) nazywać będziemy środowiskową.

Z uwagi na sposób wyznaczenia macierzy kowariancji $\mathbf{R}$, stosowanie klasyfikacji środowiskowej znajduje uzasadnienie wtedy, gdy cechy wszystkich wzorców są jednorodne w następującym sensie: odpowiednie klastry różnią się jedynie wartościami oczekiwanymi (a odpowiadające im macierze kowariancji są jednakowe). W przypadku gdy macierze kowariancji wzorców różnią się, wskazane jest zróżnicowanie sposobu pomiaru odległości stosownie do macierzy kowariancji poszczególnych wzorców [7].

Macierz kowariancji wyznaczoną na podstawie przykładów wzorca W_h oznaczymy następująco:

$$\mathbf{R}_h = \tfrac{1}{N_h - 1} \sum_{j \in W_h} \left(\mathbf{w}_j - \overline{\mathbf{w}}_h\right)\left(\mathbf{w}_j - \overline{\mathbf{w}}_h\right)^T \tag{11}$$

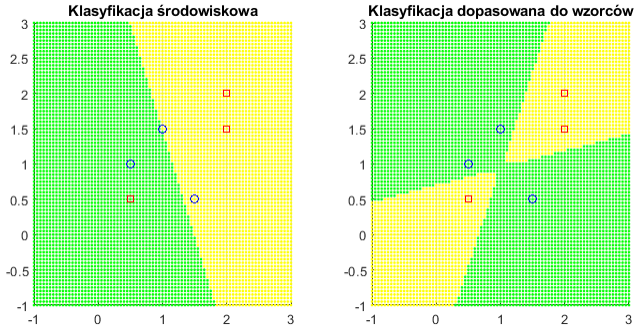
Odległość pomiędzy wektorami $\mathbf{x}$, $\mathbf{y}$ przestrzeni cech R^L zadaną wzorem:

$$\mathrm{d}_h(\mathbf{x},\mathbf{y}) = \sqrt{(\mathbf{x}-\mathbf{y})^T \mathbf{R}_h^{-1}(\mathbf{x}-\mathbf{y})}\,, \mathbf{x},\mathbf{y} \in R^L \tag{12}$$

nazywa się dopasowaną do wzorca W_h [7]. Podobnie nazywać będziemy odległość między cechą $\mathbf{x}$ a klastrem $C(W_h)$. Przykładowo dla metody centroidalnej odległość ta jest określona wzorem:

$$\mathrm{D}_h(\mathbf{x}, C(W_h)) = \mathrm{d}_h(\mathbf{x}, \overline{\mathbf{w}}_h) = \sqrt{(\mathbf{x} - \overline{\mathbf{w}}_h)^T \mathbf{R}_h^{-1} (\mathbf{x} - \overline{\mathbf{w}}_h)} \tag{13}$$

Klasyfikację opartą na wykorzystywaniu metryk (12), odpowiednio dopasowanych do poszczególnych wzorców, nazywać będziemy klasyfikacją dopasowaną do wzorców. Użyteczność stosowania takiej klasyfikacji, a więc różnicowania sposobu obliczania odległości stosownie do macierzy kowariancji wzorców, zilustrujemy przykładem przedstawionym na rysunku 1. Przykład ten dotyczy podziału przestrzeni R^2 na dwie klasy na podstawie zadanych wzorców: W_1 i W_2. Punkty $C(W_1)$ są przedstawione na rysunku jako kółka, punkty $C(W_2)$ – jako kwadraty. Punkty przestrzeni cech leżące bliżej punktów $C(W_1)$ są oznaczone kolorem ciemniejszym. W przedstawionym przykładzie klastry $C(W_1)$ i $C(W_2)$ nie są liniowo separowalne i klasyfikacja środowiskowa dała złe wyniki. Wyniki klasyfikacji dopasowanej do wzorców są zgodne z oczekiwanymi.

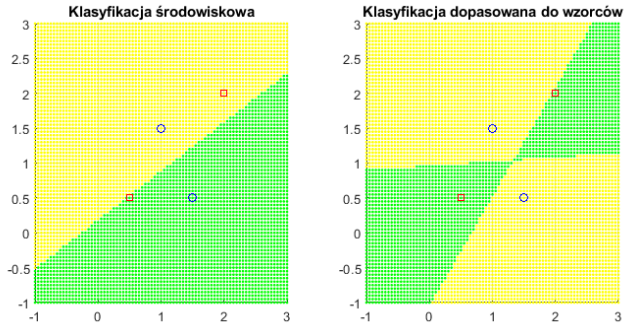


Rys. 1. Przykład klasyfikacji w przypadku, gdy macierze kowariancji wzorców są nieosobliwe i zastosowany jest wzór (13). Z lewej strony: do wyznaczania odległości wykorzystywana jest metryka bazująca na macierzy kowariancji obliczonej dla wszystkich wzorców razem, zgodnie ze wzorem (4). Z prawej strony: do wyznaczania odległości wykorzystywane są metryki bazujące na macierzach kowariancji obliczanych osobno dla cech każdego wzorca, zgodnie ze wzorem (11)

3. Metoda regularyzacji zadania klasyfikacji

Rozwiązywany w niniejszym artykule problem dotyczy regularyzacji zadania klasyfikacji dopasowanej do wzorców. Potrzeba regularyzacji pojawia

się w przypadku, gdy macierze kowariancji $\mathbf{R}_h$ wzorców są osobliwe lub źle uwarunkowane. W rozpatrywanym problemie przez złe uwarunkowanie rozumiemy bardzo dużą rozpiętość między wartościami własnymi macierzy $\mathbf{R}_h$, powodującą, że wyznacznik tej macierzy jest bliski zeru.



Rys. 2. Przykład klasyfikacji w przypadku, gdy macierze kowariancji wzorców są osobliwe i zastosowany jest wzór (14). Z lewej strony: do wyznaczania odległości wykorzystywane są metryki bazujące na macierzy kowariancji obliczonej dla wszystkich wzorców razem, zgodnie ze wzorem (4). Z prawej strony: do wyznaczania odległości wykorzystywane są metryki bazujące na macierzach kowariancji obliczanych dla cech każdego wzorca osobno, zgodnie ze wzorem (11)

Rutynowym postępowaniem w przedstawionym przypadku jest zastosowanie uogólnionej odległości Mahalanobisa, zdefiniowanej następującą zależnością [12]:

$$\mathrm{D}_h(\mathbf{x}, C(W_h)) = \mathrm{d}_h(\mathbf{x}, \overline{\mathbf{w}}_h) = \sqrt{(\mathbf{x} - \overline{\mathbf{w}}_h)^T \mathbf{R}_h^+ (\mathbf{x} - \overline{\mathbf{w}}_h)} \tag{14}$$

gdzie: $\mathbf{R}_h^+$ – pseudoinwersja Moore'a–Penrose'a macierzy kowariancji $\mathbf{R}_h$.

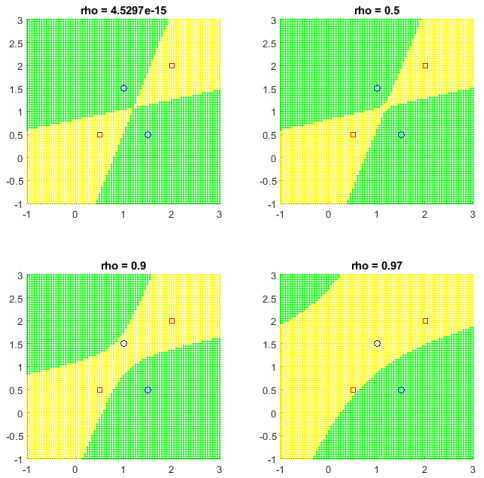
W zadaniach klasyfikacji na podstawie wzorców nieseparowalnych liniowo, postępowanie bazujące na uogólnionej odległości Mahalanobisa może jednak prowadzić do fałszywych rozwiązań. Zilustrujemy to przykładem przedstawionym na rysunku 2. Podobnie jak w przykładzie przedstawionym wcześniej, przestrzeń R^2 jest dzielona na dwie klasy na podstawie zadanych wzorców: W_1 i W_2. Punkty $C(W_1)$ są przedstawione na rysunku jako kółka, punkty $C(W_2)$ – jako kwadraty. W porównaniu z przykładem poprzednim oba klastry W_1 i W_2 są mniej liczne: każda klasa jest wskazywana tylko przez dwa przykłady. Punkty przestrzeni cech leżące bliżej punktów $C(W_1)$ są oznaczone kolorem ciemniejszym. W przedstawionym przykładzie klastry $C(W_1)$ i $C(W_2)$ nie są liniowo separowalne i obie metody klasyfikacji dały złe (nieoczekiwane)

wyniki, przy czym wynik zastosowania metody wykorzystującej metryki dopasowane jest wręcz przeciwny do oczekiwanego.

Proponowana metoda regularyzacji polega na wykorzystaniu następującej metryki:

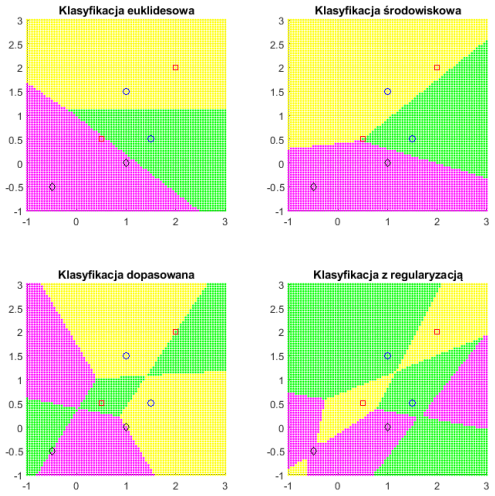
$$\mathrm{d}_h^r(\mathbf{x},\mathbf{y}) = \sqrt{(\mathbf{x}-\mathbf{y})^T[(1-\rho)\mathbf{R}_h + \rho\mathbf{R}]^{-1}(\mathbf{x}-\mathbf{y})}\,, \quad \mathbf{x},\mathbf{y} \in R^L \tag{15}$$

gdzie: $\rho \in [0,1]$ – współczynnik regularyzacji. Metryka $\mathrm{d}_h^r(\mathbf{x},\mathbf{y})$ różni się od metryki $\mathrm{d}_h(\mathbf{x},\mathbf{y})$ zastąpieniem macierzy kowariancji $\mathbf{R}_h$ kombinacją wypukłą tej macierzy i macierzy kowariancji $\mathbf{R}$. Wartość $\rho = 0$ oznacza brak regularyzacji i klasyfikację dopasowaną, wartość $\rho = 1$ oznacza przejście do klasyfikacji środowiskowej.



Rys. 3. Ilustracja wyników klasyfikacji dopasowanej z zastosowaniem regularyzacji

Uzyskane rezultaty zilustrujemy dla danych jak na rysunku 2. Na rysunku 3 przedstawiono wyniki klasyfikacji dopasowanej do wzorców z zastosowaniem regularyzacji. Poprawne wyniki klasyfikacji zostały zaobserwowane już przy wartości współczynnika regularyzacji ρ rzędu 10^{-15}. Wartość maksymalna tego współczynnika wyniosła około 0,5 (dalsze zwiększanie wartości tego współczynnika skutkuje łagodnym przechodzeniem do wyników klasyfikacji środowiskowej).

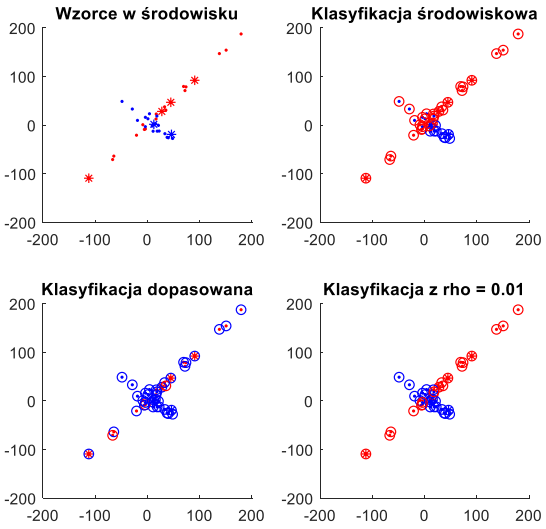


Rys. 4. Porównanie wyników klasyfikacji z regularyzacją do wyników klasyfikacji innymi metodami

Na rysunku 4 przedstawiono analogiczne wyniki dla zadania podziału przestrzeni cech na trzy klasy. Porównano wyniki klasyfikacji bazującej na metryce euklidesowej (na rysunku oznaczone tytułem: klasyfikacja euklidesowa), klasyfikacji bazującej na metryce Mahalanobisa (na rysunku oznaczonej jako klasyfikacja środowiskowa), klasyfikacji dopasowanej do wzorców, bazującej na uogólnionej metryce Mahalanobisa (oznaczonej jako klasyfikacja dopasowana) i klasyfikacji dopasowanej do wzorców z zastosowaniem regularyzacji ze współczynnikiem $\rho = 0{,}01$ (oznaczonej jako klasyfikacja z regularyzacją). Można zauważyć, że tylko wyniki ostatniej klasyfikacji były satysfakcjonujące.

W celu zilustrowania wpływu wartości współczynnika regularyzacji na jakość klasyfikacji przedstawimy wynik eksperymentu obliczeniowego, polegającego na dokonaniu podziału zbioru obiektów na dwie klasy. W eksperymencie losowane były charakterystyki N obiektów pierwszej klasy i tyle samo obiektów drugiej klasy. Spośród nich losowo wskazywane było N_1 przykładów obiektów pierwszej klasy i N_2 przykładów dla obiektów drugiej klasy. Dla obu klas obiektów charakterystyki są punktami na płaszczyźnie, losowanymi zgodnie z odpowiednio dobranymi rozkładami normalnymi. W przedstawionym na rysunku 5 przykładzie przyjęto: $N = 20$, $N_1 = 4$, $N_2 = 2$. Można zauważyć, że podprzestrzeń generowana przez dwa wskazania obiektów jest prostą i zadanie klasyfikacji dopasowanej jest źle uwarunkowane.

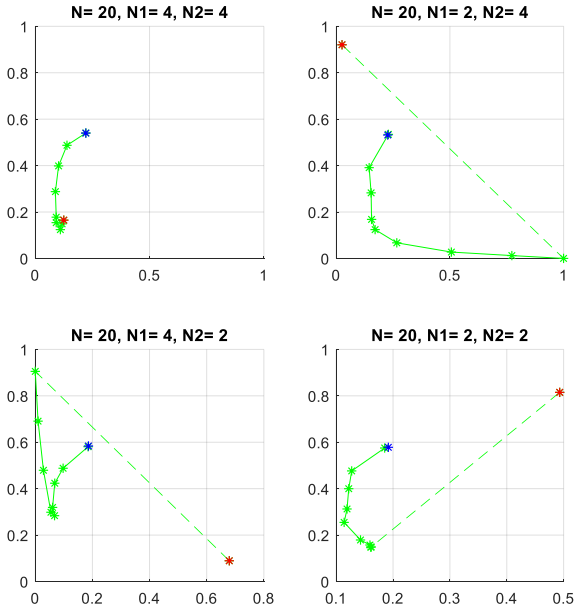
Regularyzacja tego zadania ze współczynnikiem $\rho = 0{,}01$ pozwoliła uzyskać poprawne wyniki klasyfikacji.



Rys. 5. Przykład eksperymentu obliczeniowego. Obiekty klasy 1 są oznaczone punktami koloru czerwonego, a klasy 2 – koloru niebieskiego. Wskazane przykłady są oznaczone gwiazdkami odpowiedniego koloru. Wyniki klasyfikacji są oznaczone kółkami odpowiedniego koloru

Zbiorcze wyniki omawianego eksperymentu obliczeniowego zostały przedstawione na rysunku 6. Oś odciętych przedstawia częstość błędnego zaklasyfikowania do klasy 2, zaś oś rzędnych – częstość błędnego zaklasyfikowania do klasy 1. Częstość została wyznaczona na podstawie 1000 prób. Kolorem zielonym oznaczono wyniki klasyfikacji metodą dopasowaną z regularyzacją – dla różnych wartościach współczynnika regularyzacji $\rho \in (0,1)$. Skrajny punkt dla wartości $\rho = 1$ (oznaczony kolorem niebieskim) odpowiada jakości klasyfikacji środowiskowej. Skrajny punkt dla wartości $\rho = 0$ (oznaczony kolorem czerwonym) odpowiada jakości klasyfikacji dopasowanej bez regularyzacji. Zauważalny jest brak ciągłości uzyskanych charakterystyk przy przejściu od zerowej wartości współczynnika regularyzacji ($\rho = 0$) do wartości dodatniej. Zaobserwowany w eksperymencie przeskok następował przy wartości $\rho \approx 10^{-14}$. Dla wykorzystanego środowiska obliczeniowego jest to minimalna wartość współczynnika regularyzacji. Uzyskane krzywe jakości klasyfikacji z regularyzacją dążą do krzywej ilustrującej sytuację, kiedy regularyzacja nie jest konieczna ($N_1 = 4$, $N_2 = 4$).

Jednak i w tym przypadku można zaobserwować możliwość nieznacznej poprawy jakości klasyfikacji dzięki zastosowaniu regularyzacji. W omawianym eksperymencie akceptowalne wyniki klasyfikacji były uzyskiwane dla wartości współczynnika regularyzacji z przedziału od 0,01 do 0,1.



Rys. 6. Ilustracja wpływu współczynnika regularyzacji na jakość klasyfikacji. Na osi odciętych przedstawiona jest częstość błędnego zaklasyfikowania do klasy 2, na osi rzędnych – częstość błędnego zaklasyfikowania do klasy 1

4. Wnioski

1) Proponowana metoda regularyzacji ewaluacji może być wykorzystywana wszędzie tam, gdzie cechy klasyfikowanych obiektów można przedstawić w postaci wektorów liczb rzeczywistych. W przypadku, gdy macierz kowariancji wszystkich badanych obiektów jest osobliwa (lub źle uwarunkowana), należy przeprowadzić przetwarzanie wstępne w celu wyłonienia tych cech, które zapewnią nieosobliwość ich macierzy kowariancji.
2) Zadanie pochodne klasyfikacji ma przejrzystą interpretację. Zaproponowane podejście polega na uzupełnieniu danych o wzorcach danymi o właściwościach statystycznych środowiska.

3) Algorytm obliczeniowy jest atrakcyjny ze względu na swą prostotę. Daje możliwość użycia bardziej złożonych metod wyznaczania odległości między klastrami niż metoda wykorzystana w przedstawionych w artykule przykładach. Możliwe jest także uzyskiwanie klasyfikacji bazujących na różnych zasadach oceny podobieństwa klastrów.
4) Uzyskanie w wyniku klasyfikacji trywialnych, wieloczłonowych albo zbyt obszernych klas świadczy zwykle o niespójności dokonanych wskazań wzorców klas.

Literatura

[1] BENEDIKTSSON J.A., BENEDIKTSSON K., *Hybrid consensus theoretic classification with pruning and regularization*. IEEE 1999 International Geoscience and Remote Sensing Symposium, 1999, Volume 5, pp. 2486-2488.

[2] FUKUNAGA K., *Feature Extraction Algorithm Using Distance Transformation.* IEEE Transactions on Computers C-21(1), February 1972, pp. 56-63.

[3] HSUN-HSIEN CHANG, MOURA JOSE M.F., *Classification by Cheeger Constant Regularization*. 2007 IEEE International Conference on Image Processing, 2007, Volume 2, pp. II-209-II-212.

[4] JIANGTAO PENG, LEFEI ZHANG, LUOQING LI, *Regularized set-to-set distance metric learning for hyperspectral image classification*. Pattern Recognition Letters, Volume 83, Part 2, 1 November 2016, pp. 143-151.

[5] JIM JING-YAN WANG, YI WANG, SHIGUANG ZHAO, XIN GAO, *Maximum mutual information regularized classification*. Engineering Applications of Artificial Intelligence, Volume 37, January 2015, pp. 1-8.

[6] JUN WANG, GUANGJUN YAO, GUOXIAN YU, *Semi-supervised classification by discriminative regularization*. Applied Soft Computing, Volume 58, September 2017, pp. 245-255.

[7] KWIATKOWSKI W., *Metody automatycznego rozpoznawania wzorców*. BEL Studio, Warszawa, 2010.

[8] KWIATKOWSKI W., *Wykrywanie anomalii bazujące na wskazanych przykładach*. Przegląd Teleinformatyczny, nr 1-2, 2018, s. 3-21.

[9] KWIATKOWSKI W., *Recommendations as a result of decision evaluations based on reference examples.* Teleinformatics Review, No. 1-2, 2019.

[10] MAHALANOBIS P.C., *On the generalized distance in statistics*. Proceedings of National Institute of Sciences (India), Vol. 2, No. 1, 1936, pp. 49-55.

[11] MAJUMDAR A., WARD R.K., *Classification via group sparsity promoting regularization*. 2009 IEEE International Conference on Acoustics, Speech and Signal Processing, 2009, pp. 861-864.

[12] WARMUS M., *Uogólnienie odległości Mahalanobisa*. Listy Biometryczne, Nr 30-33, 1971, s. 3-7.

[13] TAO ZHANG, CHEN GONG, WENJING JIA, XIAONING SONG, JUN SUN, XIAOJUN WU, *Supervised Image Classification with Self-Paced Regularization*. 2018 IEEE International Conference on Data Mining Workshops (ICDMW), 2018, pp. 411-414.

[14] YANG LI, DAPENG TAO, WEIFENG LIU, YANJIANG WANG, *Co-regularization for classification*. 2014 IEEE International Conference on Security, Pattern Analysis, and Cybernetics (SPAC), 2014, pp. 218-222.

[15] YATING SHEN, YUNYUN WANG, ZHIGUO MAARMUS, *Label-expanded manifold regularization for semi-supervised classification*. 12th International Conference on Intelligent Systems and Knowledge Engineering (ISKE), 2017, pp. 1-4.

[16] ZHILEI CHAI, WEI SONG, HUILING WANG, FEI LIU, *A semi-supervised auto-encoder using label and sparse regularizations for classification*. Applied Soft Computing, Volume 77, April 2019, pp. 205-217.

The regularization method in the classification task according to given examples

ABSTRACT: The article considers the problem of classification based on the given examples of classes. As the feature vector, a complete characteristic of an object is assumed. The peculiarity of the problem being solved is that the number of examples of the class may be less than the dimension of the feature vector, and most of the coordinates of the feature vector can be correlated. As a consequence, the feature covariance matrix calculated for the cluster of examples may be singular or ill-conditioned. This disenable a direct use of metrics based on this covariance matrix. The article presents a regularization method involving the additional use of statistical properties of the environment.

KEYWORDS: regularization, classification, pattern recognition, exploratory data analysis

Praca wpłynęła do redakcji: 29.04.2019 r.

PRZEGLĄD TELEINFORMATYCZNY
No. 3-4, 2019

Device authentication methods in Internet of Things networks

Michał JAROSZ

Institute of Teleinformatics and Cybersecurity, Faculty of Cybernetics, MUT,
ul. gen. Sylwestra Kaliskiego 2, 00-908 Warsaw, Poland
michal.jarosz@wat.edu.pl

ABSTRACT: The paper describes basic requirements for authentication systems used in Internet of Things networks, along with problems and attacks that may hinder or even prevent the process of authentication. The methods currently used in device authentication are also presented.

KEYWORDS: Internet of Things, IoT, device authentication, device identification

1. Introduction

The Internet of Things (IoT) is currently one of the fastest developing branches of IT. The Internet of Things means a distributed network connecting physical objects that can collect data from the environment (using sensors), interact with the environment (using actuators), and communicate with each other, other devices and computers. Data collected by these devices may be collected and analysed to develop actions resulting in savings, increased efficiency or improved products and services [5]. It is estimated that by 2021 there will be 21 billion IoT devices connected to the Internet [37], and one of the major challenges is to ensure proper device authentication [42]. This problem applies not only to Internet of Things devices used in industrial or medical environments, but also to devices in households.

The number of attacks on IoT devices is continuously growing; the reason is undoubtedly the increasing use of IoT devices in various environments, but also insufficient security of IoT devices [40]. According to respondents [34], the area that needs the most improvements is device authentication and authorisation.

Identification is a process in which an entity declares its identity. This is then followed by the authentication process. It checks whether the identity actually exists and whether the entity declaring the identity can use it [26]. The subject of this article is an Internet of Things device. The authentication process is important in the context of access control to secured resources.

The purpose of this article is to review the current methods applied in the authentication of devices used in Internet of Things networks. The second section shows examples of architecture models. The third section includes the requirements to be met by a system for authenticating Internet of Things devices. Problems and threats that occur in IoT device authentication systems are discussed (section 4) based on the four-layer architecture described in the second section. The next section (section 5) describes the current methods of device authentication in Internet of Things networks. The last section summarises the entire paper (section 6).

2. Architecture of Internet of Things systems

The system architecture shows how to divide a system into layers, each with defined functions and interactions with other layers. A model can be used to determine whether system elements of the same layer meet specific requirements. We can find many models in the literature on the construction of Internet of Things systems, but the most common architectures are as follows:

a) three-layer,
b) four-layer.

Figure 1 shows layers of the architecture models described. Other examples of architecture models are described in articles [25], [28] and [41].

Three-layer architecture [22] is the basic architecture model of Internet of Things systems. The first layer is the Perception Layer. This layer receives events from the external environment, such as temperature, humidity, speed or location. This is done using sensors that are built-in or connected to the device. The data received can be pre-processed by the device. Another layer is the Network Layer, whose task is to send data from the perception layer to the application layer. Data are transmitted by wire or wirelessly using technologies such as 3G, 4G, Wi-Fi, Zig-Bee, Bluetooth or LoRa. The last layer is the Application Layer. Elements of this layer are responsible for providing application-specific services to the user. The Application Layer does not participate in the authentication process, but may require device authentication.

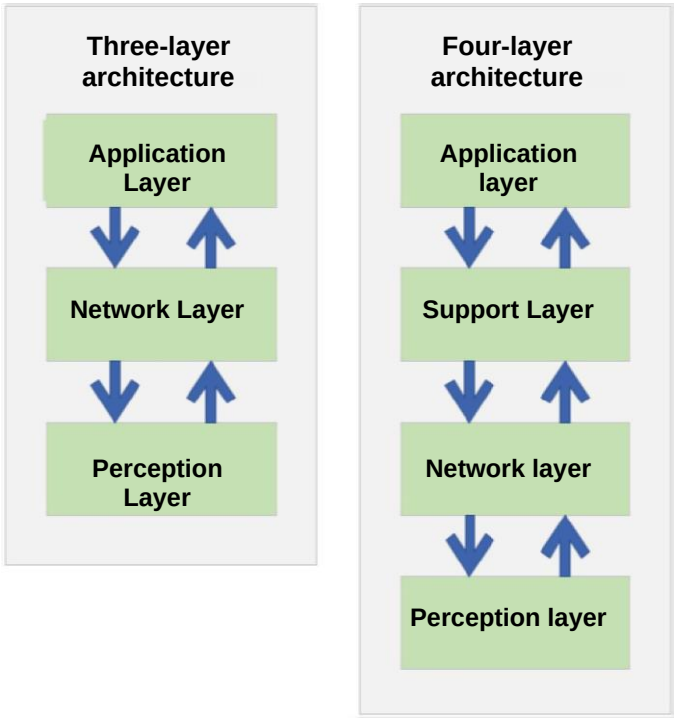


Fig. 1. Example models of Internet of Things system architectures

The four-layer architecture [39] is based on the three-layer architecture. It has the same layers as the three-layer architecture, but it features an additional Support Layer. This layer is responsible for processing and storing data received by sensors in the Perception Layer. These processes usually use cloud services, but can also be performed by a regular computer or disk array. The use of such services is advisable when Internet of Things devices do not have sufficient resources to carry out the task. This is the most common model in literature. The model is used later in this article, as it contains all the layers utilised in the authentication process.

3. Requirements for authentication systems

Authentication is a process confirming the identity of an entity or group of entities. For this purpose, the entity sends its identifier and element confirming the identity. Such an element can be [26]:

a) something known by the entity (e.g. password, PIN),
b) something owned by the entity (e.g. key, token),

c) something that the entity is (e.g. a signature based on the device network traffic),
d) entity location (physical location based on GPS or logical location based on an IP address, for example),
e) something the entity does (e.g. secret handshake).

Because the Internet of Things can consist of millions of devices, a very important requirement for the authentication system is the identifier uniqueness for the entity or group of entities. When two different entities present the same identifier, it may lead to a situation where entity A gains access to information intended for entity B, which of course is unacceptable. If the entity can be clearly identified through a device identifier or the data collected by the device, an appropriate level of privacy and protection of identifiers should be provided during their use and processing.

It should be ensured that the identifier is not changed during assignment, transfer or use [38]. The identifier must be human- and machine-readable and should not contain important information about the identified entity. It is also required to ensure the scalability of identifiers so that each device in the system receives its own identifier. It is impossible to specify in advance how many identifiers the authentication system should be able to process. Even simple environments could grow over time, and a change of the authentication system because of the limited number of processed identities means unnecessary problems. Devices come from different manufacturers and can send data to various applications (not necessarily belonging to the same organisation), so existing standards as well as limitations and capabilities of devices should be considered when developing an identifier generation system. The identifier assigner should keep track of which identifiers are used and which are not. By "disabling" unnecessary identities, it is possible to restrict network access for unauthorised devices.

The authentication process itself should be resistant to the attacks described in Part 4. In addition to these guidelines, we should keep the limitations of IoT devices in mind. These include:

- low memory capacity – the program to be executed by the device must fit into the memory in addition to the authentication system,
- low computing power – since many Internet of Things devices are equipped with processors with low computing power, authentication processes should be as fast as possible, meaning as few calculations as possible,
- low network bandwidth – Internet of Things networks use devices and protocols characterised by low power consumption. However, their limitation is low bandwidth. Therefore the device should send as little data as possible during authentication. It is also necessary to consider the

situation when the device does not have access to the Internet or an authentication server,

- the lowest possible energy consumption – Internet of Things devices can work in hard-to-reach environments where power cannot be supplied. Therefore, IoT devices are made of energy-saving components. It is essential that they are able to operate as long as possible on battery power. The authentication system should ensure the lowest possible CPU load, as far as possible, should not connect to third devices,
- inability to connect additional devices – since Internet of Things devices are small and do not have (or have few) additional ports, the authentication system should not use additional components. It should also be noted that the authentication system will be used in devices of various manufacturers, which may mean different output ports. In the event of a failure, the device can be replaced with another model or even a device from a different manufacturer.

The authentication system must not require a response from the operator, because IoT devices communicate themselves without human intervention. Credentials should be sent in encrypted form so that third parties cannot read them.

4. Problems and threats in authentication systems for Internet of Things devices

This section describes possible problems and attacks that occur in IoT device authentication systems. As mentioned earlier, a four-layer architecture model was used, as it contains all the layers necessary to present the threats. These attacks and problems are assigned to one layer, but some of them may also occur in other layers.

4.1. Support Layer

a. Storage attack – the authentication system can be based on an authentication server. The attack involves changing the credentials on the server or device. The result is the inability to authenticate the device or group of devices. The effects can be more severe if data replication is done between multiple authentication servers [4].
b. Malicious insider attack – an event when a person with authorised access to the system uses its privileges in a negative way. Such a person operates

within the network and most often has direct access to particular data of interest [4].

c. Disaster recovery – this problem occurs, for example, when the only authentication server fails and it needs to be restored. IoT devices cannot perform the authentication process during server recovery. This can be prevented by using at least 2 authentication servers, but it must be ensured that both have the same set of credentials and that the credentials are updated.
d. Brute force attack – obtaining credentials by checking all possible combinations. Rainbow tables can be used to save computing power.
e. Privacy – because Internet of Things devices can be assigned directly to a person, or the data obtained from the device can be used to clearly identify a person, the problem of credentials storage and anonymisation of data obtained from IoT devices should be taken into account.

4.2. Network Layer

a. Eavesdropping – this attack involves eavesdropping between Internet of Things devices or between a device and a server, and obtaining credentials or private data [4].
b. Replay attack – the attacker eavesdrops on the transmission between twos or between a device and a server to obtain credentials. Then the attacker tries to authenticate their data using the obtained credentials [4].
c. Denial of Service (DoS) – an attack preventing the provision of or access to services. It is usually done by sending a large number of requests to the device or by interfering with the transmission [6].
d. Man-in-the-Middle – the attacker acts as an intermediary between devices so that they do not know about its existence. The attacker can change the content of packets sent in real time [6].
e. Device heterogeneity – devices communicate with each other using different communication protocols, so it is important that the authentication system does not rely only on one communication protocol [1].

4.3. Perception Layer

a. Node capture – the attack consists in taking control of the device. If successful, the attacker can obtain credentials and also has network access with the privileges of the captured device [29].
b. Fake and malicious node – the attack involves adding an additional device to the organisation's Internet of Things network. The device sends fake

data. The purpose is to interfere with transmission in the organisation's network. The device added to the system may use the power supply of another node [4].

c. Node tempering – the attack involves replacing the device, changing the device elements to infected ones or adding infected elements to the device [6].
d. Sybil attack – a malicious node has multiple identities (new or taken over from other nodes). This way, it can send data as other nodes or participate in a voting process several times, for example [6].
e. Cryptanalysis – the field dealing with key recovery or data recovery before encryption. Attacks used in cryptanalysis include side-channel attacks [6], timing attacks [4] and brute force attacks.
f. Implementation errors – during implementation of the authentication system the programmer inadvertently makes errors in the code. The attacker can use exploit to take control of the device. In some cases, companies introduce errors into the developed systems on purpose (backdoor).
g. Configuration errors – errors made by people implementing the authentication system, e.g. weak passwords, many entities with the same password, vulnerable algorithms.
h. 0-day attack – the system has a security vulnerability unknown to the manufacturer. The vulnerability could be used to execute malicious code in a device. There is no universal form of protection against this group of attacks. One way to solve this problem may be providing the ability of using other cryptographic methods that are not susceptible to the discovered attack or enabling updates of software on the IoT device [4].

5. Device authentication methods in Internet of Things networks

Many methods of authenticating Internet of Things devices have been developed [7], [30], so this section presents only some of them, focusing on the properties of elements used in authentication systems.

Every IoT device should have its own unique identifier. This can be done manually by the user or the identifier is assigned automatically based on the device's features. If identifier is assigned manually to the device, the applicable standards can be applied, such as FIWARE (using the NGSI standard) or Watson IoT (Table 1). Identification standards are described in paper [10]. Automatic identification is carried out by analysing the device communication. An example of such identification is shown in articles [15], [33]. However, there are two problems with automatic identification based on a device communication analysis:

1) the device will probably not be identified correctly when it starts generating other traffic (e.g. updates),

2) this method is not suitable when there are several identical devices performing the same task in the network.

Table 1. Types of identifiers used in Watson IoT[1]

Customer Type	ID	Identifier Format
Applications	a	a:orgId:appId
Scalable applications	A	A:orgId:appId
Devices	d	d:orgId:deviceType:deviceId
Gateways	g	g:orgId:typeId:deviceId

Based on [35]

Some authentication systems are designed to be used only in specific cases, e.g. for medical purposes [2], [27] or in a smart home [17]. The advantage of personalised systems is the selection of appropriate methods and components for the task. For example, in the case of device authentication in a medical environment, the authentication system is adapted to better data protection compared to IoT devices in a home environment, but the latter may operate faster.

Authentication may apply to not only one communication party, but to both. When only one party is authenticated, it is called a one-way authentication, whereas authentication of both parties of communication is a two-way authentication. There may also be a situation in which a trusted third party is used for authentication (three-way authentication). The disadvantage of three-way authentication might be the increased number of packets to be generated and processed by an IoT device. The best option is to use two-way authentication - then both parties are sure that the data sent comes from a device that has permission to send data and that the data go to a trusted place.

Authentication can be based on:

1) context,
2) identity.

Ad 1) Context-based authentication has been described to some extent at the beginning of this part of the article. A device is authenticated according to its physical characteristics or behaviour. In the previously described case, the researchers have shown that identification can be based on the analysis of the device's network transmission. Then the data are used to create a fingerprint for the authentication process.

[1]orgId – organisation identifier; appId – application identifier, deviceId – device identifier (e.g. serial number), deviceType – device type identifier, typeId – gateway type identifier

Ad 2) In this type of authentication, the device sends or uses an additional element it owns in addition to the identifier. The simplest element is the password/key. However, its main disadvantage is the distribution of a new password/key, e.g. when the old password has been cracked. The key-based authentication method is used in the Directed Path Based Authentication Scheme (DPAS) [18]. The solution to the problem of long-term use of the same password for authentication may be one-time passwords [24]. One-time passwords are changed after each use. The use of one-time passwords[2] presented in paper [24] is resistant to replay attacks and cryptanalysis methods. Asymmetric cryptography can be used instead of passwords. However, it requires more computing power than symmetric cryptography. When using asymmetric cryptography instead of the RSA algorithm, many researchers have been experimenting with elliptic curve cryptography (ECC) [20], [31]. The authentication scheme presented in article [31] is resistant to replay attacks. RSA is considered a safe algorithm since it is based on the factoring of large numbers. The security of elliptic curve cryptography is based on the computational complexity of discrete logarithm search on elliptic curves. In the case of IoT devices, algorithms are based on elliptic curves, because the key used for encryption is shorter than in RSA, with the same security level [36]. The generated keys are used in HMAC *(*keyed-hash message authentication code) [19], [23]. The authentication method presented in paper [19] is resistant to brute force attacks and Men-in-the-Middle attacks. The authentication method presented in paper [23] is resistant to Man-in-the-Middle attack, DoS and cryptanalysis (including side-channel attack). In addition to asymmetric cryptography in HMAC, researchers also create their own systems [17]. The system described in article [17] is resistant to DoS attacks (DDoS), Men-in-the-Middle attacks, replay attacks and brute force attacks.

Public key infrastructure is used instead of keys only, so that public key authenticity is ensured. Examples of authentication systems for Internet of Things devices using public key infrastructure are presented in papers [21] and [32]. The authentication method presented in paper [32] is resistant to node tempering and cryptanalysis. However, in the event of a DoS attack (DDoS), such infrastructure can authenticate the compromised device, even when the certificate has already been revoked. Certificate-based device authentication is also used in the DTLS protocol applied in Internet of Things systems [11].

The element used in the authentication process can also be generated by means of hardware. This is done through a Trusted Platform Module (TPM) [8]. Such a module is responsible for operations involving cryptography (key

[2] Information about vulnerabilities and resistance to attacks has been taken from the cited articles. Many more examples of authentication systems with a list of their vulnerabilities are presented in [7].

generation and storage, encryption). Also, each module has its own unique and secret RSA private key and unique identifier. Naturally, the given IoT device must be equipped with a TPM module. A less popular solution is Physical Unclonable Function (PUF) [14]. PUF is a physical structure made at the chip production stage, which cannot be cloned or changed. It is completely random and is not known even to the manufacturer. The structure generates a response to a signal (request) sent to it. A device is authenticated according to a request-response pair. Using PUF reduces the risk of device cloning because it is impossible to create two identical PUF modules. In some solutions we can find weak PUFs, such as SRAM PUF, which is not unidirectional and mathematically unclonable [3]. They are also vulnerable to numerous attacks (some of which are described in [16]). Instead, it is recommended using strong PUFs, which can generate multiple request-response pairs. There are also systems that feature a unique serial number that can be connected to an IoT device, such as the Maxim DS2411 system. It is used in the authentication scheme presented in paper [9]. The disadvantage of such a system is that its serial number can be read and then used via a program (without involving the module) in another device, which makes it easier to replace an IoT device with another one.

Tables 2 and 3 show the main advantages and disadvantages of authentication systems utilising different elements.

Authentication systems featuring an identifier with an additional element to confirm identity should be used whenever possible. This ensures greater certainty as to the device identity.

6. Conclusion

The article presents the methods of device authentication in Internet of Things networks. Two models of the Internet of Things system architecture have been described in the initial sections. Based on the architecture model, it is easier to identify problems and threats in the authentication systems of Internet of Things devices. A proper threat identification is one of the basic elements of risk analysis in the system design process. Attacks and problems that may pose a threat to device authentication systems in Internet of Things networks have been described based on the four-layer architecture. The article also presents the basic requirements for an identifier and the authentication system itself in relation to Internet of Things devices. The last part shows the properties and methods used in the current authentication systems, including their advantages and disadvantages. It has been described which attacks a given authentication scheme is susceptible to in the mentioned authentication systems.

Table 2. Advantages and disadvantages of authentication systems

Element	Advantages	Disadvantages	Comments
Context	+ IoT device does not require configuration.	– The device can change " its behaviour", such as communication, and then authentication could fail. – It is possible to generate similar network traffic to pretend to be another device.	
Password/key	+ Simple implementation. + Symmetric cryptography is fast.	– Problem of redistributing a new password/key. – The password/key must be stored by the transmitting and receiving devices.	• Passwords can be easy to crack because they are shorter than keys and are not always created at random.
One Time Password	+ The password is used only once. + Resistant to many attacks, including replay attacks.	– In the case of access to a device with a software password generator, it is possible to clone the generator.	

Table 3. Advantages and disadvantages of authentication systems – continued

Component	Advantages	Disadvantages	Comments
RSA, ECC	+ No key distribution problem.	– Slower than symmetric cryptography. – Public key is not authenticated. – Requires more computing power than symmetric cryptography.	
Certificate (PKI)	+ Ensures public key authentication. + Ensures non-repudiation. + Easy identity management.	– Slower than symmetric cryptography. – Communication with a trusted third party is required, e.g. to track the list of revoked certificates.	• Uses asymmetric cryptography.
TPM	+ Secure storage of cryptographic keys. + Hardware-based generation of keys, random numbers.	– A special module is required. – Module damage prevents device authentication.	
PUF	+ Unclonable. + Not susceptible to physical attacks.	– A special module is required. – They may have a high bit error rate.	• Strong PUFs recommended.
Serial number	None.	– Additional module required. – Does not protect against device replacement.	

Credentials are usually stored in a database or file. Recently, there have been many articles using distributed registers [12], [13]. The advantages of distributed registers include decentralization, invariability of stored data and data replication between nodes. As presented in the introduction, the intense development of the Internet of Things forces users to utilise effective and secure authentication systems. Therefore, it is important to continue research on authentication systems and cryptographic protocols for such devices.

Literature

[1] ALI I., SABIR S., ULLAH Z., *Internet of Things Security, Device Authentication and Access Control: A Review*. International Journal of Computer Science and Information Security, Vol. 14, No 8, 2016, pp. 456-466.

[2] ALMULHIM M., ZAMAN N., *Proposing secure and lightweight authentication scheme for IoT based E-health applications*. 2018 20th International Conference on Advanced Communication Technology (ICACT), 2018, pp. 481-487.

[3] BRAEKEN A., *PUF Based Authentication Protocol for IoT*. Symmetry 2018, 10 (8), 352, 2018, pp. 1-15.

[4] BURHAN M., REHMAN R., KHAN B., BYUNG-SEO K., *IoT Elements, Layered Architectures and Security Issues: A comprehensive Survey*. Sensors 2018, pp. 1-37.

[5] DAVIES R., *The Internet of Things – Opportunities and challenges*. European Parliamentary Research Service, 2015, pp. 1-8.

[6] DEOGIRIKAR J., VIDHATE A., *Security Attacks in IoT: A Survey*. International Conference on I-SMAC, 2017, pp. 32-37.

[7] FERRAG M.A., MAGLARAS L.A., JANICKE H., JIANG J., *Authentication Protocols for Internet of Things: A Comprehensive Survey*. Hindawi, Security and Communication Networks, 2017, ID 6562953, pp. 1-41.

[8] FURTAK J., ZIELIŃSKI Z., CHUDZIKIEWICZ J., *Procedures for sensor nodes operation in the secured domain*. Concurrency and Computation: Practice and Experience, 2019, e5183, pp. 1-13.

[9] HASAN A., QUERSHI K., *Internet of Things Device Authentication Scheme using Hardware Serialization*. 2018 International Conference on Applied and Engineering Mathematics, 2018, pp. 109-114.

[10] KOO J., OH S.-R., KIM Y.-G., *Device Identification Interoperability in Heterogeneous IoT Platforms*. Sensors 2019, 2019, pp. 1-16.

[11] KOTHMAYR T., SCHMITT C., HU W., BRÜNIG M., CARLE G., *A DTLS Based End-To-End Security Architecture for the Internet of Things with Two-Way Authentication*. Local Computer Networks Workshops, 2012, pp. 956-963.

[12] LAU C., YEUNG A., YAN F., *Blockchain-based Authentication in IoT Networks*. 2018 IEEE Conference on Dependable and Secure Computing (DSC), 2018, pp. 1-8

[13] LEE C., KIM K., *Implementation of IoT System using BlockChain with Authentication and Data Protection*. 2018 International Conference on Information Networking (ICOIN), 2018, pp. 936-940.

[14] MAES R., VERBAUWHEDE I., *Physically Unclonable Functions: a Study on the State of the Art and Future Research Directions*. Towards Hardware-Intrinsic Security, 2010, pp. 1-37.

[15] MEIDAN Y., BOHADANA M., SHABTAI A., GUARNIZO J.D., OCHOA M., TIPPENHAUER N.O., ELOVICI Y., *ProfilIoT: A Machine Learning Approach for IoT Device Identification Based on Network Traffic Analysis*. SAC'17 Proceedings of the Symposium on Applied Computing, 2017, pp. 506-509.

[16] MUKHOPADHYPAY D., *PUFs as Promising Tools for Security in Internet of Things*. IEEE Design & Test, Vol. 33, Issue 3, 2016, pp. 103-115.

[17] NICANFAR H., JOKAR P., LEUNG V., *Smart Grid Authentication and Key Management for Unicast and Multicast Communications*. 2011 IEEE PES Innovative Smart Grid Technologies, 2011, https://ieeexplore.ieee.org/document/6167151

[18] NING H., LIU H., LIU Q., JI G., *Directed Path Based Authentication Scheme for the Internet of Things*. Journal of Universal Computer Science Vol. 18, No. 9, 2012, pp. 1112-1131.

[19] RABIAH A., RAMAKRISHNAN K., LIRI E., KAR K., *A Lightweight Authentication and Key Exchange Protocol for IoT*. Workshop on Decentralized IoT Security and Standards 2018, 2018, pp. 1-6.

[20] SCHIMTT C., NOACK M., STILLER B., *TinyTO: Two-way Authentication for Constrained Devices in the Internet-of-Things*. Internet of Things, 2015, pp. 239-258

[21] SCHUKAT M., CARTIJO P., *Public key infrastructures and digital certificates for the Internet of things*, 2015 26th Irish Signals and Systems Conference (ISSC), 2015, https://ieeexplore.ieee.org/abstract/document/7163785

[22] SETHI P., SARANGI S.R., *Internet of Things: Architectures, Protocols, and Applications*. Journal of Electrical and Computer Engineering, 2017, pp. 1-25.

[23] SHAH T., VENKATESAN S., *Authentication of IoT Device and IoT Server Using Security Vaults*. 2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications, 2017, pp. 819-824.

[24] SHIVRAJ V., RAJAN M., SINGH M., BALAMURALIDHAR P., *One time password authentication scheme based on elliptic curves for Internet of Things (IoT).* 2015 5th National Symposium on Information Technology: Towards New Smart World (NSITNSW), 2015, pp. 1-6.

[25] SPIESS P. I INNI, *SOA-based Integration of the Internet of Things in Enterprise Services*. 2009 IEEE International Conference on Web Services, 2009, pp. 968-975.

[26] STEWART J.M., *CompTIA Security+ Review Guide*. Sybex, Indianapolis, 2014.

[27] TASALI Q., CHOWDHURY C., VASSERMAN E., *A Flexible Authorization Architecture for Systems of Interoperable Medical Devices.* SACMAT`17, 2017, pp. 9-20.

[28] TORKAMAN A., SEYYEDI M.A., *Analyzing IoT References Architecture Model*. International Journal of Computer Science and Software Engineering, Vol. 5, Issue 8, August 2016, pp. 154-160.

[29] TRIPATHY B.K., ANURADHA J., *Internet of Things (IoT) Technologies, Applications, Challenges and Solutions*. CRC Press, Boca Raton, 2017.

[30] TRNKA M., CERNY T., STICKNEY N., *Survey of Authentication and Authorization for the Internet of Things*. Hindawi, Security and Communication Networks, 2018, ID 4351603, pp. 1-17.

[31] WANG K.H., CHEN C.M., FANG W., WU T.Y., *A secure authentication scheme for Internet of Things*. Pervasive and Mobile Computing 42, 2017, pp. 15-26.

[32] WON J., SINGLA A., BERTINO E., BOLLELLA G., *Decentralized Public Key Infrastructure for Internet-of-Things*. Milcom 2018 Track 5, 2018, pp. 1-7.

Electronic sources

[33] ALUTHGE N., *IoT device fingerprinting with sequence-based features*, 2017, https://helda.helsinki.fi/handle/10138/234247 (accessed on 12.05.2019)

[34] *An overview of the IoT Security Market Report 2017-2022*, https://iiot-world.com/reports/an-overview-of-the-iot-security-market-report-2017-2022/ (accessed on 12.05.2019)

[35] *Connecting applications, devices and gateways*, IBM, https://www.ibm.com/support/knowledgecenter/en/SSQP8H/iot/platform/reference/security/connect_devices_apps_gw.html (accessed on 12.05.2019)

[36] ECC 101: What is ECC and why would I want to use it?, https://www.globalsign.com/en/blog/elliptic-curve-cryptography/ (accessed on 20.06.2019)

[37] *Gartner Identifies Top 10 Strategic IoT Technologies and Trends*, https://www.gartner.com/en/newsroom/press-releases/2018-11-07-gartner-identifies-top-10-strategic-iot-technologies-and-trends, 2018 (accessed on 12.05.2019)

[38] *Identifiers in Internet of Things*, Alliance for Internet of Things Innovation, Version 1.0, 2018, https://aioti.eu/wp-content/uploads/2018/03/AIOTI-Identifiers_in_IoT-1_0.pdf.pdf, (accessed on 10.05.2019)

[39] Series Y: Global Information Infrastructure, Internet Protocol Aspects and Next-Generation Network. Overview of the Internet of things, TELECOMMUNICATION STANDARIZATION SECTOR OD ITU, 2012, https://www.itu.int/rec/dologin_pub.asp?lang=e&id=T-REC-Y.2060-201206-I!!PDF-E&type=items (accessed on 12.05.2019)

[40] *The Internet of Things (IoT) – Threats and Countermeasures*, https://www.cso.com.au/article/575407/internet-things-iot-threats-countermeasures/ (accessed on 12.05.2019)

[41] *The Internet of Things Reference Model*, Cisco 2014, http://cdn.iotwf.com/resources/71/IoT_Reference_Model_White_Paper_June_4_2014.pdf (accessed on 10.05.2019)

[42] Top 10 IoT security challenges, https://developer.ibm.com/articles/iot-top-10-iot-security-challenges/, 2017 (accessed on 12.05.2019)

Sposoby uwierzytelniania urządzeń w sieciach Internetu Rzeczy

STRESZCZENIE: W artykule opisano podstawowe wymagania systemów uwierzytelniania stosowanych w sieciach Internetu Rzeczy oraz problemy i ataki, które mogą utrudnić lub nawet uniemożliwić przeprowadzenie procesu uwierzytelniania. Przedstawiono również obecne metody stosowane w uwierzytelnianiu urządzeń Internetu Rzeczy.

SŁOWA KLUCZOWE: Internet Rzeczy, IoT, uwierzytelnianie urządzeń, identyfikacja urządzeń

Received by the editorial staff on: 12.07.2019

PRZEGLĄD TELEINFORMATYCZNY
NR 3-4, 2019

Sposoby uwierzytelniania urządzeń w sieciach Internetu Rzeczy

Michał JAROSZ

Instytut Teleinformatyki i Cyberbezpieczeństwa, Wydział Cybernetyki, WAT
ul. gen. Sylwestra Kaliskiego 2, 00-908 Warszawa
michal.jarosz@wat.edu.pl

STRESZCZENIE: W artykule opisano podstawowe wymagania systemów uwierzytelniania stosowanych w sieciach Internetu Rzeczy oraz problemy i ataki, które mogą utrudnić lub nawet uniemożliwić przeprowadzenie procesu uwierzytelniania. Przedstawiono również obecne metody stosowane w uwierzytelnianiu urządzeń Internetu Rzeczy.

SŁOWA KLUCZOWE: Internet Rzeczy, IoT, uwierzytelnianie urządzeń, identyfikacja urządzeń

1. Wstęp

W dzisiejszych czasach jedną z najszybciej rozwijających się gałęzi informatyki jest Internet Rzeczy (ang. *Internet of Things,* IoT). Internet Rzeczy odnosi się do rozproszonej sieci łączącej obiekty fizyczne, które są zdolne do zbierania danych z otoczenia (za pomocą sensorów), oddziaływania na otoczenie (przy użyciu aktuatorów) oraz komunikowania się ze sobą, innymi urządzeniami i komputerami. Dane zebrane przez te urządzenia mogą być gromadzone i analizowane w celu wypracowania działań, które przyniosą oszczędności, zwiększą wydajność lub ulepszą produkty i usługi [5]. Szacuje się, że do 2021 roku będzie 21 miliardów urządzeń IoT podłączonych do Internetu [37], a jednym z istotnych wyzwań jest zapewnienie odpowiedniego uwierzytelniania urządzeń [42]. Problem ten nie odnosi się tylko do urządzeń Internetu Rzeczy stosowanych w środowisku przemysłowym czy medycznym, ale także w urządzeniach wykorzystywanych w gospodarstwach domowych.

Ciągle rośnie liczba ataków na urządzenia IoT, powodem takiego stanu jest niewątpliwie coraz częstsze wykorzystanie urządzeń IoT w różnych

środowiskach, ale także niedostateczny poziom zabezpieczeń urządzeń Internetu Rzeczy [40]. Według respondentów [34] obszarem, który potrzebuje największych usprawnień jest ten związany z uwierzytelnianiem i autoryzacją urządzeń.

Identyfikacja jest to proces, w którym podmiot deklaruje swoją tożsamość. Po zadeklarowaniu przez podmiot tożsamości następuje proces uwierzytelniania. W tym procesie sprawdzane jest, czy taka tożsamość rzeczywiście istnieje oraz czy podmiot deklarujący swoją tożsamość jest podmiotem, który może jej używać [26]. W przypadku tego artykułu podmiotem jest urządzenie Internetu Rzeczy. Proces uwierzytelniania jest istotny w kontekście kontroli dostępu do chronionych zasobów.

Celem tego artykułu jest przegląd obecnych sposobów uwierzytelniania urządzeń stosowanych w sieciach Internetu Rzeczy. W punkcie drugim przedstawiono przykładowe modele architektur. Punkt trzeci zawiera wymagania, jakie musi spełniać system uwierzytelniania urządzeń Internetu Rzeczy. Na podstawie architektury czterowarstwowej opisanej w punkcie drugim omówiono problemy i zagrożenia, jakie występują w systemach uwierzytelniania urządzeń IoT (punkt 4). Następnie opisano obecnie stosowane sposoby uwierzytelniania urządzeń w sieciach Internetu Rzeczy (punkt 5). Ostatni punkt zawiera podsumowanie całej pracy (punkt 6).

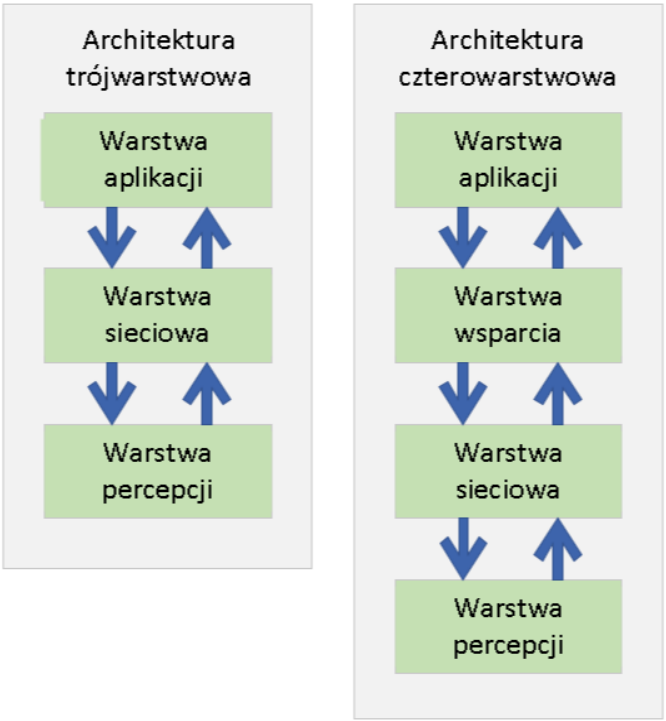
2. Architektura systemów Internetu Rzeczy

Architektura systemu przedstawia sposób podziału systemu na warstwy, z których każda ma zdefiniowane funkcje oraz relacje z innymi warstwami. Na podstawie modelu można określić, czy elementy systemu należące do tej samej warstwy spełniają określone wymagania. W literaturze dotyczącej budowy systemów Internetu Rzeczy można spotkać wiele modeli architektur, jednak do najczęściej używanych należą architektury:

a) trójwarstwowa,

b) czterowarstwowa.

Na rysunku 1 przedstawiono warstwy opisywanych modeli architektur. Inne przykłady modeli architektur zostały opisane w artykułach [25], [28], [41].



Rys. 1. Przykładowe modele architektur systemu Internetu Rzeczy

Architektura trójwarstwowa [22] jest podstawowym modelem architektury systemów Internetu Rzeczy. Pierwszą warstwą jest warstwa percepcji (ang. *Perception Layer*). W tej warstwie następuje odbiór zdarzeń ze środowiska zewnętrznego, np. temperatury, wilgotności, prędkości, lokalizacji. Odbiór następuje przy użyciu czujników, które są wbudowane lub połączone z urządzeniem. Odebrane dane mogą być wstępnie przetwarzane na urządzeniu. Kolejną warstwą jest warstwa sieciowa (ang. *Network Layer*), której zadaniem jest przesłanie danych z warstwy percepcji do warstwy aplikacji. Dane są przekazywane przewodowo lub bezprzewodowo, wykorzystując m.in. technologie 3G, 4G, Wi-Fi, Zig-Bee, Bluetooth lub LoRa. Ostatnią warstwą jest warstwa aplikacji (ang. *Applicaiton Layer*). Elementy tej warstwy są odpowiedzialne za dostarczenie użytkownikowi usług specyficznych dla aplikacji. Warstwa aplikacji nie bierze udziału w procesie uwierzytelniania, może natomiast wymagać uwierzytelnienia urządzenia.

Architektura czterowarstwowa [39] bazuje na architekturze trójwarstwowej. Posiada ona takie same warstwy, jak architektura trójwarstwowa, ale poza tymi warstwami ma dodatkową warstwę wsparcia (ang. *Support Layer*). Warstwa ta odpowiada za przetwarzanie i przechowywanie

danych odebranych przez czujniki w warstwie percepcji. Najczęściej procesy te odbywają się z wykorzystaniem usług chmurowych, ale mogą być wykonywane przez zwykły komputer lub macierz dyskową. Wykorzystanie takich usług jest wskazane w przypadku, gdy urządzenia Internetu Rzeczy nie mają wystarczających zasobów do realizacji zadania. Jest to najczęściej spotykany model w literaturze. Model ten został wykorzystany w dalszej części tego artykułu, ponieważ zawiera on wszystkie warstwy, które są wykorzystywane w procesie uwierzytelniania.

3. Wymagania systemów uwierzytelniania

Uwierzytelnianie jest to proces, w którym potwierdzana jest tożsamość podmiotu lub grupy podmiotów. W tym celu podmiot wysyła swój identyfikator oraz element potwierdzający swoją tożsamość. Takim elementem może być [26]:

a) coś, co podmiot zna (np. hasło, PIN),
b) coś, co podmiot ma (np. klucz, token),
c) coś, czym podmiot jest (np. sygnatura na podstawie ruchu sieciowego urządzenia),
d) lokalizacja podmiotu (fizyczna na podstawie GPS lub logiczna, np. na podstawie adresu IP),
e) coś, co podmiot robi (np. sekretny sposób nawiązywania połączenia (ang. *secret handshake*)).

Ponieważ Internet Rzeczy może składać się z milionów urządzeń, bardzo ważnym wymaganiem dla systemu uwierzytelniania jest unikalność identyfikatora dla podmiotu lub grupy podmiotów. Sytuacja, w której dwa różne podmioty przedstawiają się tym samym identyfikatorem może doprowadzić do sytuacji, kiedy podmiot A uzyskuje dostęp do informacji przeznaczonych dla podmiotu B, co oczywiście jest niedopuszczalne. W przypadku, gdy dzięki identyfikatorowi urządzenia lub danym zebranym przez urządzenie można jednoznacznie zidentyfikować podmiot, należy zastosować odpowiedni poziom prywatności i ochrony identyfikatorów podczas ich używania i przetwarzania.

Należy zapewnić, aby identyfikator nie był zmieniony podczas przydzielania, przekazywania i użytkowania [38]. Identyfikator musi być czytelny dla człowieka i maszyny oraz nie powinien zawierać istotnych informacji o identyfikowanym podmiocie. Trzeba również zapewnić skalowalność identyfikatorów tak, aby każde urządzenie w systemie otrzymało swój własny identyfikator. Nie można z góry określić, ile identyfikatorów system uwierzytelniania powinien móc obsłużyć. Nawet proste środowiska mogą z czasem się rozrastać, a zmiana systemu uwierzytelniania z powodu

ograniczonej liczby obsługiwanych tożsamości jest wprowadzeniem niepotrzebnych problemów. Urządzenia pochodzą od różnych producentów i mogą przesyłać dane do różnych aplikacji (niekoniecznie należących do tej samej organizacji), dlatego w opracowywaniu systemu generowania identyfikatorów należy wziąć pod uwagę istniejące standardy oraz ograniczenia i możliwości urządzeń. Organ nadający identyfikator powinien śledzić, które identyfikatory są wykorzystywane, a które nie. Dzięki „wyłączeniu" zbędnych tożsamości można ograniczyć dostęp do sieci dla nieupoważnionych urządzeń.

Sam proces uwierzytelniania powinien być odporny na ataki opisane w części 4. Poza tymi wytycznymi, należy pamiętać o ograniczeniach urządzeń Internetu Rzeczy. Można do nich zaliczyć:

- małą ilość pamięci – w pamięci poza systemem uwierzytelniania musi zmieścić się jeszcze program, który będzie wykonywany przez to urządzenie,
- niską moc obliczeniową urządzenia – ponieważ w wielu urządzeniach Internetu Rzeczy wykorzystano procesory cechujące się niską mocą obliczeniową, należy zwrócić uwagę, aby proces uwierzytelniania trwał możliwie jak najszybciej, a tym samym wykonywał jak najmniej obliczeń,
- niską przepustowość sieci – w sieciach Internetu Rzeczy wykorzystuje się urządzenia oraz protokoły cechujące się niskim poborem mocy. Mają one jednak ograniczenie w postaci niskiej przepustowości. Dlatego podczas realizacji uwierzytelniania, urządzenie powinno przesyłać jak najmniej danych. Trzeba również wziąć pod uwagę sytuację, gdy urządzenie nie ma dostępu do Internetu lub serwera uwierzytelniającego,
- jak najmniejsze zużycie energii – urządzenia Internetu Rzeczy mogą pracować w środowiskach trudno dostępnych, do których nie można doprowadzić zasilania. Dlatego urządzenia Internetu Rzeczy są zbudowane z energooszczędnych komponentów. Istotne jest, aby były one w stanie pracować jak najdłużej na zasilaniu bateryjnym. Wykorzystywany system uwierzytelniania powinien jak najmniej obciążać procesor, a także w miarę możliwości nie łączyć się z urządzeniami trzecimi,
- brak możliwości podłączenia dodatkowych urządzeń – ponieważ urządzenia Internetu Rzeczy są małe, nie mają dodatkowych portów (lub mają ich mało), wykorzystywany system uwierzytelniania nie powinien korzystać z dodatkowych komponentów. Należy zwrócić uwagę również na to, że system uwierzytelniania będzie wykorzystywany na urządzeniach różnych producentów, a te mogą nie mieć lub będą miały inne porty wyjściowe. W przypadku awarii urządzenie może zostać zamienione na inny model lub nawet na urządzenie innego producenta.

Stosowany system uwierzytelniania nie może wymagać reakcji operatora, ponieważ urządzenia Internetu Rzeczy komunikują się same bez ingerencji człowieka. Dane uwierzytelniające powinny być przesyłane w postaci zaszyfrowanej, tak aby osoby trzecie nie mogły odczytać tych danych.

4. Problemy i zagrożenia w systemach uwierzytelniania urządzeń Internetu Rzeczy

W tym punkcie opisano możliwe problemy i ataki występujące w systemach uwierzytelniania urządzeń IoT. Tak jak wcześniej wspomniano, wykorzystano czterowarstwowy model architektury, ponieważ zawiera on wszystkie warstwy niezbędne do przedstawienia zagrożeń. Opisane ataki i problemy przyporządkowano do jednej warstwy, ale niektóre z nich mogą występować również w innych warstwach.

4.1. Warstwa wsparcia

a. Atak na zasób danych (ang. *storage attack*) – system uwierzytelniania może być oparty o serwer uwierzytelniania. Atak polega na zmianie danych uwierzytelniających na serwerze lub na urządzeniu. Efektem jest brak możliwości uwierzytelniania urządzenia lub grupy urządzeń. Skutki mogą być poważniejsze, jeżeli wykonywana jest replikacja danych pomiędzy wieloma serwerami uwierzytelniającymi [4].
b. Złośliwe działanie osób uwierzytelnionych (ang. *malicious insider attack*) – zagrożenie polegające na tym, że osoba z autoryzowanym dostępem do systemu wykorzystuje swoje uprawnienia w sposób negatywny. Osoba taka działa wewnątrz sieci oraz najczęściej ma bezpośredni dostęp do danych ją interesujących [4].
c. Odtwarzanie awaryjne (ang. *disaster recovery*) – problem występuje przykładowo w sytuacji, gdy awarii ulega jedyny serwer uwierzytelniania i konieczne jest przywrócenie jego działania. Urządzenia IoT w czasie odtwarzania serwera nie mają możliwości wykonania procesu uwierzytelniania. Można temu zaradzić, wykorzystując przynajmniej 2 serwery uwierzytelniania, ale trzeba zapewnić, aby obydwa miały ten sam zestaw danych uwierzytelniających oraz dane uwierzytelniające były aktualne.
d. Atak typu brute-force – atak polegający na uzyskaniu danych uwierzytelniających poprzez sprawdzenie wszystkich możliwych kombinacji. W celu zaoszczędzenia mocy obliczeniowej można wykorzystać tęczowe tablice.

e. Problem prywatności – ponieważ urządzenia Internetu Rzeczy mogą być przypisane bezpośrednio do osoby, lub na podstawie danych uzyskanych z urządzenia można określić jednoznacznie osobę, należy wziąć pod uwagę problem przechowywania danych uwierzytelniających oraz anonimizacji danych uzyskanych od urządzeń IoT.

4.2. Warstwa sieciowa

a. Podsłuchanie transmisji (ang. *eavesdropping*) – atak polega na podsłuchaniu transmisji pomiędzy urządzeniami Internetu Rzeczy lub między urządzeniem a serwerem i uzyskaniu danych uwierzytelniających lub danych prywatnych [4].
b. Atak metodą powtórzenia (ang. *replay attack*) – atakujący podsłuchuje transmisję pomiędzy dwoma urządzeniami lub między urządzeniem a serwerem w celu uzyskania danych uwierzytelniających. Następnie próbuje uwierzytelnić się, wykorzystując uzyskane dane uwierzytelniające [4].
c. Denial of Service (DoS) – atak polegający na uniemożliwieniu realizacji lub dostępu do usług. Najczęściej jest realizowany poprzez wysłanie dużej ilości żądań do urządzenia lub poprzez zakłócenie transmisji [6].
d. Man-in-the-Middle – atakujący pełni rolę pośrednika pomiędzy urządzeniami tak, aby te urządzenia nie wiedziały o jego istnieniu. Atakujący może zmieniać zawartość przesyłanych pakietów w czasie rzeczywistym [6].
e. Problem heterogoniczności urządzeń – urządzenia komunikują się między sobą, wykorzystując różne protokoły komunikacji, dlatego ważne jest, aby system uwierzytelniania nie bazował tylko na jednym protokole komunikacji [1].

4.3. Warstwa percepcji

a. Przejęcie urządzenia (ang. *node capture*) – atak polega na przejęciu kontroli nad urządzeniem. W przypadku powodzenia atakujący może uzyskać dane uwierzytelniające, a także ma dostęp do sieci z uprawnieniami przejętego urządzenia [29].
b. Dodanie podrobionego, zainfekowanego urządzenia (ang. *fake and malicious node*) – atak polega na dodaniu dodatkowego urządzenia do sieci Internetu Rzeczy organizacji. Urządzenie wysyła podrobione dane. Celem ataku jest zakłócanie transmisji w sieci organizacji. Urządzenie dodane do systemu może korzystać z zasilania innego węzła [4].

c. Podmiana urządzenia (ang. *node tempering*) – atak polega na podmianie urządzenia, zmianie elementów urządzenia na zainfekowane lub dodaniu zainfekowanych elementów do urządzenia [6].
d. Atak typu Sybil (ang. *sybil attack*) – złośliwy węzeł jest w posiadaniu wielu tożsamości (przejętych od innych węzłów lub stworzonych). Dzięki temu może np. wysyłać dane jako inne węzły lub kilkukrotnie uczestniczyć w procesie głosowania [6].
e. Kryptoanaliza – dziedzina zajmująca się odtworzeniem klucza lub odtworzeniem danych przed zaszyfrowaniem. Do ataków wykorzystywanych w kryptoanalizie można zaliczyć m.in. atak side-channel (ang. *side-channel attack*) [6], atak czasowy (ang. *timming attack*) [4] oraz atak brute-force.
f. Błędy implementacyjne – problem ten polega na tym, że podczas implementacji systemu uwierzytelniania programista nieumyślnie popełnia błędy w kodzie. Atakujący wykorzystując exploit, może przejąć kontrolę nad urządzeniem. W niektórych przypadkach firmy specjalnie wprowadzają błędy do opracowywanych systemów (ang. *backdoor*).
g. Błędy konfiguracyjne – błędy popełniane przez osoby wdrażające system uwierzytelniania, np. słabe hasła, wiele podmiotów ma takie samo hasło, wykorzystanie podatnych algorytmów.
h. Atak typu 0-day – system posiada lukę w zabezpieczeniach, która nie jest znana producentowi. Luka może posłużyć do wykonania złośliwego kodu na urządzeniu. Przed tą grupą ataków nie ma jednoznacznej formy zabezpieczeń. Jednym ze sposobów rozwiązania tego problemu może być np. zapewnienie możliwości skorzystania z innych metod kryptograficznych, które nie są podatne na odkryty atak lub umożliwienie aktualizacji oprogramowania działającego na urządzeniu IoT [4].

5. Sposoby uwierzytelniania urządzeń w sieciach Internetu Rzeczy

Opracowano już bardzo wiele metod uwierzytelniania urządzeń Internetu Rzeczy [7], [30], dlatego w tym punkcie przedstawiono tylko niektóre z nich, skupiając się na właściwościach elementów wykorzystywanych w systemach uwierzytelniania.

Każde urządzenie IoT powinno mieć nadany swój unikalny identyfikator. Może być to wykonane manualnie przez użytkownika lub nadanie identyfikatora nastąpi automatycznie na podstawie cech danego urządzenia. W przypadku manualnego nadawania identyfikatora urządzeniu można skorzystać z obowiązujących standardów, np. FIWARE (wykorzystujący standard NGSI) czy też Watson IoT (tabela 1). Standardy identyfikacji zostały opisane w pracy [10]. Identyfikacja automatyczna wykonana jest poprzez analizę

komunikacji danego urządzenia. Przykład takiej identyfikacji przedstawiono w artykułach [15], [33]. Jednak w przypadku automatycznej identyfikacji opartej na analizie komunikacji konkretnego urządzenia występują dwa problemy:

1) urządzenie prawdopodobnie nie zostanie zidentyfikowane poprawnie, gdy zacznie generować inny ruch (np. na skutek aktualizacji),

2) sposób taki nie nadaje się, gdy w sieci jest kilka takich samych urządzeń, które wykonują to samo zadanie.

Tabela 1. Typy identyfikatorów wykorzystywanych w Watson IoT[1]

Typ Klienta	ID	Format Identyfikatora
Aplikacje	a	a:orgId:appId
Skalowalne aplikacje	A	A:orgId:appId
Urządzenia	d	d:orgId:deviceType:deviceId
Bramy	g	g:orgId:typeId:deviceId

Opracowano na podstawie [35]

Niektóre systemy uwierzytelniania są przygotowane z myślą o wykorzystaniu tylko w konkretnych przypadkach, np. medycznych [2], [27] lub w inteligentnym domu [17]. Zaletą spersonalizowanych systemów jest dobór odpowiednich metod i komponentów do wykonywanego zadania. Na przykład do uwierzytelniania urządzeń w środowisku medycznym, system uwierzytelniania przystosowany jest do zwiększonej ochrony danych niż w przypadku wykorzystania urządzeń IoT w środowisku domowym, ale może być gorszy w szybkości działania.

Uwierzytelnianie może nie dotyczyć tylko jednej strony komunikacji, ale obydwu. Kiedy uwierzytelnia się tylko jedna strona, mówimy o uwierzytelnianiu jednokierunkowym (ang. *one-way authentication*), w przypadku, gdy uwierzytelniają się obydwie strony komunikacji, jest to uwierzytelnianie dwukierunkowe (ang. *two-way authentication*). Może jeszcze zajść sytuacja, w której do uwierzytelniania wykorzystywana jest zaufana trzecia strona (ang. *three-way authentication*). Wadą uwierzytelniania wykorzystującego zaufaną trzecią stronę może być zwiększona liczba pakietów, które muszą być wygenerowane i obsłużone przez urządzenie IoT. Najlepszą opcją jest stosowanie uwierzytelniania dwukierunkowego, wtedy obydwie strony są pewne, że dane wysyłane pochodzą od urządzenia, które ma uprawnienia do wysłania danych oraz dane trafiają do zaufanego miejsca.

[1] orgId – identyfikator organizacji; appId – identyfikator aplikacji, deviceId – identyfikator urządzenia (np. numer seryjny), deviceType – identyfikator typu urządzenia, typeId – identyfikator typu bramy.

Uwierzytelnianie może nastąpić na podstawie:

1) kontekstu,
2) tożsamości.

Ad 1) Uwierzytelnianie na podstawie kontekstu zostało opisane w pewnym stopniu na początku tej części artykułu. Urządzenie uwierzytelniane jest na podstawie jego cech fizycznych lub zachowania. W opisywanym wcześniej przypadku badacze pokazali, że możliwa jest identyfikacja na podstawie analizy transmisji sieciowej urządzenia. Na podstawie danych tworzony jest fingerprint, który następnie jest wykorzystywany w procesie uwierzytelniania.

Ad 2) W tym typie uwierzytelniania urządzenie poza identyfikatorem wysyła lub wykorzystuje dodatkowy element, który jest w posiadaniu tego urządzenia. Najprostszym elementem jest hasło/klucz. Jednak jego główną wadą jest problem dystrybucji nowego hasła/klucza na przykład w momencie, gdy stare hasło zostanie złamane. Sposób uwierzytelniania w oparciu o klucz został wykorzystany w protokole DPAS (*Directed Path Based Authentication Scheme*) [18]. Rozwiązaniem problemu wykorzystania tego samego hasła do uwierzytelniania przez dłuższy czas może być wykorzystanie haseł jednorazowych (ang. *One Time Password*) [24]. Hasła jednorazowe są zmieniane po każdym użyciu. Wykorzystanie haseł jednorazowych[2] przedstawione w pracy [24] jest odporne na atak metodą powtórzenia oraz metody kryptoanalizy. Zamiast haseł można skorzystać z kryptografii asymetrycznej. Wymaga ona jednak większej mocy obliczeniowej niż kryptografia symetryczna. W przypadku wykorzystania kryptografii asymetrycznej zamiast wykorzystywania algorytmu RSA wielu badaczy eksperymentuje z wykorzystaniem kryptografii krzywych eliptycznych (ang. *Elliptic Curve Cryptography*, (ECC)) [20], [31]. Schemat uwierzytelniania przedstawiony w artykule [31] jest odporny na atak metodą powtórzenia. RSA jest uważany za bezpieczny algorytm, ponieważ opiera się na faktoryzacji dużych liczb. Bezpieczeństwo kryptografii krzywych eliptycznych jest natomiast oparte na złożoności obliczeniowej poszukiwania logarytmu dyskretnego na krzywych eliptycznych. W przypadku urządzeń IoT wykorzystuje się algorytmy oparte na krzywych eliptycznych, ponieważ klucz używany do szyfrowania jest krótszy niż w RSA, przy jednakowym poziomie bezpieczeństwa [36]. Wygenerowane klucze są wykorzystywane w uwierzytelnianiu wykorzystującym HMAC *(*ang. *keyed-Hash Message Authentication Code*) [19], [23]. Sposób uwierzytelniania zaprezentowany w pracy [19] jest odporny na atak brute-force i atak Men-in-the-Middle. Sposób uwierzytelniania przedstawiony

[2] Informacje o podatnościach i odporności na ataki zaczerpnięto z przytoczonych artykułów. O wiele więcej przykładowych systemów uwierzytelniania wraz z zestawieniem ich podatności przedstawiono w pracy [7].

w pracy [23] odporny jest na atak Man-in-the-Middle, DoS oraz kryptoanalizę (m.in. atak side-channel). Poza wykorzystaniem kryptografii asymetrycznej w HMAC badacze tworzą też własne systemy [17]. System opisany w artykule [17] jest odporny na ataki DoS (DDoS), Men-in-the-Middle, ataki metodą powtórzenia i ataki brute force.

Zamiast stosowania samych kluczy wykorzystuje się infrastrukturę klucza publicznego, dzięki czemu zyskujemy autentyczność klucza publicznego. Przykładowe systemy uwierzytelniania urządzeń Internetu Rzeczy wykorzystujące infrastrukturę klucza publicznego zostały przedstawione w pracach [21], [32]. Sposób uwierzytelniania przedstawiony w pracy [32] jest odporny między innymi na podmianę urządzenia oraz kryptoanalizę. Jednak w przypadku ataku DoS (DDoS) przedstawiona infrastruktura może uwierzytelnić skradzione urządzenie, pomimo że certyfikat jest już unieważniony. Uwierzytelnianie urządzeń na podstawie certyfikatów wykorzystywane jest także w protokole DTLS stosowanym w systemach Internetu Rzeczy [11].

Element wykorzystywany w procesie uwierzytelniania może być generowany również sprzętowo. Wykorzystywany jest tu moduł TPM (ang. *Trusted Platform Module*) [8]. Moduł taki jest odpowiedzialny za operacje związane z kryptografią (generowanie i przechowywanie kluczy, szyfrowanie). Poza tym każdy moduł ma swój własny unikalny oraz tajny klucz prywatny RSA oraz unikalny identyfikator. Oczywiście urządzenie IoT musi być wyposażone w moduł TPM. Mniejszą popularnością cieszy się PUF (ang. *Physical Unclonable Function*) [14]. PUF jest to fizyczna struktura powstała na etapie produkcji chipu, której nie można sklonować czy też zmienić. Jest ona w pełni losowa i nie jest znana nawet producentowi. Wytworzona struktura generuje odpowiedź do sygnału (żądania) wysłanego do tej struktury. Urządzenie jest uwierzytelniane na podstawie pary żądanie-odpowiedź. Używanie PUF zmniejsza ryzyko sklonowania urządzenia, ponieważ nie można stworzyć dwóch identycznych modułów PUF. W niektórych rozwiązaniach można spotkać wykorzystanie słabych PUF, np. SRAM PUF, która nie jest jednokierunkowa oraz matematycznie nieklonowalna [3]. Są one również podatne na liczne ataki (niektóre ataki zostały opisane w pracy [16]). Zamiast nich zaleca się wykorzystanie silnych PUF, czyli takich, które umożliwiają wygenerowanie wielu par żądanie-odpowiedź. Istnieją także układy, które zawierają unikalny numer seryjny i które można podłączyć do urządzania IoT, np. układ Maxim DS2411. Został on wykorzystany w schemacie uwierzytelniania przedstawionym w pracy [9]. Wadą takiego układu jest to, że można z niego odczytać numer seryjny i następnie wykorzystać taki numer programowo (bez użycia takiego modułu) w innym urządzeniu, co ułatwia podmianę urządzenia IoT na inne.

W tabelach 2 i 3 przedstawiono główne zalety oraz wady systemów uwierzytelniania wykorzystujących różne elementy.

Tabela 2. Zalety i wady systemów uwierzytelniania

Nazwa elementu	Zalety	Wady	Uwagi
Kontekst	+ Urządzenie IoT nie wymaga konfiguracji.	– Urządzenie może zmienić „swoje zachowanie”, np. komunikację i wtedy uwierzytelnianie może nie nastąpić. – Można wygenerować podobny ruch sieciowy i tym samym podszyć się pod inne urządzenie.	
Hasło/klucz	+ Prosta implementacja. + Kryptografia symetryczna jest szybka.	– Problem redystrybucji nowego hasła/ klucza. – Hasło/klucz musi być przechowane/y przez urządzenie nadające i odbierające.	• Hasła mogą być łatwe do złamania, ponieważ są krótsze od kluczy oraz nie zawsze powstają w sposób losowy.
One Time Password	+ Hasło wykorzystywane jest tylko raz. + Odporne na wiele ataków, m.in. atak metodą powtórzenia.	– W przypadku dostępu do urządzenia z programowym generatorem haseł możliwe jest sklonowanie generatora.	

Tabela 3. Zalety i wady systemów uwierzytelniania – ciąg dalszy

Nazwa elementu	Zalety	Wady	Uwagi
RSA, ECC	+ Brak problemu dystrybucji kluczy.	– Są wolniejsze niż kryptografia symetryczna. – Klucz publiczny nie jest uwierzytelniony. – Wymaga większej mocy obliczeniowej niż kryptografia symetryczna.	
Certyfikat (PKI)	+ Zapewnia uwierzytelnianie klucza publicznego. + Zapewnia niezaprzeczalność. + Łatwe zarządzanie tożsamościami.	– Jest wolniejszy niż kryptografia symetryczna. – Wymagana łączność z zaufaną trzecią stroną, np. w celu śledzenia listy odwołanych certyfikatów.	• Wykorzystuje kryptografię asymetryczną.
TPM	+ Umożliwia bezpieczne przechowywanie kluczy kryptograficznych. + Umożliwia sprzętowe generowanie kluczy, liczb losowych.	– Wymagany jest specjalny moduł. – Uszkodzenie modułu uniemożliwia uwierzytelnianie urządzenia.	
PUF	+ Nieklonowalny. + Nie jest podatny na ataki fizyczne.	– Wymagany jest specjalny moduł. – Mogą mieć wysoki współczynnik błędnych bitów.	• Zalecany jest wybór silnych PUF.
Numer seryjny	Brak.	– Wymagany dodatkowy moduł. – Nie chroni przed podmianą urządzenia.	

Jeżeli jest to tylko możliwe, powinno się stosować systemy uwierzytelniania wykorzystujące identyfikator wraz z dodatkowym elementem do potwierdzania swojej tożsamości. Dzięki temu zyskujemy większą pewność co do tożsamości urządzenia.

6. Podsumowanie

W artykule przedstawiono sposoby uwierzytelniania urządzeń w sieciach Internetu Rzeczy. Na początku opisano dwa modele architektury systemów Internetu Rzeczy. Na podstawie modelu architektury łatwiej jest zidentyfikować problemy oraz zagrożenia występujące w systemach uwierzytelniania urządzeń Internetu Rzeczy. Odpowiednia identyfikacja zagrożeń jest jednym z podstawowych elementów analizy ryzyka w procesie projektowania systemów. Na podstawie architektury czterowarstwowej przedstawiono ataki i problemy, które mogą stanowić zagrożenie dla systemów uwierzytelniania urządzeń w sieciach Internetu Rzeczy. W artykule przedstawiono również podstawowe wymagania względem identyfikatora oraz samego systemu uwierzytelniania względem wymagań urządzeń Internetu Rzeczy. W ostatniej części przedstawiono, jakie właściwości oraz metody są stosowane w obecnych systemach uwierzytelniania. Przedstawiono także ich zalety i wady. Dla przytoczonych systemów uwierzytelniania opisano, na jakie ataki dany schemat uwierzytelniania jest podatny, a na jakie nie.

Dane uwierzytelniające przechowywane są najczęściej w bazie danych lub pliku. Obecnie pojawia się wiele artykułów wykorzystujących rejestry rozproszone [12], [13]. Do zalet rejestrów rozproszonych należą m.in. decentralizacja, niezmienność przechowywanych danych, a także replikacja danych pomiędzy węzłami. Jak przedstawiono we wstępie, intensywny rozwój Internetu Rzeczy zmusza użytkowników do stosowania skutecznych i bezpiecznych systemów uwierzytelniania. Dlatego istotne są dalsze badania nad systemami uwierzytelniania, a także protokołami kryptograficznymi dla tego typu urządzeń.

Literatura

[1] ALI I., SABIR S., ULLAH Z., *Internet of Things Security, Device Authentication and Access Control: A Review*. International Journal of Computer Science and Information Security, Vol. 14, No. 8, 2016, pp. 456-466.

[2] ALMULHIM M., ZAMAN N., *Proposing secure and lightweight authentication scheme for IoT based E-health applications*. 2018 20th International Conference on Advanced Communication Technology (ICACT), 2018, pp. 481-487.

[3] BRAEKEN A., *PUF Based Authentication Protocol for IoT*. Symmetry 2018, 10 (8), 352, 2018, pp. 1-15.

[4] BURHAN M., REHMAN R., KHAN B., BYUNG-SEO K., *IoT Elements, Layered Architectures and Security Issues: A comprehensive Survey*. Sensors 2018, 2018, pp. 1-37.

[5] DAVIES R., *The Internet of Things – Opportunities and challenges*. European Parliamentary Research Service, 2015, pp. 1-8.

[6] DEOGIRIKAR J., VIDHATE A., *Security Attacks in IoT: A Survey*. International Conference on I-SMAC, 2017, pp. 32-37.

[7] FERRAG M. I INNI, *Authentication Protocols for Internet of Things: A Comprehensive Survey*. Hindawi, Security and Communication Networks, 2017, ID 6562953, pp. 1-41.

[8] FURTAK J., ZIELIŃSKI Z., CHUDZIKIEWICZ J., *Procedures for sensor nodes operation in the secured domain*. Concurrency and Computation: Practice and Experience, 2019, e5183, pp. 1-13.

[9] HASAN A., QUERSHI K., *Internet of Things Device Authentication Scheme using Hardware Serialization*. 2018 International Conference on Applied and Engineering Mathematics, 2018, pp. 109-114.

[10] KOO J., OH S.-R., KIM Y.-G., *Device Identification Interoperability in Heterogeneous IoT Platforms*. Sensors 2019, 2019, pp. 1-16.

[11] KOTHMAYR T. I INNI, *A DTLS Based End-To-End Security Architecture for the Internet of Things with Two-Way Authentication*. Local Computer Networks Workshops, 2012, pp. 956-963.

[12] LAU C., YEUNG A., YAN F., *Blockchain-based Authentication in IoT Networks*. 2018, IEEE Conference on Dependable and Secure Computing (DSC), 2018, pp. 1-8.

[13] LEE C., KIM K., *Implementation of IoT System using BlockChain with Authentication and Data Protection*. 2018 International Conference on Information Networking (ICOIN), 2018, pp. 936-940.

[14] MAES R., VERBAUWHEDE I., *Physically Unclonable Functions: a Study on the State of the Art and Future Research Directions*. Towards Hardware-Intrinsic Security, 2010, pp. 1-37.

[15] MEIDAN Y. I INNI, *ProfilIoT: A Machine Learning Approach for IoT Device Identification Based on Network Traffic Analysis*. SAC‘17 Proceedings of the Symposium on Applied Computing, 2017, pp. 506-509.

[16] MUKHOPADHYPAY D., *PUFs as Promising Tools for Security in Internet of Things*. IEEE Design & Test, Volume 33, Issue 3, 2016, pp. 103-115.

[17] NICANFAR H., JOKAR P., LEUNG V., *Smart Grid Authentication and Key Management for Unicast and Multicast Communications*. 2011 IEEE PES Innovative Smart Grid Technologies, 2011, https://ieeexplore.ieee.org/document/6167151

[18] NING H., LIU H., LIU Q., JI G., *Directed Path Based Authentication Scheme for the Internet of Things*. Journal of Universal Computer Science, Vol. 18, No. 9, 2012, pp. 1112-1131.

[19] RABIAH A., RAMAKRISHNAN K., LIRI E., KAR K., *A Lightweight Authentication and Key Exchange Protocol for IoT*. Workshop on Decentralized IoT Security and Standards 2018, 2018, pp. 1-6.

[20] SCHIMTT C., NOACK M., STILLER B., *TinyTO: Two-way Authentication for Constrained Devices in the Internet-of-Things*. Internet of Things, 2015, pp. 239-258

[21] SCHUKAT M., CARTIJO P., *Public key infrastructures and digital certificates for the Internet of things*. 2015 26th Irish Signals and Systems Conference (ISSC), 2015, https://ieeexplore.ieee.org/abstract/document/7163785

[22] SETHI P., SARANGI S.R., *Internet of Things: Architectures, Protocols, and Applications*. Journal of Electrical and Computer Engineering, 2017, pp. 1-25.

[23] SHAH T., VENKATESAN S., *Authentication of IoT Device and IoT Server Using Security Vaults*. 2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications, 2017, pp. 819-824.

[24] SHIVRAJ V., RAJAN M., SINGH M., BALAMURALIDHAR P., *One time password authentication scheme based on elliptic curves for Internet of Things (IoT)*. 2015 5th National Symposium on Information Technology: Towards New Smart World (NSITNSW), 2015, pp. 1-6.

[25] SPIESS P. I INNI, *SOA-based Integration of the Internet of Things in Enterprise Services*. 2009 IEEE International Conference on Web Services, 2009, pp. 968-975.

[26] STEWART J.M., *CompTIA Security+ Review Guide*. Sybex, Indianapolis, 2014.

[27] TASALI Q., CHOWDHURY C., VASSERMAN E., *A Flexible Authorization Architecture for Systems of Interoperable Medical Devices*. SACMAT'17, 2017, pp. 9-20.

[28] TORKAMAN A., SEYYEDI M.A., *Analyzing IoT References Architecture Model*. International Journal of Computer Science and Software Engineering, Vol. 5, Issue 8, August 2016, pp. 154-160.

[29] TRIPATHY B.K., ANURADHA J., *Internet of Things (IoT) Technologies, Applications, Challenges and Solutions*. CRC Press, Boca Raton, 2017.

[30] TRNKA M., CERNY T., STICKNEY N., *Survey of Authentication and Authorization for the Internet of Things*. Hindawi, Security and Communication Networks, 2018, ID 4351603, pp. 1-17.

[31] WANG K.H., CHEN C.M., FANG W., WU T.Y., *A secure authentication scheme for Internet of Things*. Pervasive and Mobile Computing 42, 2017, pp. 15-26.

[32] WON J., SINGLA A., BERTINO E., BOLLELLA G., *Decentralized Public Key Infrastructure for Internet-of-Things*. Milcom, 2018 Track 5, 2018, pp. 1-7.

Źródła elektroniczne

[33] ALUTHGE N., *IoT device fingerprinting with sequence-based features*, 2017, https://helda.helsinki.fi/handle/10138/234247 (dostęp 12.05.2019)

[34] *An overview of the IoT Security Market Report 2017-2022*, https://iiot-world.com/reports/an-overview-of-the-iot-security-market-report-2017-2022/ (dostęp 12.05.2019)

[35] *Connecting applications, devices and gateways*, IBM, https://www.ibm.com/support/knowledgecenter/en/SSQP8H/iot/platform/reference/security/connect_devices_apps_gw.html (dostęp 12.05.2019)

[36] ECC 101: What is ECC and why would I want to use it?, https://www.globalsign.com/en/blog/elliptic-curve-cryptography/ (dostęp 20.06.2019)

[37] *Gartner Identifies Top 10 Strategic IoT Technologies and Trends*, https://www.gartner.com/en/newsroom/press-releases/2018-11-07-gartner-identifies-top-10-strategic-iot-technologies-and-trends, 2018 (dostęp 12.05.2019)

[38] *Identifiers in Internet of Things*, Alliance for Internet of Things Innovation, Version 1.0, 2018, https://aioti.eu/wp-content/uploads/2018/03/AIOTI-Identifiers_in_IoT-1_0.pdf.pdf, (dostęp 10.05.2019)

[39] Series Y: Global Information Infrastructure, Internet Protocol Aspects and Next-Generation Network. Overview of the Internet of things, TELECOMMUNICATION STANDARIZATION SECTOR OD ITU, 2012, https://www.itu.int/rec/dologin_pub.asp?lang=e&id=T-REC-Y.2060-201206-I!!PDF-E&type=items (dostęp 12.05.2019)

[40] *The Internet of Things (IoT) – Threats and Countermeasures*, https://www.cso.com.au/article/575407/internet-things-iot-threats-countermeasures/ (dostęp 12.05.2019)

[41] *The Internet of Things Reference Model*, Cisco 2014, http://cdn.iotwf.com/resources/71/IoT_Reference_Model_White_Paper_June_4_2014.pdf (dostęp 10.05.2019)

[42] Top 10 IoT security challenges, https://developer.ibm.com/articles/iot-top-10-iot-security-challenges/, 2017 (dostęp 12.05.2019)

Device authentication methods in Internet of Things networks

ABSTRACT: The paper describes basic requirements for authentication systems used in Internet of Things networks, along with problems and attacks that may hinder or even prevent the process of authentication. The methods currently used in device authentication are also presented.

KEYWORDS: Internet of Things, IoT, device authentication, device identification

Praca wpłynęła do redakcji: 12.07.2019 r.

TELEINFORMATICS REVIEW
No. 3-4, 2019

Risk of undesired changes to significant information quality criteria

Krzysztof LIDERMAN

Institute of Teleinformatics and Cybersecurity, Faculty of Cybernetics, MUT
ul. gen. Sylwestra Kaliskiego 2, 00-908 Warsaw, Poland
krzysztof.liderman@wat.edu.pl

ABSTRACT: The article presents a method of estimating the risk of an undesirable change in the information quality criterion of secrecy, meaning estimating the risk of a certain class of information security incidents. The qualitative risk estimation method is adopted and the impact of a descriptive grade composition method on the results is discussed. Considerations on the possibilities of interpreting variables used in the risk estimation and establishing the range of their actual values were also presented. Additionally, the paper describes how the identified range of actual variable values translates into levels used in the risk estimation.

KEYWORDS: incident, risk estimation, information security

1. Introduction

The problem presented herein concerns the estimation of the risk of a specific class of information security incidents. In this article, an information security incident means an event or a series of events (resulting in threat execution) that causes or may cause an undesired change in the value of significant information quality criteria[1]. The issue of the incident and its

[1] Only events that have caused an undesired change in the value of significant information quality criteria are considered in a risk analysis (also in this article). This means that attacks that were stopped by IPS and did not cause damage are not taken into consideration. However, according to the rules of the art, such events are also classified as incidents by IT specialists and security departments.

handling is described in paper [1], among others. Since 2018, the perception and handling of incidents has been significantly impacted by the Act on the National Cyber Security System [10], with six regulations assigned to it, of which [6] and [7] are most important from the perspective of the subject matter described herein. The said Act and the accompanying regulations are an implementation under Polish law of the EU NIS Directive (*Network and Information Systems Directive* [3]). The incident occurrence and handling process can be illustrated through the so-called ICOM **(I**nput, **C**ontrol **O**utput, **M**echanism) cube – see Figure 1.

Which of the information quality criteria, mentioned in the title, from among the elementary criteria are significant to the information resources of a given organisation and what their required values are should be:

- determined in a risk analysis,
- approved by the organisation's management,
- entered into the appropriate documents, such as the security policy.

Secrecy, integrity and availability are usually the basic set of criteria from which significant criteria are selected. Further criteria are also those regarding actions on information resources, such as accountability, non-repudiation, etc. The above criteria set out the basic classes of incidents - this article presents the issue of estimating the risk of undesired changes in the value of one of these elementary criteria - secrecy.

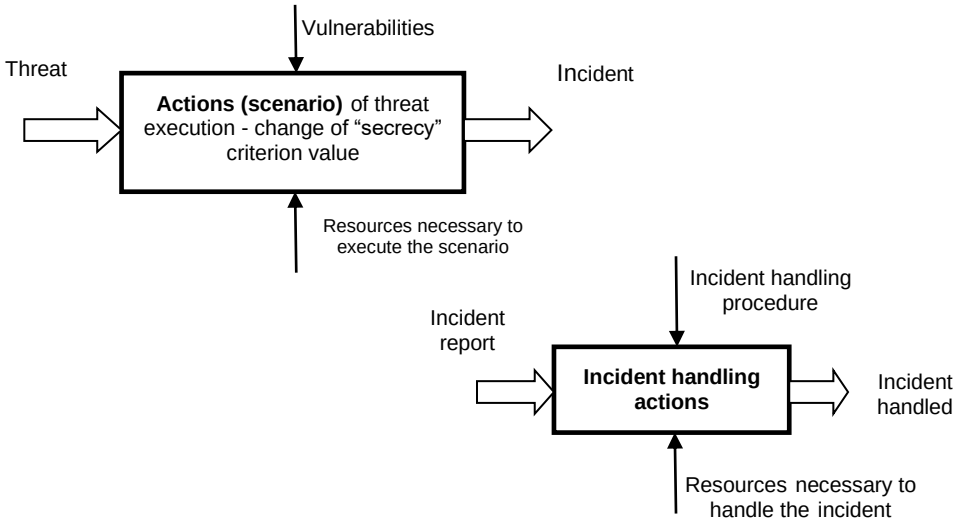


Fig. 1. Incident and incident handling

Secrecy indicates the required level (strength) of information resource protection against any information being obtained about these resources in an unauthorised manner. This level is agreed by the entities that exchange information. An example of the issue of information resource secrecy in legal frameworks is the Act [9] and the Regulation [8], where the value set of the security level is established as:

$$\{top\_secret, secret, confidential, classified\}$$

2. Descriptive grades in risk estimation and grade composition

The existence of a risk of an information security incident for (any) information resource means the occurrence of a threat affecting the significant information quality criteria (secrecy, integrity, availability, etc.) for that resource; the magnitude of the risk is established by combining the value the assessment of the possibility of events caused by the threat, and the value of the assessment of damage resulting from these events, i.e. the incident effects.

Risk estimation consists in determining the possibility of threat execution (familiarity is also needed for this purpose, as shown in Figure 1, vulnerabilities) and potential losses. Such estimation includes two basic parameters: the possibility or probability of threat execution and the measurable effects of such an event. In risk estimation, the choice is basically limited to two methods [1]:

1. **Quantitative** risk analysis method, where a random event measure is applied – probability expressed by a number in the range of [0, 1].
2. **Qualitative** risk analysis method, which uses descriptive, arbitrarily selected measures expressing the possibility of the event occurring.

This article assumes that risk estimation is carried out using a qualitative method. The detailed assumptions are as follows:

- information resource $z_i \in Z$, where $Z=\{z_i|$ and$\in[1, n]\}$ is a set of resources subject to risk analysis;
- z_i has vulnerabilities $p_j \in P$, where $P=\{p_j| j\in[1, m]\}$ is a set of vulnerabilities of the resources in set Z;
- vulnerability can be used by threat $d_k \in D$, where $D=\{d_k| k\in[1, r]\}$ is a set of identified threats that may affect resources Z.
- risk analysis will be carried out in the resource variant.

The following should be specified:

1) Uniform symbolic grades for properties (attributes) of: threats $d_k \in D$, vulnerability $p_j \in P$, damage, losses and risk. These features describing risk components are as follows:
   - possibility of threat execution, hereinafter referred to as PTE,
   - degree of vulnerability, hereinafter referred to as DV,
   - loss value, hereinafter referred to as LV,
   - magnitude of risk, hereinafter referred to as RISK.
2) The method for assigning grades and calculating generalised grades.

It is proposed to adopt the following system K for assigning descriptive grades from the GRADE set to selected features in the FEATURE set:

$$K = \langle \text{FEATURE, GRADE, } \textit{Procedure} \rangle$$

where:

- FEATURE – a set of features: PTE, DV, LV, RISK.
- GRADE – a set of descriptive grades. This article assumes that it is a three-element set {L, M, H}, where:

  H – HIGH possibility, degree or loss,

  M – MEDIUM possibility, degree or loss,

  L – LOW possibility, degree or loss.
- *Procedure* – a method for assigning descriptive grades from the GRADE set to features of the FEATURE set (e.g. by a decision of experts after a "brainstorming session").

The following method for descriptive grade composition (Algorithm 1) is recommended (as stated in [2]):

ALGORITHM 1

Assuming the following:

1) set A of grades in ascending order, i.e. $A=\{q_1, \ldots, q_i, \ldots, q_n\}$ where $i \in N$ is the (position) number of element q_j in set A and $q_j < q_{j+1}$;
2) "element_selected" – any arbitrarily indicated element of set A.
3) number $r \in R$ equals the remainder of the average of position numbers of the selected elements.

   IF $r = 0$ THEN

   $q = q_k$ where k is the average of position numbers of elements selected from set A

   ELSE ($r \neq 0$):

   IF $r \geq 0.5$ THEN

$q = q_{\lceil k \rceil}$

ELSE ($r < 0.5$)

$q = q_{\lfloor k \rfloor}$

END OF ALGORITHM 1

In the presented algorithm, the *floor*, *integer part*, *feature* or *entier* of real number *x*, marked as $\lfloor x \rfloor$, is the largest integer not greater than *x*. However, the *ceiling* or *upper feature* of real number *x* is the smallest integer not less than *x*, marked as $\lceil x \rceil$.

EXAMPLE 1

The symbol ☼ means the descriptive grade composition operation as per Algorithm 1. For example, if the following descriptive grade values are adopted:

N (negligible), L (low), M (medium), H (high), C (catastrophic), i.e.

$A=\{N, L, M, H, C\}$

Where $N< L< M< H< C$ i.e. $q_1=N$, $q_2=L$, $q_3=M$, $q_4=H$, $q_5=C$,

then for q=N☼C☼L☼C:

$(1+5+2+5)/4=3.\mathbf{25}$, i.e. $r < 0.5$,

therefore: $q_1 ☼ q_5 ☼ q_2 ☼ q_5 = q_{\lfloor 3.25 \rfloor} = q_3$ i.e. N☼C☼L☼C=M

******** (end of example)

The following should be done to estimate the risk of losses caused by an incident, for resource $z_i \in Z$, specific threat $d_k \in D$ and vulnerability $p_j \in P$ (Algorithm 2):

ALGORITHM 2

Assuming that *grade*: FEATURE→GRADE

Using the grades from the set {L, M, H} estimate the following:

1) possibility of threat execution (PTE) "as such"[2] for $d_k \in D$, i.e. give the *grade* (PTE) value;
2) degree of vulnerability DV $p_j \in P$ for resource $z_i \in Z$, which can be used by the threat, i.e. the *grade* value (DV) should be provided;

[2] The threat's "potentiality" is estimated at this stage. For example, when assessing the possibility of hardware theft from the organisation's office, the bars, locks, alarm systems etc. of the building are not taken into account (this affects the vulnerability to theft, as considered at the next stage); only the fact that the target is located in a district where there are a lot of thieves is considered.

3) as per formula (1) the RISK of occurrence of an event such that threat $d_k \in D$ uses vulnerability $p_j \in P$ to cause damage with loss value LV:

$$grade(\text{RISK}) = grade(\text{PTE}) \divideontimes grade(\text{DV}) \divideontimes grade(\text{LV}) \qquad (1)$$

END OF ALGORITHM 2

The interpretation of elements of Algorithm 2 for the issue of undesired changes in the significant information quality value, i.e. secrecy, is as follows:

- $d_k \in D$ is a threat to the information resource secrecy;
- $p_j \in P$ is the vulnerability to actions that violate the information resource secrecy;
- PTE is the possibility of executing a threat resulting in an undesired change in the value of the "secrecy" quality criterion;
- LV is the value of losses caused by damage relating to an undesired change in the value of the "secrecy" quality criterion;
- RISK is the risk of an incident of an undesired change in the value of the "secrecy" quality criterion of resource $z_i \in Z$.

The next chapter presents considerations on the possibility of determining (indicating and describing) elements of sets D and P as well as variables PTE, DV and LV and assigning values (descriptive grades) to these variables.

3. Setting variables and the values of their descriptive grades

According to the subject of the article, information resource secrecy is subject to security. Before estimating the risks, the meaning of "secrecy" should be clarified and agreed with all stakeholders[3]. As we can see in EXAMPLE 2, this will also have an impact on the magnitude of the estimated effects of threat execution.

EXAMPLE 2

In the sentence *"I cannot give you this document because it is secret"*, the word "secret" may mean one of the following options, depending on the circumstances:

1. The document is "secret" in the common sense of the word, if its owner does not wish to disseminate it for various reasons.

[3] "implicit" would be an equally good term, but following other publications, the term "secrecy" is also used herein.

2. The document is assigned the "secret" clause under the Act on the Protection of Classified Information (JoL. of 2010, No. 182, item 1228).
3. The document is "secret" within the meaning of the regulations in force at the organisation (these might be only internal regulations) on the dissemination of information, for example, when it contains information constituting trade secrets (within the meaning of Art. 11(4) of the Act of 16/04/1993 *"on combating unfair competition"*; JoL.93.47.211).

******** (end of example)

It is also necessary to establish and reach a consensus among stakeholders as to the nature of "secrecy". Is secrecy something indivisible (something is or is not secret, there are no other options) or whether "secrecy" can be somehow graded. The globally prevailing view is that secrecy can be graded. This was significantly influenced by early works in the field of security, made in the 1970s, in particular those by D. Bell and L. La Padula (for example, see the descriptions of the results of these works in Chapter 2 of [1]]). The grading of "secrecy" has also been entered into the Polish Act on *the Protection of Classified Information* [9], which features four values for the "secrecy" criterion. This view is also adopted in this article.

However, such an approach raises some interpretative complications – how should the effects of a secrecy incident be estimated, assuming that grade values are established and possible levels of secrecy are determined? As already mentioned, "secrecy" specifies the required protection strength of information resource security against any information being obtained about these resources in an unauthorised manner. Levels (labels assigned to items) specify sets of requirements assigned to them. Assuming that the SECRET variable can take values from the set[4]:

{secret, confidential, classified}

the phrase "undesired change in the SECRECY information quality criterion" means a change from SECRECY to ~SECRECY[5]. For example, an information resource as a set on a disk array was secured in as per the requirements for secrecy at the "confidential" level (i.e. SECRECY=confidential), but after the threat execution, the condition is not met, i.e. ~(SECRECY=confidential), e.g. it has been demonstrated that the required safeguards can be bypassed or the safeguards preventing unauthorised access to the information have been broken. Does this mean that the intruder can read only information classified as

[4] This three-element set of the SECRET variable is used later in the article: tables and examples.

[5] The ~ symbol is a sentence-forming functor, "is false".

"confidential", but not "classified", or that they can read both "confidential" and "classified" information, classified below, assuming that the set of information classified in such a way is on this same disk as the confidential set? The answer to this question requires an analysis of possible threat scenarios, considering the actual allocation of the secured information resources and applied safeguards. Another practical consequence of adopting the possible grading of "secrecy" is the need to take into account the classification[6] of am information resource when assigning grades as part of risk estimation - which is best represented in Table 5 (last column) and Table 10.

This article does not consider the type of an information resource by its carrier (electronic, paper, microfilm tape, etc.), although such a preliminary classification of resources might be useful for comprehensive and detailed risk analysis. This would systematise the identification of possible methods of threat execution, e.g. by the tools necessary to execute the threat and possible vulnerabilities.

A uniform method for describing threats is also good for practical reasons. Table 1 presents a proposal for such a method. The provisions of standard ISO/IEC WD 29115 [4], for example, can be used when determining the estimated effects and the possibility of threat execution. Although this standard applies to identity and authentication, it includes some guidelines on what to look for when considering the impact of a violation of information resource secrecy. According to the standard, the potential impact of incorrect authentication applies to:

1. Discomfort, trouble or damage to reputation or position.
2. Financial loss or liability.
3. Damage to the entity, its plans or public interests.
4. Leakage of sensitive information or unauthorised access.
5. Personal security,
6. Violations of civil or criminal law.

The strength of each of the above factors is set on a scale of values: low, moderate, significant, high (i.e. the set of grades differs from the one adopted herein, but it does not matter for the considerations here). The organisation is to determine, based on the estimation of the risk specific to the organisation, what their interpretation is, e.g. what level of financial losses is low, moderate, etc. ISO/IEC WD 29115 does not specify how to carry out risk estimation – it can be done as recommended by PN-ISO/IEC 27005:2010 [5].

[6] A classified resource is a resource for which a security class has been determined by specifying the required level of significant security quality criterion and category. If SECRECY is a significant quality criterion, and the set of values for this variable consists of three labels {secret, confidential, classified}, a sample security class can be as follows: ⟨confidential, ABW_documents ⟩.

The situation is different in the case of entities (organisations) subject to the Act on the *National Cyber Security System* ([10], hereinafter referred to as NCS), which states the following:

Art. 6. The Council of Ministers shall determine, by regulation:

1) a list of key services referred to in Art. 5(2)(1) to assign the key service to a given sector, subsector and the type of entity listed in Annex 1 to the Act, and the importance of the service for maintaining critical social or economic activity;
2) **thresholds for the relevance of the disruptive incident effect** on the provision of key services provided in the list of key services, taking into account:
   a) number of users dependent on the key service provided by the entity,
   b) dependence of other sectors, referred to in Annex 1 to the Act, on the service provided by the entity,
   c) impact the incident could have, due to its scale and duration, on economic and social activities or public security,
   d) market share of the key service provider,
   e) geographical scope of the area that could be affected by the incident,
   f) entity's ability to maintain a sufficient level of the key service, taking into account the availability of alternative ways of providing it,
   g) other factors specific to a given sector or subsector, if applicable – in order to provide protection against the threat to human life or health, significant property losses and reduction of the quality of the key service provided.

Tab. 1. Threat description sheet template (example)

SHEET No. ……. DESCRIPTION OF THREAT TO SECRECY OF INFORMATION RESOURCE $z_i \in Z$ **Threat ID:** [threat symbol]	
Threat:	[one-sentence descriptive name of threat, e.g. *actions of an intruder – employee of this organisation*]
Threat execution scenario:	[a few sentences of description in words or a block diagram]
Resource owner:	[identification data]
Possible (estimated) effects/damage if the threat is executed:	[a few sentences of description or a list specification]
Possible (estimated) losses if the threat is executed:	[amount in specified currency or description]
Threat potential:	[a few sentences of description]

To the entities specified in the NCS, Art. 6 shall mean the obligation to describe the incident using the method contained therein. It seems that to improve the handling of incidents on both the national and European scale, the proposed incident description method should also be used by organisations (entities) that are not subject to NCS regulations.

It should also be clarified in preliminary arrangements what may be a threat to the information resource secrecy. In this case[7], it is the so-called human factor, manifested as intentional or erroneous actions. This must be determined primarily to estimate the value of the possibility of threat execution (PTE factor) and other risk factors (DV, LV). This can be done using the following table, for example:

Tab. 2. Threats to the information resource secrecy (example description)

Type of action	Who	Motive
INTENTIONAL	intruder, employee	benefits (financial, ideological, psychological, etc.), revenge, curiosity, blackmail, etc.
ERRONEOUS	employee	none

In the case of risk estimation, the possibility of exposing the organisation and its information resources to the threat execution through intentional actions, requires an organisation description in terms of its attractiveness to the intruder. The description should include factors that affect the intruder's motivation. This can be done by adopting a certain set of features (hereinafter referred to as ZC) as ordered four values describing the organisation from this perspective:

$$ZC = \langle BS, OA, PM, AT \rangle$$

where:

- $BS=\{bs_i | \ i \in [1,m]\}$ is a non-empty set of features describing the organisation's "business" size;
- $OA=\{oa_j | \ j \in [1,n]\}$ is a non-empty set of features describing the area of the organisation's activities;
- $PM=\{pm_k | \ k \in [1,l]\}$ is a non-empty set of features describing the impact of the organisation's activities on public mood;
- $AT=\{at_p | \ p \in [1,r]\}$ is a non-empty set of features describing the industry's "attractiveness" for an intruder.

[7] Unlike threats to the availability of an information resource, where the most common are failures of infrastructure in which the information resource is embedded, and disasters and adverse natural phenomena affecting the infrastructure and the resource itself (such as floods or fires).

These sets of features should be specified by experts or imposed by significant regulations[8] to obtain analysis repeatability. Let's assume that the above feature sets are specified as follows:

- BS = {large, medium, small, micro enterprise};
- OA = {global, local};
- PM = {significant, moderate, low, none};
- AT is set in the predefined Table 3:

Tab. 3. Type of organisation and "attractiveness" for an intruder (example)

Type of organisation	Attractiveness
Telecoms	**big**
Media companies	
Public administration	
….	
Chain stores	**moderate**
Defense industry	
Medicine	
….	
Other	**low**
….	

As already mentioned, determining the set of features (how many and what elements), their specification and assignment of grade values should be done by a group of experts, e.g. through brainstorming, or should refer to known official regulations. It should be noted that the description provided will not apply if the intruder's motivation is revenge or the intention to cause harm to a particular organisation (e.g. the intruder was paid to do so). In such cases, it should be assumed that the PTE value is high. It should also be taken into account that the "attractiveness" of the target is only one of the elements affecting the intruder's motivation. It is certainly reduced by high penalties for this type of crime, the effectiveness of their prosecution and the belief that there are strong safeguards to be broken (although this factor may very well be a motivation for an intruder who likes a challenge).

EXAMPLE 3

For a large media company operating on a global scale, a sample set ZC_p of feature values can be as follows:

[8] The question remains which entity would issue such regulations. Government Security Centre? Ministry of Digitisation?

$$ZC_p = \{large, global, moderate, big\}$$

This set of features describing the organisation should be translated into grades adopted for risk estimation, i.e. an interpretation table similar to Table 4 should be made.

Tab. 4. A set of features describing the organisation, their values and corresponding grades (example)

FEATURE	Possible values of FEATURE variable	Grade
BS	large	H
	medium	M
	small, micro-enterprise	L
OA	global	H
	local	L
PM	significant	H
	moderate	M
	low none	L
AT	"telecoms" class, "chain stores" class	H
	-	M
	"other" class	L

Then the PTE value – the possibility of the organisation being exposed by intentional actions of an intruder – is set by the formula:

$$grade[PTE(ZC_p)] = grade[PTE(\{large, global, moderate, big\})] = PTE(\{H, H, M, H\})$$

where *grade* is a function (usually heuristic) assigning grades from the set of grades (in this article – set {L, M, H}) to the elements of the set of features (here: {large, global, moderate, big}), i.e.

$$grade: PTE(ZC) \rightarrow PTE(GRADE)$$

where: $GRADE=\{grade_i | \text{ and} \in [1,n]\}=\{L, M, H\}$.

Adopting the method for grade composition as per Algorithm 1 in this example results in a high possibility of threat execution:

$$alg(PTE(\{H, H, M, H\}))=PTE(H)$$

******** (end of example)

When estimating the risk of the organisation being exposed to erroneous actions of its employees, historical data should be available regarding the errors that resulted in the incident relating to the information resource secrecy in order to determine the PTE value. Such data, including both the type of error and its frequency, should be collected by the organisation. If there are no such data, the data from organisations of a similar company profile may be used, provided they are available. The third option is to use generalised statistical data published by various organisations involved in information security (e.g. CERT). Naturally, such records include only cases of detected errors and may not be adequate to the actual situation of the organisation for which the risk is estimated.

Errors may result in the disclosure of the information resource content to unauthorised entities, divulging the information about the existence of an undisclosed resource in the system, disclosure of all or some entities authorised to access such a resource, the possibility of leading to said situations by performing an unauthorised operation in the system, etc. If a table of possible effects has been prepared (for example, developed as a result of expert brainstorming), effects should be assigned a frequency of occurrence based on historical data and assigned significant grades based on interpretation tables (e.g. such as Tables 6-9), depending on the classification of the resource affected by the incident. An example of the description is shown in Table 5.

Assuming that disclosure of the content of an information resource classified as *secret* or *confidential*, or divulging of information about the existence of a resource classified as *secret* in the system is not permissible under any circumstances, and specifying the thresholds for the frequency of specific events, interpretation tables of descriptive grades for PTE can be made. This type of assumptions-decisions regarding the thresholds for the frequency of a specific event must be made by the management board of the organisation or its security department. Examples of such descriptions are shown in Tables 6-9. Having considered the contents of the interpretation tables, the last column of Table 5 may be filled in.

Considering the discussion of effects earlier in this chapter (when proposing the threat description), it can be assumed that damage caused by intentional or erroneous actions depend on the following factors:

1. Security class assigned to the information resource (designated *secret*, *confidential* and *classified* in this article).
2. Number of users depending on the resource/service affected by the incident.
3. Dependence of other organisations (or sectors within the meaning of NCS) on the resource/service affected by the incident.

Tab. 5. Effects of errors, empirical data on frequency and PTE grade according to interpretation Tables 6-9 (example)

No.	Effects of a secrecy error	Incident frequency	Grade for PTE (based on Tables 6-9)	
1	Disclosure of the content of an undisclosed resource to unauthorised entities	Once every two years	*secret*	H
			confidential	H
			classified	L
2	Divulging information about the existence of an undisclosed resource in the system	Once every three years	*secret*	H
			confidential	M
			classified	M
3	Disclosure of all or some entities authorised to access an undisclosed resource	Twice a year	*secret*	H
			confidential	M
			classified	L
4	Possibility of situations 1-3 occurring by performing an unauthorised operation in the system	Five times a year	*secret*	H
			confidential	H
			classified	H
5	….	….	….	

Tab. 6. Interpretation of descriptive grades for the possibility of threat execution (PTE) for the error of disclosing the content of an undisclosed information resource to unauthorised entities

GRADE	INTERPRETATION
H	Whenever the error is made for *secret* and *confidential*, more than once a year for *classified*
M	*Classified* once a year
L	*Classified* once every two years

Tab. 7. Interpretation of descriptive grades for the possibility of threat execution (PTE) for the error of divulging information about the existence of an undisclosed resource in the system

GRADE	INTERPRETATION
H	For *secret* whenever the error is made
M	For *confidential* and *classified* whenever the error is made
L	Never

Tab. 8. Interpretation of descriptive grades for the possibility of threat execution (PTE) for the error of disclosure of all or some entities authorised to access an undisclosed resource

GRADE	INTERPRETATION
H	For *secret* regardless of frequency
M	For *confidential* regardless of frequency
L	For *classified* regardless of frequency

Tab. 9. Interpretation of descriptive grades for the possibility of threat execution (PTE) for the error of possibility of situations 1-4 in Tab. 5 occurring by performing an unauthorised operation in the system

GRADE	INTERPRETATION
H	For *secret* regardless of frequency, for *confidential* when once a year or more, for *classified* when more than three times a year
M	For *confidential* when not more than once every two years, for *classified* when two or three times a year
L	For *confidential* when not more than once every three years, for *classified* when not more than once a year

4. Impact the incident could have, due to its scale and duration, on economic and social activities or public security.
5. Market share of the organisation affected by the incident.
6. Geographical scope of the area that could be affected by the incident.
7. Violations of civil or criminal law.
8. Impact on personal security.
9. Impact on the organisation's image.
10. Impact on the organisation's plans or public interests.

Unlike the estimates for the loss of availability of an information resource (see example 3.6 in [1], for example), in case of violation of its secrecy, damage is difficult to translate into losses measured in a particular currency. This leads to complications in determining the content of the interpretation table for losses (LV). Therefore, it is proposed not to include losses to interpret the LV factor in formula (1) as the extent of damage assessed by experts. Assuming that the list of factors affecting the extent of damage is limited to the ten said factors, an interpretation table similar to Table 10 should be developed, using normative and legal guidelines and expert opinions.

The generalised damage value should be estimated using Algorithm 1, following the example shown for PTE in example 3.

Tab. 10. Specification of possible damage (LV) and grade values (example)

No.	Factors affecting the extent of damage	Extent of damage	Grade
1	Security class of the resource affected by the incident	*secret*	H
		confidential	M
		classified	L
2	Number of users affected by the incident	Sector-specific actions of the organisation, according to NCS, for details see Regulation[9]	
3	Dependence of other organisations	as above	
4	Impact on economic and social activities or public security	as above	
5	Market share	as above	
6	Geographical scope of the incident	as above	
7	Violations of civil or criminal law	Assessment by the Legal Department	
8	Impact on personal security	Assessment by the Security Department	
9	Impact on the organisation's image	Assessment by the Management Board	
10	Impact on the organisation's plans or public interests	as above	

There must be a corresponding vulnerability for the threat to be executed. In practice, the set of vulnerabilities is based on the results of the operation of security scanners (detecting vulnerabilities in software and configuration files), penetration tests[10], local inspections, documentation reviews and expert consultations. Then, the degree of vulnerability for all elements of the above set is determined (usually using expert assessments). The results can be presented in tables, as shown in Tables 11 and 12[11].

[9] Regulation of the Council of Ministers of 31/10/2018 *on the thresholds for considering an incident serious* JoL. item 2180.

[10] They should also include testing the staff's resistance to social engineering and resistance to physical security penetration.

[11] Symbols $\wedge$, $\vee$, and $\sim$ are sentence-forming functors "and", "or" and "is false", respectively.

Tab. 11. Interpretation of descriptive grades for vulnerability[12] to intentional actions of an intruder (example)

GRADE	INTERPRETATION
H	~ (safeguards required for "classified", "confidential", "secret" security levels) ∧ ~ (correct safeguard configuration)
M	[~ (safeguards required for "classified", "confidential", "secret" security levels) ∧ (correct safeguard configuration)] ∨ [(safeguards required for "classified", "confidential", "secret" security levels) ∧ ~ (correct safeguard configuration)]
L	(safeguards required for "classified", "confidential", "secret" security levels) ∧ (correct safeguard configuration)

Tab. 12. Interpretation of descriptive grades for vulnerability to erroneous actions of an employee (example)

GRADE	INTERPRETATION
H	~ (proper employee training in information resource security) ∧~(supervision over employee operations)
M	[~ (proper employee training in information resource security) ∧ (supervision over employee operations)] ∨ [(proper employee training in information resource security) ∧ ~(supervision over employee operations)]
L	(proper employee training in information resource security) ∧ (supervision over employee operations)

EXAMPLE 4

In the case of an organisation of the *gov.pl* domain, it was decided to estimate the risk of exposure of its information resources to intentional actions of intruders and errors by employees aimed at violating the secrecy of these resources, resulting in information security incidents. The estimates apply to resources classified as *secret*, *confidential* and *classified*. The organisation authorities, with employees of its Security Department and risk analysis specialists engaged under a contract of mandate[13], determined the following, based on historical data and expert estimates:

1. "Employee error" incidents usually resulted in two effects:
   a) disclosure of the content of a confidential information resource to unauthorised entities;

[12] In this and the next table, the number of vulnerabilities is limited to two. In practice, their number depends on the identification results, and the method for their composition to obtain the interpretation of the grades depend on the knowledge and decisions of a risk analyst or a supporting expert.

[13] Specialists proposed a three-element set of grades: high (H), medium (M), low (L) and grade composition using Algorithm 1.

   b) disclosure of some entities authorised to access a resource classified as "secret".
2. "Intentional action" incidents usually resulted in two effects:
   a) disclosure of all entities authorised to access a resource classified as "confidential".
   b) disclosure of the content of a classified information resource to unauthorised entities.
   c) there were no actions motivated by revenge or ordered actions. It was considered that such actions would also be unlikely in the future.
3. Damage caused by intentional actions of intruders and employee errors, incurred in the past and identified as possible in the future, relate to:
   a) impact on social activities or public security,
   b) violations of civil and/or criminal law,
   c) impact on personal security,
   d) impact on the organisation's image,
   e) impact on the organisation's public interests.

When estimating damage, the security class (KB in Table 13) of the resource affected by the incident was also taken into account. Risk estimators used the Interpretation Table 10. For damage in items a-e in the above lists, the organisation's Legal Department, the Security Department and the Management Board set the grade values as in Table 13. The resultant grade (last column in Table 13) was obtained using Algorithm 1 – the whole damage estimation can be presented as in Table 13.

4. On the basis of local inspections, analysis of documentation and review of safeguard configuration files, the specialists found that all safeguards required for information resources were used, but some of them were misconfigured. In addition, deficiencies were found in the supervision of employee operations on sensitive resources, although the the security training of employees was highly rated. Assuming that these were all the vulnerabilities found, and that the interpretation of grades for DV is given in Tables 11 and 12, the level of vulnerability for both types of incidents is at level M.
5. It was found that in the past, there were cases (errors) of disclosing information resource content classified as *confidential* to unauthorised persons. It was also found that over the past six years, information about who has access to an information resource classified as *secret* was disclosed twice to unauthorised persons due to error by an employee. Tables 6 and 8 were used to estimate the PTE values for these cases.

6. For intentional actions, the possibility of threat execution PTE was estimated based on the set of features ZC (see example 3). The set of feature values was set at {medium, local, significant "telecoms"}, which translates into the set of grades {M, L, H, H}, so the resultant grade of alg{M, L, H, H} = M.
7. The results of the risk estimation are shown in Table 14.

Tab. 13. LV damage estimation (for example 4)

INCIDENT	INCIDENT TYPE	a	b	c	d	e	KB	alg{.}
Intentional actions	disclosure of all entities authorised to access a resource classified as "confidential"	L	L	M	L	L	M	**L**
	disclosure of the content of a classified information resource to unauthorised entities	M	L	L	M	L	L	**L**
Employee error	disclosure of the content of a confidential information resource to unauthorised entities	M	M	L	H	M	M	**M**
	disclosure of some entities authorised to access a resource classified as "secret"	H	H	H	L	M	H	**H**

Tab. 14. Risk estimation (for example 4)

THREAT	INCIDENT TYPE	PTE	DV	LV	RISK
Intentional actions	disclosure of all entities authorised to access a resource classified as "confidential"	M	M	L	**M** (R'_{515})
	disclosure of the content of a classified information resource to unauthorised entities	M	M	L	**M** (R'_{515})
Employee error	disclosure of the content of a classified information resource to unauthorised entities	H	M	M	**M** (R'_{205})
	disclosure of some entities authorised to access a resource classified as "secret"	H	M	H	**H** (R'_{204})

******** (end of example)

4. Methods for descriptive grade composition and risk interpretation

The analyst can choose any formally correct method for grade composition – currently there are no norms, standards or regulations that would explicitly impose or otherwise govern this issue. In many applications (for example, see NIST SP 800-53 [11]), it is recommended to apply the formula max{GRADE}, because of its simplicity and the fact that it is sufficient for many practical problems; it selects the maximum grade from the set of composed grades as the resultant. The disadvantage of this method for grade composition is the migration of resultant grades towards the highest grades, contrary to the formula min{GRADE}, where the resultant grades migrate towards the lowest grade – this issue is presented in Table 15.

Assuming that:

$$\otimes = \max\{grade_1, \ldots, grade_j, \ldots, grade_k\} \quad \text{for } j\in[1,k] \text{ where: } grade_j \in GRADE \quad (3)$$

$$\varnothing = \min\{grade_1, \ldots, grade_j, \ldots, grade_k\} \quad \text{for } j\in[1,k] \text{ where: } grade_j \in GRADE \quad (4)$$

$$\sun = alg\{grade_1, \ldots, grade_j, \ldots, grade_k\} \quad \text{for } j\in[1,k] \text{ where: } grade_j \in GRADE \quad (5)$$

where:

- alg{...} means the grade composition as per Algorithm 1;
- $\otimes$, $\varnothing$, $\sun$ are symbols for descriptive grade composition as per specific formulas or algorithms.

Tab. 15. Resultant values for composition of two descriptive grades using different methods

No.	A	B	C=A⊗/∅B
1	H	H	**H** H
2	H	M	**H** M
3	H	L	**H** **L**
4	M	H	**H** M
5	M	M	M M
6	M	L	M **L**
7	L	H	**H** **L**
8	L	M	M **L**
9	L	L	L **L**

Tab. 16. Estimations of the risk value using composition operations ⊗ (column 5) **and composition operations ☼** (column 6)

No.	PTE	DV	LV	RISK ⊗	RISK' ☼
1	2	3	4	5	6
1	H	H	H M L	$R_{101} \rightarrow H$ $R_{102} \rightarrow H$ $R_{103} \rightarrow H$	$R'_{101} \rightarrow H$ $R'_{102} \rightarrow H$ $R'_{103} \rightarrow M$
2	H	M	H M L	$R_{204} \rightarrow H$ $R_{205} \rightarrow H$ $R_{206} \rightarrow H$	$R'_{204} \rightarrow H$ $R'_{205} \rightarrow M$ $R'_{206} \rightarrow M$
3	H	L	H M L	$R_{307} \rightarrow H$ $R_{308} \rightarrow H$ $R_{309} \rightarrow H$	$R'_{307} \rightarrow M$ $R'_{308} \rightarrow M$ $R'_{309} \rightarrow M$
4	M	H	H M L	$R_{410} \rightarrow H$ $R_{411} \rightarrow H$ $R_{412} \rightarrow H$	$R'_{410} \rightarrow H$ $R'_{411} \rightarrow M$ $R'_{412} \rightarrow M$
5	M	M	H M L	$R_{513} \rightarrow H$ $R_{514} \rightarrow M$ $R_{515} \rightarrow M$	$R'_{513} \rightarrow M$ $R'_{514} \rightarrow M$ $R'_{515} \rightarrow M$
6	M	L	H M L	$R_{616} \rightarrow H$ $R_{617} \rightarrow M$ $R_{618} \rightarrow M$	$R'_{616} \rightarrow M$ $R'_{617} \rightarrow M$ $R'_{618} \rightarrow L$
7	L	H	H M L	$R_{719} \rightarrow H$ $R_{720} \rightarrow H$ $R_{721} \rightarrow H$	$R'_{719} \rightarrow M$ $R'_{720} \rightarrow M$ $R'_{721} \rightarrow M$
8	L	M	H M L	$R_{822} \rightarrow H$ $R_{823} \rightarrow M$ $R_{824} \rightarrow M$	$R'_{822} \rightarrow M$ $R'_{823} \rightarrow M$ $R'_{824} \rightarrow L$
9	L	L	H M L	$R_{925} \rightarrow H$ $R_{926} \rightarrow M$ $R_{927} \rightarrow L$	$R'_{925} \rightarrow M$ $R'_{926} \rightarrow L$ $R'_{927} \rightarrow L$

Comments on Table 16:

1. The number yy in subscript R_{xyy} is the ordinal number of risk.
2. The number x in subscript R_{xyy} is the row number in Table 16.
3. The apostrophe in symbol R'_{xyy} means that the risk was estimated using operation ☼. The absence of an apostrophe means that the risk was estimated using operation ⊗.

Further considerations on risk estimation refer to the results obtained by applying Algorithm 1 (see column 6 in Table 16). The following conclusions can be drawn from Table 16 regarding the theoretical risk minimisation options:

1. There are <u>three</u> options for minimising the risk value (through the impact on PTE, DV and LV) for the risk set:

$$\{R'_{101}, R'_{102}, R'_{204}, R'_{205}, R'_{410}, R'_{411}, R'_{513}, R'_{514}\}$$

2. Only <u>two</u> options for minimising the risk value exist for risk sets:
   - PTE and LV minimisation: $\{R'_{307}, R'_{308}, R'_{616}, R'_{617}\}$

- PTE and DV minimisation: $\{R'_{103}, R'_{206}, R'_{412}, R'_{515}\}$
- LV and DV minimisation: $\{R'_{719}, R'_{720}, R'_{822}, R'_{823}\}$

3. Only <u>one</u> option for minimising the risk value exist for risk sets:
   - PTE minimisation: $\{R'_{309}, R'_{618}\}$
   - DV minimisation: $\{R'_{721}, R'_{824}\}$
   - LV minimisation: $\{R'_{925}, R'_{926}\}$
4. <u>No</u> options to minimise the risk (the values of all components, i.e. PTE, DV and LV, are at a low level, meaning the risk is minimal) exist for $\{R'_{927}\}$

Assuming an acceptable risk value L, the set of acceptable risk for the grade composition method consists of the following: $\{R'_{618}, R'_{824}, R'_{926}, R'_{927}\}$. However, for elements $\{R'_{618}, R'_{824}, R'_{926}\}$, there are also potential options to reduce the risk value – see the shaded values in item 3. This situation does not occur when estimating the risk value as per formula max{.} – in this case there is only one acceptable risk, in which the value of all components is L (R_{927} in Table 16).

Sometimes, to make decisions on how to minimise risk, it is necessary to know what influences the increase in risk or, looking at the issue differently, which elements composing the risk are at a low level and can be ignored. And so, based on Table 16, it can be concluded that:

1. The set of risk values for which the possibility of threat execution is low, i.e. grade{PTE}=L, is composed of nine elements:

$$\{R'_{719}, R'_{720}, R'_{721}, R'_{822}, R'_{823}, R'_{824}, R'_{925}, R'_{926}, R'_{927}\}$$

2. The set of risk values for which the level of vulnerability is low, i.e. grade{DV}=L, is composed of nine elements:

$$\{R'_{307}, R'_{308}, R'_{309}, R'_{616}, R'_{617}, R'_{618}, R'_{925}, R'_{926}, R'_{927}\}$$

3. The set of risk values for which the level of damage is low, i.e. grade{LV}=L, is composed of nine elements:

$$\{R'_{103}, R'_{206}, R'_{309}, R'_{412}, R'_{515}, R'_{618}, R'_{721}, R'_{824}, R'_{927}\}$$

<u>EXAMPLE 5</u>

Referring the considerations of this chapter to EXAMPLE 4 (see Table 14, last column), the following risk minimisation methods can be recommended for identified incidents:

1. Incident: *Disclosure of all entities authorised to access a resource classified as "confidential"* – **R'_{515}**, **PTE and DV minimisation**.

   <u>For PTE</u>: it is impossible to reduce the target's attractiveness for the intruder. Only the intruder's <u>motivation</u> can be weakened by setting high penalties and

through effective prosecution of this type of crime, but such undertakings are beyond the scope of ordinary organisations – they require action by government administration and (usually) changes in the law.

For DV: EXAMPLE 4 shows (see item 4 of the example – findings of specialists) that the vulnerability found was *safeguard misconfiguration*. Recommended actions: **improve safeguard configuration**.

2. Incident: *Disclosure of the content of a classified information resource to unauthorised entities* – **R'$_{515}$**, **PTE and DV minimisation**.

   For PTE: comment as in item 1.

   For DV: comment as in item 1.

3. Incident: *Disclosure of the content of a classified information resource to unauthorised entities* – **R'$_{205}$, PTE, DV and LV minimisation.**

   For PTE: improve the control system for operations on sensitive resources, improve training for persons who have access to sensitive information (despite being rated as good!), verify the rules for allowing employees to work with sensitive information.

   For DV: EXAMPLE 4 shows (see item 4 of the example – findings of specialists) that the vulnerability found was a *lack of proper supervision over employee operations on sensitive resources*. Therefore, the recommended action: **improve supervision of employee operations on sensitive resources**.

   For LV: the incident affects *social activities and public security, violation of civil and/or criminal law, organisation's public interests.* Minimising these damages requires coordinated actions by the organisation's Management Board, its lawyers and people responsible for PR.

4. Incident: *Disclosure of some entities authorised to access a resource classified as "secret"* – **R'$_{204}$**, **PTE, DV and LV minimisation**.

   For PTE: comment as in item 3.

   For DV: comment as in item 3.

   For LV: the incident affects *social activities and public security, violation of civil and/or criminal law, personal security and organisation's public interests.* Minimising these damages requires coordinated actions by the organisation's Management Board, its lawyers and people responsible for PR. In addition, the organisation's Security Department should provide personal security to those who have access to information classified as "secret".

******** (end of example)

5. Conclusion

The article presents a method of estimating the risk of an undesirable change in the information quality criterion of secrecy, meaning estimating the risk of a certain class of information security incidents. Knowledge about risk, its value, the value of components and their practical relevance (interpretation) is the basis for both building a security system (risk minimisation) and actions related to the handling of incidents caused by the execution of threats for which the risk was estimated.

In the case of risk estimation (or more broadly – risk analysis) of information security for specific organisations, the estimates usually apply to secrecy, integrity and availability of information resources. This article describes a proposal for such an estimate for the secrecy of information resources. An example of risk estimation for availability is shown in Chapter 3.4 in [1]. Risk estimations regarding the integrity of an information resource will be the subject of a separate article.

Literature

[1] LIDERMAN K.: *Bezpieczeństwo informacyjne*. *Nowe wyzwania*. PWN, 2017.

[2] MALIK A., *Propozycja doboru i składania ocen opisowych w jakościowym szacowaniu ryzyka systemów informacyjnych*. Praca dyplomowa. Politechnika Warszawska. Podyplomowe Studium Bezpieczeństwa Systemów Informatycznych. 2011.

[3] Dyrektywa Parlamentu Europejskiego i Rady (UE) 2016/1148 z dn. 6.07.2016 r. *w sprawie środków na rzecz wysokiego wspólnego poziomu bezpieczeństwa sieci i systemów informatycznych na terytorium Unii* (NIS).

[4] ISO/IEC WD 29115:2019 Information technology – Security techniques – *Entity authentication assurance Framework*.

[5] PN-ISO/IEC 27005:2010 – Technika informatyczna – Techniki bezpieczeństwa – *Zarządzanie ryzykiem w bezpieczeństwie informacji*.

[6] Rozp. Rady Ministrów z dn. 31 października 2018 r. *w sprawie progów uznania incydentu za poważny*. Dz. U. poz. 2180.

[7] Rozporządzenie Rady Ministrów z dn. 11 września 2018 r. *w sprawie wykazu usług kluczowych oraz progów istotności skutku zakłócającego incydentu dla świadczenia usług kluczowych*. Dz. U. poz. 1806.

[8] Rozporządzenie Prezesa Rady Ministrów z dn. 20 lipca 2011 r. *w sprawie podstawowych wymagań bezpieczeństwa teleinformatycznego.* Dz. U. z 2011 r. nr 159, poz. 948.

[9] Ustawa z dn. 2.08.2010 r. *o ochronie informacji niejawnej*. Dz. U. 182/10, poz. 1228.

[10] Ustawa z dn. 05.07.2018 r. *o krajowym systemie cyberbezpieczeństwa.* Dz. U. 2018, poz. 1560.

[11] SP-800-53 Rev.4: *Recommended Security Controls for Federal Information System*. April 2013.

Ryzyko niepożądanej zmiany istotnych kryteriów jakości informacji

STRESZCZENIE: W artykule przedstawiono sposób szacowania ryzyka niepożądanej zmiany kryterium jakości informacji, jakim jest tajność, czyli szacowania ryzyka wystąpienia pewnej klasy incydentów z zakresu bezpieczeństwa informacyjnego. Przyjęto jakościową metodę szacowania ryzyka i przedyskutowano wpływ wyboru metody składania ocen opisowych na uzyskane wyniki. Przedstawiono także rozważania na temat możliwości interpretacji zmiennych użytych w szacowaniu ryzyka oraz ustalenia zakresu ich rzeczywistych wartości. Opisano także, jak zidentyfikowany zakres rzeczywistych wartości tych zmiennych przełożyć na oceny użyte w szacowaniu ryzyka.

SŁOWA KLUCZOWE: incydent, szacowanie ryzyka, bezpieczeństwo informacyjne

Received by the editorial staff on: 12.11.2019

PRZEGLĄD TELEINFORMATYCZNY
NR 3-4, 2019

Ryzyko niepożądanej zmiany istotnych kryteriów jakości informacji

Krzysztof LIDERMAN

Instytut Teleinformatyki i Cyberbezpieczeństwa, Wydział Cybernetyki, WAT
ul. gen. Sylwestra Kaliskiego 2, 00-908 Warszawa,
krzysztof.liderman@wat.edu.pl

STRESZCZENIE: W artykule przedstawiono sposób szacowania ryzyka niepożądanej zmiany kryterium jakości informacji, jakim jest tajność, czyli szacowania ryzyka wystąpienia pewnej klasy incydentów z zakresu bezpieczeństwa informacyjnego. Przyjęto jakościową metodę szacowania ryzyka i przedyskutowano wpływ wyboru metody składania ocen opisowych na uzyskane wyniki. Przedstawiono także rozważania na temat możliwości interpretacji zmiennych użytych w szacowaniu ryzyka oraz ustalenia zakresu ich rzeczywistych wartości. Opisano także, jak zidentyfikowany zakres rzeczywistych wartości tych zmiennych przełożyć na oceny użyte w szacowaniu ryzyka.

SŁOWA KLUCZOWE: incydent, szacowanie ryzyka, bezpieczeństwo informacyjne

1. Wstęp

Przedstawiony w tym artykule problem dotyczy szacowania ryzyka wystąpienia pewnej klasy incydentów z zakresu bezpieczeństwa informacyjnego. Incydentem z zakresu bezpieczeństwa informacyjnego nazywa się tu zdarzenie lub ciąg zdarzeń (składających się na realizację zagrożenia), które powoduje lub może spowodować niepożądaną zmianę wartości istotnych kryteriów jakości informacji[1]. Zagadnienie incydentu oraz jego obsługi jest

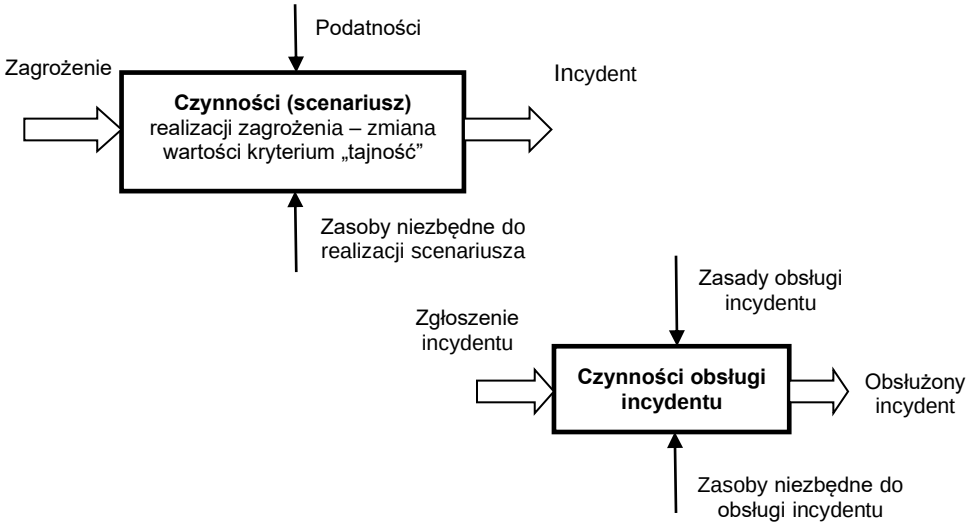
[1] Przy analizie ryzyka rozważane są tylko takie zdarzenia (także w tym artykule), które spowodowały niepożądaną zmianę wartości istotnych kryteriów jakości informacji. Oznacza to, że nie są brane pod uwagę np. ataki, które zostały powstrzymane przez IPS-y i nie spowodowały szkód. Jednak zgodnie z regułami sztuki, takie zdarzenia są przez informatyków i działy bezpieczeństwa także klasyfikowane jako incydenty.

opisane m.in. w pracy [1]. Od roku 2018 istotny wpływ na sposób postrzegania i obsługę incydentów ma ustawa o krajowym systemie cyberbezpieczeństwa [10] wraz z przypisanymi do niej sześcioma rozporządzeniami, z których najważniejsze z perspektywy tematyki opisywanej w tym artykule są [6] i [7]. Wspomniana ustawa i towarzyszące jej rozporządzenia są implementacją na gruncie prawa polskiego unijnej dyrektywy NIS (*Network and Information Systems Directive* [3]). Proces zarówno powstawania incydentu, jak i jego obsługi można zobrazować za pomocą tzw. kostki ICOM (**I**nput, **C**ontrol, **O**utput, **M**echanizm) – patrz rysunek 1.

Jakie, wspomniane w tytule artykułu, kryteria jakości informacji spośród kryteriów elementarnych są istotne dla zasobów informacyjnych danej organizacji i jakie są ich wymagane wartości, powinno zostać:

- określone podczas analizy ryzyka,
- zatwierdzone przez kierownictwo tej organizacji,
- wpisane do odpowiednich dokumentów, np. polityki bezpieczeństwa.

Podstawowy zbiór kryteriów, z których wybierane są istotne, tworzą zwykle tajność, integralność, dostępność. W dalszej kolejności także kryteria dotyczące działań na zasobach informacyjnych, np. rozliczalność, niezaprzeczalność itp. Wymienione kryteria wyznaczają podstawowe klasy incydentów – w tym artykule zostanie przedstawione zagadnienie szacowania ryzyka wystąpienia incydentów niepożądanej zmiany wartości jednego z tych kryteriów elementarnych – tajności.



Rys. 1. Incydent i obsługa incydentu

Tajność wskazuje wymagany stopień (siłę) ochrony zasobów informacyjnych przed nieuprawnionym uzyskaniem o tych zasobach jakichkolwiek informacji. Stopień ten jest uzgadniany przez podmioty wymieniające informację. Przykładem ujęcia w ramy prawne zagadnienia tajności zasobów informacyjnych jest ustawa [9] i rozporządzenie [8], gdzie ustalono zbiór wartości stopnia ochrony jako:

{*ściśle_tajne, tajne, poufne, zastrzeżone*}

2. Oceny opisowe w szacowaniu ryzyka i składanie ocen

Istnienie ryzyka wystąpienia incydentu z zakresu bezpieczeństwa informacyjnego dla (jakiegoś) zasobu informacyjnego oznacza występowanie zagrożenia oddziałującego na istotne kryteria jakości informacji (tajność, integralność, dostępność itp.) tego zasobu, a wielkość ryzyka określa się, łącząc wartość oceny możliwości wystąpienia zdarzeń powodowanych przez zagrożenie i wartość oceny wielkości szkód powstałych w wyniku tych zdarzeń, czyli skutków incydentu.

Szacowanie ryzyka polega na określeniu możliwości realizacji zagrożeń (do tego potrzebna jest także znajomość, jak pokazano to na rys. 1, podatności) oraz potencjalnych strat. Oszacowanie takie uwzględnia dwa podstawowe parametry: możliwość lub prawdopodobieństwo realizacji zagrożenia i mierzalne skutki tego zdarzenia. W oszacowaniach ryzyka wybór sprowadza się w zasadzie do dwóch metod [1]:

1. Metody **ilościowej** analizy ryzyka, gdzie operuje się miarą zdarzenia losowego – prawdopodobieństwem wyrażonym liczbą z przedziału [0, 1].
2. Metody **jakościowej** analizy ryzyka, gdzie operuje się opisowymi, arbitralnie dobieranymi miarami wyrażającymi możliwość zajścia zdarzenia.

W tym artykule założono, że szacowanie ryzyka odbywa się z wykorzystaniem metody jakościowej. Założenia szczegółowe są następujące:

- dany jest zasób informacyjny $z_i \in Z$, gdzie $Z=\{z_i|\ i\in[1, n]\}$ to zbiór zasobów podlegających analizie ryzyka;
- z_i ma podatności $p_j \in P$, gdzie $P=\{p_j|\ j\in[1, m]\}$ to zbiór podatności zasobów należących do zbioru Z;
- podatność może być wykorzystana przez zagrożenie $d_k \in D$, gdzie $D=\{d_k|\ k\in[1, r]\}$ to zbiór zidentyfikowanych zagrożeń mogących oddziaływać na zasoby Z,
- analiza ryzyka zostanie przeprowadzona w wariancie zasobowym.

Należy określić:

1) Jednolite oceny symboliczne dla właściwości (cech): zagrożeń $d_k \in D$, podatności $p_j \in P$, szkód i strat oraz ryzyka. Te cechy opisujące składowe ryzyka to:
   - możliwość realizacji zagrożenia, oznaczona dalej jako MRZ,
   - stopień podatności, oznaczony dalej jako PZ,
   - wielkość strat, oznaczona dalej jako ST,
   - wielkość ryzyka, oznaczona dalej jako RYZYKO.
2) Sposób przypisywania ocen i wyliczania ocen uogólnionych.

Proponuje się przyjęcie następującego systemu K przypisywania ocen opisowych ze zbioru OCENA wybranym cechom ze zbioru CECHA:

$$K = \langle \text{CECHA, OCENA, } \textit{Procedura} \rangle$$

gdzie:

- CECHA – zbiór cech: MRZ, PZ, ST, RYZYKO.
- OCENA – zbiór ocen opisowych. W tym artykule przyjęto, że jest to zbiór trójelementowy {N, S, W}, gdzie:
  W – możliwość, stopień lub strata WYSOKA,
  S – możliwość, stopień lub strata ŚREDNIA,
  N – możliwość, stopień lub strata NISKA.
- *Procedura* – podaje sposób przypisania ocen opisowych ze zbioru OCENA cechom ze zbioru CECHA (np. decyzją ekspertów w ramach sesji „burzy mózgów").

Rekomenduje się (za [2]) następujący sposób składania ocen opisowych (algorytm 1):

ALGORYTM 1

Niech będą dane:

1) zbiór A ocen uporządkowanych rosnąco, tj. $A=\{q_1, \ldots, q_i, \ldots, q_n\}$, gdzie $i \in N$ jest numerem (pozycji) elementu q_j w zbiorze A i $q_j < q_{j+1}$;
2) „element_wybrany" – dowolny, wskazany arbitralnie element zbioru A.
3) liczba $r \in R$ równa reszcie ze średniej numerów pozycji wybranych elementów.

JEŻELI $r = 0$ TO

$q = q_k$ gdzie k jest średnią z numerów pozycji elementów wybranych ze zbioru A

W PRZECIWNYM RAZIE ($r \neq 0$):

JEŻELI r ≥0,5 TO

$q = q_{\lceil k \rceil}$

W PRZECIWNYM RAZIE (r < 0,5)

$q = q_{\lfloor k \rfloor}$

KONIEC ALGORYTMU 1

W przedstawionym algorytmie *podłoga*, *część całkowita*, *cecha* lub *entier* liczby rzeczywistej *x*, oznaczana $\lfloor x \rfloor$, to największa liczba całkowita nie większa od *x*. Natomiast *sufit* lub *cecha górna* liczby rzeczywistej *x*, to najmniejsza liczba całkowita nie mniejsza od *x*, oznaczana symbolem $\lceil x \rceil$.

PRZYKŁAD 1

Niech symbol ☼ oznacza operację składania ocen opisowych zgodnie z algorytmem 1. Jeżeli np. przyjęte będą następujące wartości ocen opisowych:

Z (znikome), N (niskie), S (średnie), W (wysokie), K (katastrofalne), czyli

A={Z, N, S, W, K}

gdzie

Z< N< S< W< K, czyli q_1=Z, q_2=N, q_3=S, q_4=W, q_5=K,

to dla q=Z☼K☼N☼K:

(1+5+2+5)/4=3,**25**, czyli r < 0,5,

zatem: q_1 ☼q_5 ☼q_2 ☼q_5 = $q_{\lfloor 3,25 \rfloor}$ =q_3 tzn. Z☼K☼N☼K=S

******** (koniec przykładu)

W celu oszacowania ryzyka powstania strat na skutek zajścia incydentu, dla zasobu $z_i \in Z$, konkretnego zagrożenia $d_k \in D$ i podatności $p_j \in P$, należy (algorytm 2):

ALGORYTM 2

Niech *ocena*: CECHA→OCENA

Używając ocen ze zbioru {N, S, W}, oszacować:

1) możliwość realizacji zagrożenia (MRZ) „jako takiego”[2] dla $d_k \in D$, tzn. podać wartość *ocena*(MRZ);

[2] Na tym etapie szacuje się „potencjalność” zagrożenia. Na przykład oceniając możliwość kradzieży sprzętu komputerowego z siedziby organizacji, nie bierze się pod uwagę krat, zamków, systemów alarmowych itp., w które wyposażony jest budynek (to wpływa na podatność na kradzież, co rozpatrywane jest na etapie kolejnym), tylko bierze się pod uwagę to, że obiekt ten znajduje się w dzielnicy, w której mieszka dużo złodziei.

2) stopień PZ podatności $p_j \in P$ zasobu $z_i \in Z$, która to podatność może być wykorzystana przez zagrożenie, tzn. podać wartość *ocena*(PZ);
3) według formuły (1) RYZYKO zajścia zdarzenia takiego, że zagrożenie $d_k \in D$ wykorzysta podatność $p_j \in P$ do spowodowania szkody o wartości strat ST:

$$ocena(\text{RYZYKO}) = ocena(\text{MRZ}) \☼ ocena(\text{PZ}) \☼ ocena(\text{ST}) \quad (1)$$

KONIEC ALGORYTMU 2

Interpretacja elementów algorytmu 2 dla zagadnienia niepożądanej zmiany wartości istotnego kryterium jakości informacji, jakim jest tajność, jest następująca:

- $d_k \in D$ to zagrożenie dla tajności zasobu informacyjnego;
- $p_j \in P$ to podatność na działania naruszające tajność zasobu informacyjnego;
- MRZ to możliwość realizacji zagrożenia skutkującego niepożądaną zmianą wartości kryterium jakości „tajność";
- ST to wartość strat spowodowanych szkodami związanymi z niepożądaną zmianą wartości kryterium jakości „tajność";
- RYZYKO to ryzyko zaistnienia incydentu niepożądanej zmiany wartości kryterium jakości „tajność" zasobu $z_i \in Z$.

W kolejnym rozdziale są przedstawione rozważania na temat możliwości ustalenia (wskazania i opisu) elementów zbiorów D i P oraz zmiennych MRZ, PZ i ST i przypisywania wartości (ocen opisowych) tym zmiennym.

3. Ustalanie zmiennych i wartości ich ocen opisowych

Zgodnie z tematyką artykułu, ochronie podlega tajność zasobu informacyjnego. Przed przystąpieniem do szacowania ryzyka należy sprecyzować i uzgodnić ze wszystkimi zainteresowanymi stronami znaczenie terminu „tajność"[3]. Jak można bowiem zauważyć chociażby na podstawie PRZYKŁADU 2, będzie to miało wpływ m.in. na wielkość szacowanych skutków realizacji zagrożenia.

PRZYKŁAD 2

W zdaniu „Nie mogę dać Tobie tego dokumentu, bo jest tajny" w zależności od okoliczności, słowo „tajny" może oznaczać jedną z niżej wymienionych możliwości:

[3] Równie dobrym określeniem byłaby „niejawność", ale ze względu na inne publikacje również w tej jest stosowany termin „tajność".

1. Dokument jest „tajny” w potocznym rozumieniu tego słowa, jego właściciel z różnych powodów nie chce go upowszechniać.
2. Dokument ma nadaną klauzulę „tajne” na podstawie ustawy o ochronie informacji niejawnych (Dz.U. z 2010 r., nr 182, poz. 1228).
3. Dokument jest „tajny” w rozumieniu obowiązujących w danej organizacji przepisów (być może tylko wewnętrznych) dotyczących upowszechniania informacji, bo np. zawiera informacje stanowiące prawnie chronione tajemnice przedsiębiorstwa (w rozumieniu art. 11 pkt. 4 ustawy z dn. 16.04.1993 „o zwalczaniu nieuczciwej konkurencji”; Dz.U.93.47.211).

******** (koniec przykładu)

Należy również ustalić i uzyskać konsensus wśród interesariuszy co do tego, jaka jest natura „tajności”. Czy tajność to jest coś niepodzielnego (coś jest albo nie jest tajne, innych możliwości nie ma), czy też „tajność” można jakoś stopniować. Na świecie przeważa pogląd, że tajność podlega stopniowaniu. W niemałym stopniu przyczyniły się do tego wczesne, z początku lat 70 XX wieku, prace z dziedziny bezpieczeństwa, w szczególności prace D. Bella i L. La Paduli (patrz np. opisy wyników tych prac w rozdz. 2 w [1]). Stopniowanie „tajności” zostało zapisane też w polskiej ustawie *o ochronie informacji niejawnych* [9], gdzie wprowadzono cztery wartości dla kryterium „tajność”. Pogląd ten jest przyjęty również w tym artykule.

Jednak takie podejście rodzi pewne komplikacje interpretacyjne – jak oszacować skutki incydentu dotyczącego tajności, zakładając, że zostały ustalone wartości ocen oraz ustalono możliwe stopnie tajności? Jak zostało wspomniane wcześniej, „tajność” określa wymaganą siłę zabezpieczeń chroniących zasoby informacyjne przed nieuprawnioną możliwością uzyskania o tych zasobach jakichkolwiek informacji. Stopnie (w postaci etykiet przypisywanych obiektom) określają zbiory wymagań do nich przypisane. Zakładając, że zmienna TAJNOŚĆ może przyjmować wartości ze zbioru[4]:

{tajne, poufne, zastrzeżone}

to fraza „niepożądana zmiana kryterium jakości informacji TAJNOŚĆ” oznacza zmianę z TAJNOŚĆ na ~TAJNOŚĆ[5]. Na przykład zasób informacyjny w postaci zbioru na dysku macierzy dyskowej był chroniony zgodnie z wymaganiami na tajność na poziomie „poufne” (czyli TAJNOŚĆ = poufne), po realizacji zagrożenia ten warunek nie jest spełniony, tj. ~(TAJNOŚĆ = poufne), np. wykazano, że można ominąć wymagane zabezpieczenia lub przełamano zabezpieczenia uniemożliwiające nieuprawniony

[4] Taki trójelementowy zbiór wartości zmiennej TAJNOŚĆ jest stosowany w dalszej części artykułu: w tabelach i przykładach.

[5] Symbol ~ to funktor zdaniotwórczy „nieprawda, że”.

odczyt informacji. Czy oznacza to, że intruz może odczytać tylko informacje klasyfikowane jako „poufne", a „zastrzeżone" nie, czy też, że może odczytać zarówno „poufne", jak i niżej klasyfikowane „zastrzeżone", zakładając, że zbiór tak klasyfikowanych informacji znajduje się na tym samym dysku co zbiór poufny? Odpowiedź na to pytanie wymaga przeanalizowania możliwych scenariuszy realizacji zagrożenia, uwzględniających rzeczywistą alokację chronionych zasobów informacyjnych oraz zastosowanych zabezpieczeń. Inną praktyczną konsekwencją przyjęcia możliwości stopniowania „tajności" jest konieczność uwzględniania klasyfikacji[6] zasobu informacyjnego przy przypisywaniu ocen w ramach szacowania ryzyka – co najlepiej widać w tabeli 5 (ostatnia kolumna) i tabeli 10.

W tym artykule nie rozważa się typów zasobu informacyjnego ze względu na jego nośnik (elektroniczny, papierowy, na taśmie mikrofilmowej itd.), chociaż przy kompleksowej i szczegółowej analizie ryzyka przydatna może być taka wstępna kategoryzacja zasobów. Będzie to systematyzowało identyfikację możliwych sposobów realizacji zagrożenia, np. ze względu na niezbędne do realizacji zagrożenia narzędzia oraz możliwe podatności.

Ze względów praktycznych warto także przyjąć jednolity sposób opisu zagrożenia. Propozycję takiego sposobu prezentuje tabela 1. Przy określaniu szacowanych skutków oraz potencjału realizacji zagrożenia można wspomóc się zapisami np. normy ISO/IEC WD 29115 [4]. Norma ta dotyczy co prawda tożsamości i uwierzytelniania, ale można w niej znaleźć pewne wskazówki na co należy zwrócić uwagę także przy rozważaniu wpływu naruszenia tajności zasobu informacyjnego. Według zapisów tej normy, potencjalny wpływ błędnego uwierzytelnienia dotyczy:

1. Niewygody, dolegliwości lub uszczerbku na reputacji lub pozycji.
2. Straty finansowej lub odpowiedzialności podmiotu.
3. Szkody dla podmiotu, jego planów lub publicznych interesów.
4. Wycieku informacji wrażliwych lub nieuprawnionego dostępu do nich.
5. Bezpieczeństwa osobowego,
6. Naruszenia prawa cywilnego lub karnego.

Siła każdego z ww. czynników jest określona w skali wartości: niski, umiarkowany, znaczący, wysoki (czyli zbiór ocen różni się od przyjętego w tym artykule, ale dla prowadzonych tutaj rozważań nie ma to znaczenia). Do organizacji należy określenie, na podstawie oszacowania ryzyka właściwego dla

[6] Zasób klasyfikowany to taki zasób, dla którego wyznaczono klasę bezpieczeństwa określając wymagany poziom istotnego kryterium jakości dotyczącego bezpieczeństwa i kategorię. Jeżeli tym istotnym kryterium jakości jest TAJNOŚĆ, a zbiór wartości tej zmiennej tworzą trzy etykiety {tajne, poufne, zastrzeżone}, to przykładowa klasa bezpieczeństwa może mieć postać: ⟨poufne, dokumenty_ABW⟩.

tej organizacji, jaka jest ich interpretacja, np. jaki poziom strat finansowych oznacza niski, jaki umiarkowany itd. Norma ISO/IEC WD 29115 nie określa sposobu przeprowadzenia szacowania ryzyka – można ją wykonać np. tak, jak zaleca norma PN-ISO/IEC 27005:2010 [5].

Tab. 1. Wzorzec arkusza opisu zagrożenia (przykład)

ARKUSZ nr ……. OPISU ZAGROŻENIA DLA TAJNOŚCI ZASOBU INFORMACYJNEGO $z_i \in Z$ **Identyfikator zagrożenia:** [symbol zagrożenia]	
Zagrożenie:	[jednozdaniowa nazwa opisowa zagrożenia, np. *działania intruza – pracownika tej organizacji*]
Scenariusz realizacji zagrożenia:	[kilkuzdaniowy opis słowny lub w postaci schematu blokowego]
Właściciel zasobu:	[dane identyfikacyjne]
Możliwe (szacowane) skutki/szkody realizacji zagrożenia:	[kilkuzdaniowy opis słowny lub specyfikacja w postaci listy]
Możliwe (szacowane) straty w przypadku realizacji zagrożenia:	[kwota w określonej walucie lub opisowo]
Potencjał zagrożenia:	[kilkuzdaniowy opis słowny]

Inna sytuacja jest w przypadku podmiotów (organizacji) podlegających ustawie o *krajowym systemie cyberbezpieczeństwa* ([10], dalej w skrócie KSC), gdzie zapisano:

> Art. 6. Rada Ministrów określi, w drodze rozporządzenia:
> 1) wykaz usług kluczowych, o których mowa w art. 5 ust. 2 pkt 1, kierując się przyporządkowaniem usługi kluczowej do danego sektora, podsektora i rodzaju podmiotu wymienionych w załączniku nr 1 do ustawy oraz znaczeniem usługi dla utrzymania krytycznej działalności społecznej lub gospodarczej;
> 2) **progi istotności skutku zakłócającego incydentu** dla świadczenia usług kluczowych, wymienionych w wykazie usług kluczowych, uwzględniając:
> a) liczbę użytkowników zależnych od usługi kluczowej świadczonej przez dany podmiot,
> b) zależność innych sektorów, o których mowa w załączniku nr 1 do ustawy, od usługi świadczonej przez ten podmiot,
> c) wpływ, jaki mógłby mieć incydent, ze względu na jego skalę i czas trwania na działalność gospodarczą i społeczną lub bezpieczeństwo publiczne,
> d) udział podmiotu świadczącego usługę kluczową w rynku,
> e) zasięg geograficzny obszaru, którego mógłby dotyczyć incydent,

f) zdolność podmiotu do utrzymywania wystarczającego poziomu świadczenia usługi kluczowej, przy uwzględnieniu dostępności alternatywnych sposobów jej świadczenia,

g) inne czynniki charakterystyczne dla danego sektora lub podsektora, jeżeli występują – kierując się potrzebą zapewnienia ochrony przed zagrożeniem życia lub zdrowia ludzi, znacznymi stratami majątkowymi oraz obniżeniem jakości świadczonej usługi kluczowej.

Art. 6 oznacza, dla określonych w KSC podmiotów, obligatoryjność zawartego w nim sposobu opisu incydentu. Wydaje się, że dla usprawnienia procesu obsługi incydentów w skali nie tylko krajowej ale i europejskiej, proponowany sposób opisu incydentu powinny wykorzystać także organizacje (podmioty) nie podlegające pod regulacje KSC.

W ramach wstępnych ustaleń należy także sprecyzować, co może być zagrożeniem dla tajności zasobu informacyjnego. W tym przypadku[7] jest to tzw. czynnik ludzki, który przejawia się w formie działań celowych lub błędnych. Ustalenie tego jest niezbędne przede wszystkim do oszacowania wartości możliwości realizacji zagrożenia (czynnika MRZ) oraz pozostałych czynników ryzyka (PZ, ST). Takie ustalenia można przeprowadzić, posługując się np. następującą tabelą:

Tab. 2. Zagrożenia dla tajności zasobu informacyjnego (przykład opisu)

Rodzaj działania	Kto	Motyw
CELOWE	intruz, pracownik	korzyści (finansowe, ideowe, psychologiczne, itp.), zemsta, ciekawość, szantaż, itp.
BŁĘDNE	pracownik	brak

W przypadku szacowania ryzyka możliwości narażenia organizacji i jej zasobów informacyjnych na realizację zagrożenia działaniami celowymi potrzebny jest opis organizacji pod kątem jej atrakcyjności dla intruza. Opis ten powinien uwzględniać czynniki wpływające na motywację intruza. Można to zrobić poprzez przyjęcie pewnego zestawu cech (dalej symbolicznie ZC) w postaci czwórki uporządkowanej opisującej z tej perspektywy organizację:

$$ZC = \langle WO, OD, NS, AT \rangle$$

gdzie:

- $WO=\{wo_i | \ i \in [1,m]\}$ jest niepustym zbiorem cech opisujących wielkość „biznesową” organizacji;

[7] W odróżnieniu od zagrożeń dla dostępności zasobu informacyjnego, gdzie najczęstszymi są awarie infrastruktury, w której jest osadzony zasób informacyjny oraz oddziałujące na tę infrastrukturę i sam zasób katastrofy i niekorzystne zjawiska naturalne (np. powodzie lub pożary).

- $OD=\{od_j|\ j\in[1,n]\}$ jest niepustym zbiorem cech opisujących obszar działalności organizacji;
- $NS=\{ns_k|\ k\in[1,l]\}$ jest niepustym zbiorem cech opisujących wpływ działalności organizacji na nastroje społeczeństwa;
- $AT=\{at_p|\ p\in[1,r]\}$ jest niepustym zbiorem cech opisujących „atrakcyjność" branży dla intruza.

Te zbiory cech powinny zostać skonkretyzowane przez ekspertów lub, żeby uzyskać powtarzalność analizy, narzucone przez odpowiednie regulacje[8]. Załóżmy, że ww. zbiory cech zostały skonkretyzowane następująco:

- WO = {wielka, średnia, mała, mikroprzedsiębiorstwo};
- OD = {globalny, lokalny};
- NS = {znaczący, umiarkowany, niski, brak};
- AT jest zadany predefiniowaną tabelą 3:

Tab. 3. Rodzaj organizacji a „atrakcyjność" dla intruza (przykład)

Rodzaj organizacji	Atrakcyjność
Telekomy	**duża**
Spółki medialne	
Administracja publiczna	
….	
Sieci sklepów	**umiarkowana**
Przemysł obronny	
Medycyna	
….	
Inne	**mała**
….	

Jak już wspomniano, ustalenie zbioru cech (ile i jakich elementów), ich skonkretyzowanie oraz przypisanie wartości ocen, powinno odbyć się w gronie ekspertów np. metodą „burzy mózgów" lub odnosić się do znanych oficjalnych regulacji. Należy zauważyć, że przedstawiony opis nie będzie miał zastosowania, gdy motywacją intruza jest zemsta lub zamiar wyrządzenia szkody określonej organizacji (np. zapłacono za to intruzowi). W tych przypadkach należy przyjąć, że wartość MRZ jest wysoka. Trzeba wziąć też pod uwagę, że „atrakcyjność" obiektu ataku jest tylko jednym z elementów wpływających na motywację intruza. Na pewno osłabią ją wysokie kary za tego

[8] Otwarte pozostaje pytanie, jaki podmiot miałby takie regulacje wydać. RCB? Ministerstwo Cyfryzacji?

typu przestępstwa, skuteczność ich ścigania oraz przeświadczenie o wysokiej sile zabezpieczeń, które należy przełamać (chociaż akurat ten czynnik równie dobrze może być motywacją dla intruza lubiącego wyzwania).

PRZYKŁAD 3

Dla wielkiej spółki medialnej działającej w skali globalnej, przykładowy zbiór ZC_p wartości cech może mieć postać:

$$ZC_p = \{\text{wielka, globalny, umiarkowany, duża}\}$$

Ten zbiór cech opisujących organizację należy przełożyć na oceny przyjęte do szacowania ryzyka, tzn. należy skonstruować tablicę interpretacyjną na wzór tabeli 4.

Tab. 4. Zbiór cech opisujących organizację i ich wartości oraz odpowiadające im oceny (przykład)

CECHA	Możliwe wartości zmiennej CECHA	Ocena
WO	wielka	W
	średnia	S
	mała, mikroprzedsiębiorstwo	N
OD	globalny	W
	lokalny	N
NS	znaczący	W
	umiarkowany	S
	niski żaden	N
AT	klasa „telekomy”, klasa „sieci sklepów”	W
	–	S
	klasa „inne”	N

Wtedy wartość MRZ – możliwości narażenia organizacji na celowe działania intruza określa formuła:

$$\textit{ocena}[MRZ(ZC_p)] = \textit{ocena}[MRZ(\{\text{wielka, globalny, umiarkowany, duża}\})] = MRZ(\{W, W, S, W\})$$

gdzie *ocena* jest funkcją (zwykle heurystyczną) przypisującą oceny ze zbioru ocen (w tym artykule – ze zbioru {N, S, W}) elementom zbioru cech (tutaj: {wielka, globalny, umiarkowany, duża}), tj.:

$$\textit{ocena}: MRZ(ZC) \rightarrow MRZ(OCENA)$$

gdzie: $OCENA=\{ocena_i \mid i \in [1,n]\}=\{N, S, W\}$

Przyjmując w tym przykładzie sposób składania ocen według algorytmu 1, otrzymuje się wysoką możliwość realizacji zagrożenia:

alg(MRZ({W, W, S, W}))=MRZ(W)

******** (koniec przykładu)

W przypadku szacowania ryzyka narażenia organizacji na błędne działania jej pracowników, aby ustalić wartość MRZ, należy dysponować danymi historycznymi dotyczącymi popełnionych błędów, których skutkiem był incydent dotyczący tajności zasobu informacyjnego. Takie dane, obejmujące zarówno typ błędu, jak i jego częstość, powinny być zbierane przez organizację. Jeżeli takich danych brak, to można skorzystać z danych organizacji o podobnym profilu działalności, o ile takie dane są dostępne. Trzecią możliwością jest wykorzystanie uogólnionych danych statystycznych, publikowanych przez różne organizacje zajmujące się bezpieczeństwem informacyjnym (np. CERT). Oczywiście, takie zapisy obejmują tylko przypadki błędów wykrytych oraz mogą być nieadekwatne do rzeczywistej sytuacji organizacji, dla której jest szacowane ryzyko.

Skutkiem błędów może być ujawnienie zawartości zasobu informacyjnego nieuprawnionym podmiotom, upublicznienie informacji o istnieniu niejawnego zasobu w systemie, ujawnienie wszystkich bądź niektórych podmiotów, które są uprawnione do dostępu do takiego zasobu, możliwość doprowadzenia do ww. sytuacji przez wykonanie niedozwolonej operacji w systemie itp. Mając przygotowaną tabelę możliwych skutków (opracowaną np. w ramach eksperckiej „burzy mózgów"), należy przypisać im, na podstawie danych historycznych, częstości występowania oraz na podstawie tabel interpretacyjnych (np. takich jak tabele 6-9) przyporządkować odpowiednie oceny w zależności od klasyfikacji zasobu, którego taki incydent dotyczy. Przykład opisu prezentuje tabela 5.

Zakładając, że ujawnienie zawartości zasobu informacyjnego klasyfikowanego jako *tajne* bądź *poufne* nie jest dopuszczalne w żadnych okolicznościach, podobnie jak upublicznienie informacji o istnieniu w systemie zasobu klasyfikowanego jako *tajne*, oraz określając progi częstości realizacji określonych zdarzeń, można skonstruować tabele interpretacyjne ocen opisowych dla MRZ. Tego typu założenia-decyzje, dotyczące progów częstości realizacji określonego zdarzenia musi podjąć zarząd organizacji lub jej dział bezpieczeństwa. Przykłady takich opisów prezentują tabele 6-9. Po uwzględnieniu zawartości tabel interpretacyjnych można wypełnić ostatnią kolumnę tabeli 5.

Tab. 5. Skutki błędów, dane empiryczne o częstości i ocena MRZ po uwzględnieniu tabel interpretacyjnych 6-9 (przykład)

Lp.	Skutki błędu dotyczącego tajności	Częstość incydentu	Ocena dla MRZ (na podstawie tabel 6-9)	
1	Ujawnienie zawartości zasobu niejawnego nieuprawnionym podmiotom	Raz na dwa lata	*tajne*	W
			poufne	W
			zastrzeżone	N
2	Upublicznienie informacji o istnieniu niejawnego zasobu w systemie	Raz na trzy lata	*tajne*	W
			poufne	S
			zastrzeżone	S
3	Ujawnienie wszystkich bądź niektórych podmiotów, które są uprawnione do dostępu do zasobu niejawnego	Dwa razy w ciągu roku	*tajne*	W
			poufne	S
			zastrzeżone	N
4	Możliwość doprowadzenia do sytuacji 1-3 przez wykonanie niedozwolonej operacji w systemie	Pięć razy w ciągu roku	*tajne*	W
			poufne	W
			zastrzeżone	W
5	….	….	….	

Tab. 6. Interpretacja ocen opisowych dla możliwości realizacji zagrożenia (MRZ) dla błędu ujawnienia zawartości niejawnego zasobu informacyjnego nieuprawnionym podmiotom

OCENA	INTERPRETACJA
W	Dla *tajne* i *poufne* zawsze, gdy błąd popełniono, *zastrzeżone* więcej niż raz na rok
S	*zastrzeżone* raz na rok
N	*zastrzeżone* raz na dwa lata

Tab. 7. Interpretacja ocen opisowych dla możliwości realizacji zagrożenia (MRZ) dla błędu upublicznienia informacji o istnieniu niejawnego zasobu w systemie

OCENA	INTERPRETACJA
W	Dla *tajne* zawsze, gdy popełniono błąd
S	Dla *poufne* i *zastrzeżone* zawsze, gdy popełniono błąd
N	nigdy

Tab. 8. Interpretacja ocen opisowych dla możliwości realizacji zagrożenia (MRZ) dla błędu ujawnienia wszystkich bądź niektórych podmiotów które są uprawnione do dostępu do zasobu niejawnego

OCENA	INTERPRETACJA
W	Dla *tajne* bez względu na częstość
S	Dla *poufne* bez względu na częstość
N	Dla *zastrzeżone* bez względu na częstość

Tab. 9. Interpretacja ocen opisowych dla możliwości realizacji zagrożenia (MRZ) dla błędu możliwość doprowadzenia do sytuacji z pozycji 1-4 tab. 5 przez wykonanie niedozwolonej operacji w systemie

OCENA	INTERPRETACJA
W	Dla *tajne* bez względu na częstość, dla *poufne*, gdy raz na rok lub więcej, dla *zastrzeżone,* gdy więcej niż trzy razy na rok
S	dla *poufne*, gdy nie więcej niż raz na dwa lata, dla *zastrzeżone,* gdy dwa lub trzy razy na rok
N	dla *poufne*, gdy nie więcej niż raz na trzy lata, dla *zastrzeżone,* gdy najwyżej raz na rok

Uwzględniając dyskusję skutków przeprowadzoną wcześniej w tym rozdziale (przy propozycji opisu zagrożenia), można przyjąć, że szkody, powstałe na skutek zarówno działań celowych, jak i błędnych, będą zależały od następujących czynników:

1. Nadanej zasobowi informacyjnemu klasy bezpieczeństwa (w tym artykule te klasy są wyznaczone etykietami *tajne*, *poufne*, *zastrzeżone*).
2. Liczby użytkowników zależnych od zasobu/usługi, której dotyczy incydent.
3. Zależności innych organizacji (lub sektorów w rozumieniu KSC), od zasobu/usługi, której dotyczy incydent.
4. Wpływu, jaki mógłby mieć incydent, ze względu na jego skalę i czas trwania na działalność gospodarczą i społeczną lub bezpieczeństwo publiczne.
5. Udziału organizacji, której dotyczy incydent, w rynku.
6. Zasięgu geograficznego obszaru, którego mógłby dotyczyć incydent.
7. Naruszenia prawa cywilnego lub karnego.
8. Wpływu na bezpieczeństwo osobowe.
9. Wpływu na wizerunek organizacji.
10. Wpływu na plany lub publiczne interesy organizacji.

W odróżnieniu od oszacowań prowadzonych dla utraty dostępności zasobu informacyjnego (patrz np. przykład 3.6 w [1]), w przypadku naruszenia jego tajności szkody są trudno przekładalne na straty mierzone w określonej walucie. Prowadzi to do komplikacji w ustalaniu zawartości tabeli

interpretacyjnej dla strat (ST). Dlatego proponuje się zrezygnować z operowania stratami, a czynnik ST w formule (1) interpretować jako ocenianą ekspercko wielkość szkody. Zakładając, że lista czynników wpływających na wielkość szkody ogranicza się do dziesięciu ww. czynników, należy opracować, posługując się wytycznymi normatywnymi, prawnymi i opiniami ekspertów, tabelę interpretacyjną na wzór tabeli 10.

Tab. 10. Specyfikacja możliwych szkód (ST) i wartości ocen (przykład)

Lp.	Czynniki wpływające na rozmiar szkody	Rozmiar szkody	Ocena
1	Klasa bezpieczeństwa zasobu, którego dotyczył incydent	*tajne*	W
		poufne	S
		zastrzeżone	N
2	Liczba użytkowników dotkniętych skutkami incydentu	Zależne od sektora działania organizacji, zgodnie z KSC, szczegóły por. Rozporządzenie[9]	
3	Zależność innych organizacji	Jw.	
4	Wpływ na działalność gospodarczą i społeczną lub bezpieczeństwo publiczne	Jw.	
5	Udział w rynku	Jw.	
6	Zasięg geograficzny incydentu	Jw.	
7	Naruszenia prawa cywilnego lub karnego	Ocena w gestii Działu Prawnego organizacji	
8	Wpływ na bezpieczeństwo osobowe	Ocena w gestii Działu Bezpieczeństwa organizacji	
9	Wpływ na wizerunek organizacji	Ocena w gestii Zarządu	
10	Wpływ na plany lub publiczne interesy organizacji	Jw.	

Uogólnioną wartość szkody należy oszacować, posługując się algorytmem 1, na wzór tego, co pokazano dla MRZ w przykładzie 3.

Żeby zagrożenie mogło się zrealizować, musi istnieć odpowiadająca mu podatność. Zbiór podatności ustala się w praktyce na podstawie wyników działania skanerów bezpieczeństwa (wykrywających podatności w oprogramowaniu i plikach konfiguracyjnych), testów penetracyjnych[10], wizji

[9] Rozporządzenie Rady Ministrów z dn. 31.10.2018 r. *w sprawie progów uznania incydentu za poważny*. Dz.U. poz. 2180.

[10] Powinny obejmować także badanie odporności personelu na działania socjotechniczne oraz odporność na penetrację zabezpieczeń fizycznych.

lokalnych, przeglądu dokumentacji i konsultacji eksperckich. Następnie ustala się (zwykle korzystając z ocen ekspertów) stopień podatności dla wszystkich elementów ww. zbioru. Wyniki można przedstawić np. w tabelach, tak jak to obrazują tabele 11 i 12[11].

Tab. 11. Interpretacja ocen opisowych dla podatności[12] na działania celowe intruza (przykład)

OCENA	INTERPRETACJA
W	~ (wymagane zabezpieczenia dla poziomu ochrony „zastrzeżone", „poufne", „tajne") ∧ ~ (poprawna konfiguracja zabezpieczeń)
S	[~ (wymagane zabezpieczenia dla poziomu ochrony „zastrzeżone", „poufne", „tajne") ∧ (poprawna konfiguracja zabezpieczeń)] ∨ [(wymagane zabezpieczenia dla poziomu ochrony „zastrzeżone", „poufne", „tajne") ∧ ~ (poprawna konfiguracja zabezpieczeń)]
N	(wymagane zabezpieczenia dla poziomu ochrony „zastrzeżone", „poufne", „tajne") ∧ (poprawna konfiguracja zabezpieczeń)

Tab. 12. Interpretacja ocen opisowych dla podatności na błędne działania pracownika (przykład)

OCENA	INTERPRETACJA
W	~ (właściwe szkolenie pracowników z zakresu ochrony zasobów informacyjnych) ∧ ~ (nadzór nad działaniami pracowników)
S	[~ (właściwe szkolenie pracowników z zakresu ochrony zasobów informacyjnych) ∧ (nadzór nad działaniami pracowników)] ∨ [(właściwe szkolenie pracowników z zakresu ochrony zasobów informacyjnych) ∧ ~ (nadzór nad działaniami pracowników)]
N	(właściwe szkolenie pracowników z zakresu ochrony zasobów informacyjnych) ∧ (nadzór nad działaniami pracowników)

PRZYKŁAD 4

Dla organizacji z domeny *gov.pl* postanowiono oszacować ryzyko narażenia eksploatowanych w niej zasobów informacyjnych na nakierowane na naruszenie tajności tych zasobów działania celowe intruzów i błędy pracowników, powodujące incydenty z zakresu bezpieczeństwa informacyjnego. Oszacowania

[11] Symbole ∧, ∨, oraz ~, to funktory zdaniotwórcze odpowiednio „i", „lub" oraz „nieprawda, że".

[12] W tej tabeli i kolejnej liczbę podatności ograniczono do dwóch. W praktyce ich liczba będzie zależna od wyników identyfikacji a sposób ich złożenia, w celu uzyskania interpretacji ocen, będzie zależał od wiedzy i decyzji analityka ryzyka lub wspierającego go eksperta dziedzinowego.

dotyczą zasobów klasyfikowanych jako *tajne*, *poufne* i *zastrzeżone*. Władze organizacji, wspólnie z pracownikami jej Działu Bezpieczeństwa oraz zatrudnionymi na umowę-zlecenie specjalistami od analizy ryzyka[13], ustaliły na podstawie danych historycznych oraz oszacowań eksperckich co następuje:

1. Incydenty typu „błąd pracowniczy" powodowały zwykle dwa skutki:
   a) ujawnienie zawartości poufnego zasobu informacyjnego nieuprawnionym podmiotom;
   b) ujawnienie niektórych podmiotów, które są uprawnione do dostępu do zasobu klasyfikowanego jako „tajne".
2. Incydenty typu „działanie celowe" powodowały zwykle dwa skutki:
   a) ujawnienie wszystkich podmiotów, które są uprawnione do dostępu do zasobu klasyfikowanego jako „poufne";
   b) ujawnienie zawartości zastrzeżonego zasobu informacyjnego nieuprawnionym podmiotom;
   c) nie występowały działania motywowane zemstą lub działania na zamówienie. Uznano, że również w przyszłości takie działania są mało prawdopodobne.
3. Poniesione w przeszłości i zidentyfikowane jako możliwe w przyszłości szkody wywołane działaniami celowymi intruzów i błędami pracowników, dotyczą:
   a) wpływu na działalność społeczną i bezpieczeństwo publiczne,
   b) naruszenia prawa cywilnego i/lub karnego,
   c) wpływu na bezpieczeństwo osobowe,
   d) wpływu na wizerunek organizacji,
   e) wpływu na publiczne interesy organizacji.

Przy szacowaniu szkód wzięto pod uwagę także klasę bezpieczeństwa (KB w tabeli 13) zasobu, którego dotyczył incydent. Szacujący ryzyko posłużyli się tablicą interpretacyjną 10. Dla szkód z punktów a-e ww. listy, Dział Prawny, Dział Bezpieczeństwa i Zarząd organizacji ustalili wartości ocen jak w tabeli 13. Ocenę wypadkową (ostatnia kolumna w tabeli 13) otrzymano, posługując się algorytmem 1 – całość szacowania szkód można przedstawić następująco:

4. Specjaliści na podstawie wizji lokalnych, analizy dokumentacji oraz przeglądu plików konfiguracyjnych zabezpieczeń stwierdzili, że były zastosowane wszystkie wymagane zabezpieczenia zasobów informacyjnych, ale część z nich była źle skonfigurowana. Poza tym stwierdzono braki w nadzorze nad działaniami pracowników na zasobach wrażliwych, chociaż poziom wyszkolenia pracowników z zakresu bezpieczeństwa oceniono

[13] Specjaliści zaproponowali trójelementowy zbiór ocen: wysokie (W), średnie (S), niskie (N) i składanie ocen z wykorzystaniem algorytmu 1.

wysoko. Zakładając, że były to wszystkie stwierdzone podatności oraz że interpretacja ocen dla PZ jest dana tabelami 11 i 12, to poziom podatności dla obu typów incydentów jest na poziomie S.

Tab. 13. Oszacowania szkód ST (dla przykładu 4)

INCYDENT	RODZAJ INCYDENTU	a	b	c	d	e	KB	alg{.}
Działania celowe	ujawnienie wszystkich podmiotów, które są uprawnione do dostępu do zasobu klasyfikowanego jako „poufne”	N	N	S	N	N	S	**N**
	ujawnienie zawartości zastrzeżonego zasobu informacyjnego nieuprawnionym podmiotom	S	N	N	S	N	N	**N**
Błąd pracownika	ujawnienie zawartości poufnego zasobu informacyjnego nieuprawnionym podmiotom	S	S	N	W	S	S	**S**
	ujawnienie niektórych podmiotów, które są uprawnione do dostępu do zasobu klasyfikowanego jako „tajne”	W	W	W	N	S	W	**W**

5. Stwierdzono, że w przeszłości zdarzały się przypadki (błędy) udostępnienia zawartości zasobu informacyjnego klasyfikowanego jako *poufne* nieuprawnionym osobom. Stwierdzono także, że w ciągu minionych sześciu lat dwukrotnie ujawniono na skutek błędu pracownika osobom nieuprawnionym, kto ma dostęp do zasobu informacyjnego klasyfikowanego jako *tajne*. Przy szacowaniu wartości MRZ dla tych przypadków posłużono się tabelami 6 i 8.
6. Dla działań celowych możliwość realizacji zagrożenia MRZ oszacowano na podstawie zbioru cech ZC (patrz przykład 3). Zbiór wartości cech ustalono na {średni, lokalny, duży, „telekomy”}, co przekłada się na zbiór ocen {S, N, W, W}, skąd ocena wypadkowa alg{S, N, W, W} = S.
7. Wyniki szacowania ryzyka są zamieszczone w tabeli 14.

Tab. 14. Oszacowania ryzyka (dla przykładu 4)

ZAGROŻENIE	RODZAJ INCYDENTU	MRZ	PZ	ST	RYZYKO
Działania celowe	ujawnienie wszystkich podmiotów, które są uprawnione do dostępu do zasobu klasyfikowanego jako „poufne”	S	S	N	**S** (R'_{515})
	ujawnienie zawartości zastrzeżonego zasobu informacyjnego nieuprawnionym podmiotom	S	S	N	**S** (R'_{515})
Błąd pracownika	ujawnienie zawartości poufnego zasobu informacyjnego nieuprawnionym podmiotom	W	S	S	**S** (R'_{205})
	ujawnienie niektórych podmiotów, które są uprawnione do dostępu do zasobu klasyfikowanego jako „tajne”	W	S	W	**W** (R'_{204})

******** (koniec przykładu)

4. Metody składania ocen opisowych i interpretacja ryzyka

Analityk ma możliwość wybrania dowolnego, byle poprawnego formalnie, sposobu składania ocen – nie ma obecnie norm, standardów i przepisów, które by jednoznacznie narzucały czy w inny sposób regulowały to zagadnienie. W wielu zastosowaniach (patrz np. NIST SP 800-53 [11]) rekomenduje się, ze względu na prostotę oraz to, że dla wielu praktycznych problemów jest wystarczająca, formułę max{OCENA}, która wybiera jako wypadkową ocenę maksymalną ze zbioru składanych ocen. Wadą takiego sposobu składania ocen jest migrowanie ocen wypadkowych w kierunku ocen najwyższych, odwrotnie niż przy formule min{OCENA}, gdzie oceny wypadkowe migrują w kierunku oceny najniższej – problem ten przedstawia zawartość tabeli 15.

Niech:

$\otimes = \max\{ocena_1, \ldots, ocena_j, \ldots, ocena_k\}$ dla $j \in [1,k]$ gdzie: $ocena_j \in OCENA$ (3)

$\varnothing = \min\{ocena_1, \ldots, ocena_j, \ldots, ocena_k\}$ dla $j \in [1,k]$ gdzie: $ocena_j \in OCENA$ (4)

$\Sigma = alg\{ocena_1, \ldots, ocena_j, \ldots, ocena_k\}$ dla $j \in [1,k]$ gdzie: $ocena_j \in OCENA$ (5)

gdzie:

- alg{…} oznacza składanie ocen zgodnie z algorytmem 1;
- ⊗, ∅, ☼ to symbole składania ocen opisowych według określonych formuł lub algorytmów.

Tab. 15. Wartości wynikowe przy składaniu dwóch ocen opisowych różnymi metodami

Lp.	A	B	C=A⊗/∅B
1	W	W	**W** W
2	W	S	**W** S
3	W	N	**W** **N**
4	S	W	**W** S
5	S	S	S S
6	S	N	S **N**
7	N	W	**W** **N**
8	N	S	S **N**
9	N	N	N **N**

Dalsze rozważania o szacowaniu ryzyka odnoszą się do wyników uzyskanych w rezultacie stosowania algorytmu 1 (patrz kolumna nr 6 w tabeli 16). Z tabeli 16 można wyciągnąć następujące wnioski dotyczące teoretycznych możliwości minimalizacji ryzyka:

1. Trzy możliwości minimalizacji wartości ryzyka (poprzez oddziaływanie na MRZ, PZ i ST) istnieją dla zbioru ryzyka:

   $\{R'_{101}, R'_{102}, R'_{204}, R'_{205}, R'_{410}, R'_{411}, R'_{513}, R'_{514}\}$

2. Tylko dwie możliwości minimalizacji wartości ryzyka istnieją dla zbiorów ryzyka:
   - minimalizacja MRZ i ST: $\{R'_{307}, R'_{308}, R'_{616}, R'_{617}\}$
   - minimalizacja MRZ i PZ: $\{R'_{103}, R'_{206}, R'_{412}, R'_{515}\}$
   - minimalizacja ST i PZ: $\{R'_{719}, R'_{720}, R'_{822}, R'_{823}\}$
3. Tylko jedna możliwość minimalizacji wartości ryzyka istnieje dla zbiorów ryzyka:
   - minimalizacja MRZ: $\{R'_{309}, R'_{618}\}$
   - minimalizacja PZ: $\{R'_{721}, R'_{824}\}$
   - minimalizacja ST: $\{R'_{925}, R'_{926}\}$

   - Brak możliwości minimalizacji ryzyka (wartości wszystkich składowych, tj. MRZ, PZ, ST są na poziomie niskim, czyli ryzyko jest minimalne) występuje dla $\{R'_{927}\}$

Tab. 16. Oszacowania wartości ryzyka z użyciem operacji składania ⊗ (kolumna 5) **oraz operacji składania ☼** (kolumna 6)

Lp.	MRZ	PZ	ST	RYZYKO ⊗	RYZYKO' ☼
1	2	3	4	5	6
1	W	W	W	$R_{101} \rightarrow W$	$R'_{101} \rightarrow W$
			S	$R_{102} \rightarrow W$	$R'_{102} \rightarrow W$
			N	$R_{103} \rightarrow W$	$R'_{103} \rightarrow S$
2	W	S	W	$R_{204} \rightarrow W$	$R'_{204} \rightarrow W$
			S	$R_{205} \rightarrow W$	$R'_{205} \rightarrow S$
			N	$R_{206} \rightarrow W$	$R'_{206} \rightarrow S$
3	W	N	W	$R_{307} \rightarrow W$	$R'_{307} \rightarrow S$
			S	$R_{308} \rightarrow W$	$R'_{308} \rightarrow S$
			N	$R_{309} \rightarrow W$	$R'_{309} \rightarrow S$
4	S	W	W	$R_{410} \rightarrow W$	$R'_{410} \rightarrow W$
			S	$R_{411} \rightarrow W$	$R'_{411} \rightarrow S$
			N	$R_{412} \rightarrow W$	$R'_{412} \rightarrow S$
5	S	S	W	$R_{513} \rightarrow W$	$R'_{513} \rightarrow S$
			S	$R_{514} \rightarrow S$	$R'_{514} \rightarrow S$
			N	$R_{515} \rightarrow S$	$R'_{515} \rightarrow S$
6	S	N	W	$R_{616} \rightarrow W$	$R'_{616} \rightarrow S$
			S	$R_{617} \rightarrow S$	$R'_{617} \rightarrow S$
			N	$R_{618} \rightarrow S$	$R'_{618} \rightarrow N$
7	N	W	W	$R_{719} \rightarrow W$	$R'_{719} \rightarrow S$
			S	$R_{720} \rightarrow W$	$R'_{720} \rightarrow S$
			N	$R_{721} \rightarrow W$	$R'_{721} \rightarrow S$
8	N	S	W	$R_{822} \rightarrow W$	$R'_{822} \rightarrow S$
			S	$R_{823} \rightarrow S$	$R'_{823} \rightarrow S$
			N	$R_{824} \rightarrow S$	$R'_{824} \rightarrow N$
9	N	N	W	$R_{925} \rightarrow W$	$R'_{925} \rightarrow S$
			S	$R_{926} \rightarrow S$	$R'_{926} \rightarrow N$
			N	$R_{927} \rightarrow N$	$R'_{927} \rightarrow N$

Uwagi do tabeli 16:

1. Liczba w miejscu yy w indeksie dolnym R_{xyy} jest liczbą porządkową ryzyka.
2. Liczba w miejscu x w indeksie dolnym R_{xyy} jest numerem wiersza w tabeli 16.
3. Apostrof w symbolu R'_{xyy} oznacza, że ryzyko było szacowane przy użyciu operacji ☼. Brak apostrofu oznacza, że ryzyko było szacowane przy użyciu operacji ⊗.

Przyjmując za ryzyko akceptowalne ryzyko o wartości N, to dla użytego sposobu składania ocen zbiór ryzyka akceptowalnego tworzą elementy: {R'_{618}, R'_{824}, R'_{926}, R'_{927}}. Jednak dla elementów {R'_{618}, R'_{824}, R'_{926}} istnieją jeszcze możliwości zmniejszenia wartości ryzyka – patrz zacieniowane wartości w punkcie 3. Sytuacja taka nie występuje przy szacowaniu wartości ryzyka według formuły max{.} – w takim przypadku jest tylko jedno ryzyko akceptowalne, którego wszystkie składowe mają wartość N (R_{927} w tabeli 16).

Czasami dla podjęcia decyzji o sposobach minimalizacji ryzyka potrzebna jest wiedza o tym, co wpływa na podwyższenie ryzyka lub patrząc na zagadnienie inaczej, które elementy składające się na ryzyko są na poziomie niskim i można się nimi nie zajmować. I tak, na podstawie zawartości tabeli 16 można stwierdzić, że:

1. Zbiór wartości ryzyka, dla którego poziom możliwości realizacji zagrożenia jest niski, tj. ocena{MRZ} = N, tworzy dziewięć elementów:

   $\{R'_{719}, R'_{720}, R'_{721}, R'_{822}, R'_{823}, R'_{824}, R'_{925}, R'_{926}, R'_{927}\}$

2. Zbiór wartości ryzyka dla którego poziom podatności jest niski, tj. ocena{PZ} = N, tworzy dziewięć elementów:

   $\{R'_{307}, R'_{308}, R'_{309}, R'_{616}, R'_{617}, R'_{618}, R'_{925}, R'_{926}, R'_{927}\}$

3. Zbiór wartości ryzyka, dla którego poziom szkód jest niski, tj. ocena{ST} = N, tworzy dziewięć elementów:

   $\{R'_{103}, R'_{206}, R'_{309}, R'_{412}, R'_{515}, R'_{618}, R'_{721}, R'_{824}, R'_{927}\}$

PRZYKŁAD 5

Odnosząc rozważania z tego rozdziału do treści PRZYKŁADU 4 (patrz tabela 14, ostatnia kolumna), można dla zidentyfikowanych incydentów zalecić następujące sposoby minimalizacji ryzyka:

1. Incydent: *Ujawnienie wszystkich podmiotów, które są uprawnione do dostępu do zasobu klasyfikowanego jako „poufne”* – **R'_{515}, minimalizacja MRZ i PZ**.

   W zakresie MRZ: nie da się osłabić atrakcyjności obiektu ataku dla intruza. Można jedynie wpłynąć na osłabienie jego motywacji poprzez ustalenie wysokich kar i skuteczność ścigania tego typu przestępstwa, ale takie przedsięwzięcia są poza zakresem kompetencji zwykłych organizacji – wymagają działań na szczeblu administracji rządowej i (zwykle) zmian w prawie.

   W zakresie PZ: z treści PRZYKŁADU 4 wynika (patrz punkt 4 przykładu – ustalenia specjalistów) że stwierdzoną podatnością była *zła konfiguracja zabezpieczeń*. Zalecane działanie: **poprawić konfigurację zabezpieczeń**.

2. Incydent: *Ujawnienie zawartości zastrzeżonego zasobu informacyjnego nieuprawnionym podmiotom* – **R'_{515}, minimalizacja MRZ i PZ**.

   W zakresie MRZ: komentarz jak w punkcie 1.

   W zakresie PZ: komentarz jak w punkcie 1.

3. Incydent: *Ujawnienie zawartości poufnego zasobu informacyjnego nieuprawnionym podmiotom* – **R'_{205}, minimalizacja MRZ, PZ i ST**.

W zakresie MRZ: poprawić system kontroli działań na zasobach wrażliwych, udoskonalić szkolenia dla osób mających dostęp do informacji wrażliwych (pomimo że oceniono je jako dobre!), zweryfikować zasady dopuszczania pracowników do pracy z informacją wrażliwą.

W zakresie PZ: z treści PRZYKŁADU 4 wynika (patrz punkt 4 przykładu – ustalenia specjalistów), że stwierdzoną podatnością był *brak właściwego nadzoru nad działaniami pracowników na zasobach wrażliwych*. Zatem zalecane działanie: **poprawić nadzór nad działaniami pracowników na zasobach wrażliwych**.

W zakresie ST: incydent ma wpływ na *działalność społeczną i publiczne bezpieczeństwo, naruszenia prawa cywilnego i/lub karnego, publiczne interesy organizacji*. Minimalizacja tych szkód wymaga skoordynowanych działań Zarządu organizacji, jej prawników oraz osób odpowiedzialnych za PR.

4. Incydent: *Ujawnienie niektórych podmiotów, które są uprawnione do dostępu do zasobu klasyfikowanego jako „tajne”* – **R'_{204}, minimalizacja MRZ, PZ i ST**.

   W zakresie MRZ: komentarz jak w punkcie 3.

   W zakresie PZ: komentarz jak w punkcie 3.

   W zakresie ST: incydent ma wpływ na *działalność społeczną i publiczne bezpieczeństwo*, naruszenia *prawa cywilnego i/lub karnego, bezpieczeństwo osobowe, publiczne interesy organizacji*. Minimalizacja tych szkód wymaga skoordynowanych działań Zarządu organizacji, jej prawników oraz osób odpowiedzialnych za PR. Poza tym Dział Bezpieczeństwa organizacji powinien objąć ochroną osobistą osoby mające dostęp do informacji klasyfikowanych jako „tajne”.

******** (koniec przykładu)

5. Podsumowanie

W artykule przedstawiono sposób szacowania ryzyka niepożądanej zmiany kryterium jakości informacji, jakim jest tajność, czyli szacowania ryzyka wystąpienia pewnej klasy incydentów z zakresu bezpieczeństwa informacyjnego. Wiedza o ryzyku – jego wartości, wartości elementów składowych oraz ich praktycznego znaczenia (interpretacji) jest podstawą zarówno budowania systemu zabezpieczeń (minimalizacji ryzyka), jak i działań w zakresie obsługi incydentów wywołanych realizacją zagrożeń, dla których było oszacowane ryzyko.

W przypadku szacowania ryzyka (lub szerzej – analizy ryzyka) z zakresu bezpieczeństwa informacyjnego dla konkretnych organizacji, zwykle oszacowania dotyczą tajności, integralności i dostępności zasobów informacyjnych. W tym artykule opisano propozycję takiego oszacowania dla tajności zasobów informacyjnych. Przykład oszacowania ryzyka dla dostępności jest zamieszczony w rozdz. 3.4 w [1]. Oszacowania ryzyka dotyczące integralności zasobu informacyjnego będą przedmiotem odrębnego artykułu.

Literatura

1. LIDERMAN K., *Bezpieczeństwo informacyjne*. *Nowe wyzwania*. PWN. 2017.
2. MALIK A., *Propozycja doboru i składania ocen opisowych w jakościowym szacowaniu ryzyka systemów informacyjnych*. Praca dyplomowa. Politechnika Warszawska. Podyplomowe Studium Bezpieczeństwa Systemów Informatycznych. 2011.
3. Dyrektywa Parlamentu Europejskiego i Rady (UE) 2016/1148 z dn. 6.07.2016 r. *w sprawie środków na rzecz wysokiego wspólnego poziomu bezpieczeństwa sieci i systemów informatycznych na terytorium Unii* (NIS).
4. ISO/IEC WD 29115:2019 Information technology – Security techniques – *Entity authentication assurance Framework*.
5. PN-ISO/IEC 27005:2010 – Technika informatyczna – Techniki bezpieczeństwa – *Zarządzanie ryzykiem w bezpieczeństwie informacji*.
6. Rozporządzenie Rady Ministrów z dn. 31 października 2018 r. *w sprawie progów uznania incydentu za poważny*. Dz. U. poz. 2180.
7. Rozporządzenie Rady Ministrów z dn. 11 września 2018 r. *w sprawie wykazu usług kluczowych oraz progów istotności skutku zakłócającego incydentu dla świadczenia usług kluczowych*. Dz. U. poz. 1806.
8. Rozporządzenie Prezesa Rady Ministrów z dn. 20 lipca 2011 r. *w sprawie podstawowych wymagań bezpieczeństwa teleinformatycznego.* Dz. U. z 2011 r. nr 159, poz.948.
9. Ustawa z dn. 02.08.2010 *o ochronie informacji niejawnej.* Dz. U. 182/10 poz. 1228.
10. Ustawa z dn. 05.07.2018 r. *o krajowym systemie cyberbezpieczeństwa.* Dz. U. poz. 1560.
11. SP-800-53 Rev.4: *Recommended Security Controls for Federal Information System*. April 2013.

Risk of undesired changes to significant information quality criteria

ABSTRACT: The paper presents a method of estimating the risk of an undesirable change in the information quality criterion of secrecy, meaning estimating the risk of a certain class of information security incidents. The qualitative risk estimation method is adopted and the impact of a descriptive grade composition method on the results is discussed. Considerations on the possibilities of interpreting variables used in the risk estimation and establishing the range of their actual values were also presented. Additionally, the paper describes how the identified range of actual variable values translates into levels used in the risk estimation.

KEYWORDS: incident, risk estimate, information security

Praca wpłynęła do redakcji: 12.11.2019 r.

Experimental research on the impact of similarity function selection on the quality of keyword spotting in speech signal

Łukasz LASZKO

Institute of Teleinformatics and Cybersecurity, Faculty of Cybernetics, MUT
ul. gen. Sylwestra Kaliskiego 2, 00-908 Warsaw, Poland
lukasz.laszko@wat.edu.pl

ABSTRACT: The paper describes an evaluation of the application of selected similarity functions in the task of keyword spotting. Experiments were carried out in the Polish language. The research results can be used to improve already existing keyword spotting methods, or to develop new ones.

KEYWORDS: keyword spotting, signal similarity, quality of detection, dynamic time warping, textual query

1. Introduction

The task of keyword spotting (KWS) consists of query-by-example[1] in the registered spontaneous speech signal. The purpose of the task is achieved by indicating the points in the speech signal where the given word occurs. These indications should usually minimise the probability of false peace and false alarm [22].

The task of KWS is part of the field known as *information retrieval* [50][2]. In this field, it is defined as follows:

a) a speech signal, which is by definition generated by different speakers,

b) the searched word that is set in text form,

[1] The following terms are also used: *keyword or key-word spotting*, *key-phrase detection* [74] or *spoken term detection* [59].

[2] Specifically in the field of sound KWS is sometimes considered part of *Audio IR* [15], *Multimedia IR* [63], [56]. Yet another view is presented in [29].

c) the reference signal which is obtained by converting text-to-speech by using recordings of natural speakers or by speech synthesisers,
d) pattern search in the speech signal which is based on comparing the tested signal with the reference signal,
e) the comparison that applies to signals, not text (string of phonetic symbols).

One of the essential problems to solve is determining the similarity between the models of two signals: utterance and reference signal (the so-called query) [17]. An analysis of publications from the last twenty years has allowed the author to observe that usually this similarity is established in the metric space of speech signal features R^N. The features applied are acoustic coefficients such as mel-frequency cepstral coefficients (MFCC). The assessment of similarity between signal models is based on the distance between them in R^N, with the shorter distance meaning greater similarity. The most commonly applied metric in KWS tasks is the cosine metric [28], [77], [68].

The choice of metric is usually arbitrary and not discussed in publications by researchers. As noted in [17], this may be caused by the properties of the metric itself. However, significant differences in interpretation occur for Euclidean and cosine metrics, for example. This has had an impact on the direction of research described herein.

The purpose of the author's research was to determine the impact of similarity function selection on the quality of keyword spotting in speech signal. This article describes the results of comparative research obtained by the author for using the keyword spotting method introduced in paper [42]. The research was conducted for the Polish language analogously to the research reported in [44], using the same corpus of Polish speech [35].

2. Similarity of words in a speech signal

2.1. Similarity assessment methods

The following approaches can be distinguished for setting the similarity of two speech signals [64], [27][3]:

[33] Own study on [64] pp. 190-193, [27] pp. 22-37. Other classification of approaches is show in [1], for example.

Categorical (ontological) similarity – making an assessment based on a classification according to known conceptual categories (e.g. voiced sound).

Similarity of attributes – where analysed words have identical or similar features (properties), and the numerical values of the features show slight differences (i.e. are similar), such as formant frequencies.

Similarity of relations – where there are identical or similar relations, such as proportions, between the analysed words.

Similarity of causal (semantic) relations – where the analysed words have the same (similar) contexts, e.g. given words define the same subject in a sentence.

In the case of keyword spotting tasks, similarity is usually set according to speech signal attributes (i.e. similarity of attributes). Such attributes (speech signal features) are most often acoustic coefficients, such as: MFCC [55], human-factor cepstral coefficients (HFCC) [74], relative spectral-perceptual linear prediction (RASTA-PLP) [71], [32], [19] and others, referred to in paper [53], for example. The issue of selecting the similarity function may depend on the adopted features that represent the compared signals.

2.2. Similarity assessment

The solution of KWS task can be approached in two ways: using speech recognition methods [72] or speech processing methods [59].

Through the use of speech recognition methods, proper keyword spotting is done in the sphere of text (string of phonetic symbols) obtained by analysing words from the recording. Determining the similarity of words then comes down to calculating the distance between the strings of symbols, based on the Levenshtein distance, for example, as in paper [79]. In this case, the word with the lowest Levenshtein distance from the textual query is indicated.

Other measures are used instead of the Levenshtein distance, such as:

- Damerau-Levenshtein distance [4],
- Jaro-Winkler distance [70],
- Hamming distance [75] and
- LCS (longest common subsequence) [42].

When speech processing methods are used, keyword spotting is done in the sphere of signal. The speech signal for the given textual query is obtained through *text-to-speech* synthesis. The resulting signal sample vector is converted into a feature vector. Further, depending on the signal model, there are the following approaches to assessing word similarity:

1) If the signal representation is a single vector (e.g. MFCC), the similarity assessment is based on:
   a) distances between vectors, typically a cosine distance, although other distances are also used, such as:
      - Euclidean distance [34], [25],
      - cosine-Euclidean distance [22],
      - log-cosine distance [18],
      - Manhattan distance [20],
      - sigma distance[18],
   b) correlation coefficient (with zero meaning no similarity); typically this is the Pearson correlation, although Kendall or Spearman correlations are also used[4] [33], [48], [39].
2) If the signal model is a group (cluster) of vectors (such as a set of frame group features), inferring about the similarity of two signals requires defining similarity between clusters. The similarity assessment is based on the distance between clusters, while the *distance* understood in this way does not usually meet the metrics axiom[5]. The following approaches are used in this case:
   a) setting the distance based on cluster elements (e.g. between central elements of clusters), for which Euclidean distance or other distances based on Minkowski distance [67] are applied,
   b) setting the distribution of elements in the cluster based on the distance, including a probabilistic model, for which the Kullback-Leibler distance is often used [26], [30], even though others are also used, such as:
      - Bhattacharyya distance [1], [16], [5], [3], [31],
      - Mahalanobis distance [3], [38],
      - Hellinger distance [45], [23], [31], [58] and
      - divergences: *f-divergence*, Jensen–Shannon divergence, etc. [57], [62].

This article describes research in relation to the latter approach, i.e. speech processing methods are used to solve the KWS task (cf. [42]), and the research task is to choose the similarity function.

[4] Also known as rank correlation. Ranks are the numbers of subsequent observations in the ordered statistical sample.

[5] Cf. e.g. [78] p. 39.

2.3. Similarity function assessment

In Table 1 there is a list of similarity functions used in the described research. The similarity function is one of the important components of the methods applied in KWS tasks and has a direct impact on the quality of keyword spotting. It is therefore appropriate to apply the similarity function which results in the highest quality results when used in a particular method.

Tab. 1. List of similarity functions tested

No.	Basis for defining the similarity function[6]:
1	Bhattacharyya distance (K_{bha})
2	Chebyshev distance (K_{che})
3	correlation-based distance (K_{cor})
4	cosine distance (K_{cos})
5	Euclidean distance (K_{euc})
6	Hellinger distance (K_{hel})
7	symmetrical Kullback–Leibler distance(K_{skl})
8	Manhattan distance (K_{man})
9	Mahalanobis distance (K_{mah})
10	Minkowski distance (K_{min})
11	standardized Euclidean distance (K_{seu})
12	Spearman distance (K_{spr})

2.3.1. Indicators of keyword spotting quality in KWS tasks

The quality of spotting can be measured using basic indicators directly related to the number of results achieved [61]. These include:

- TP (true positive) – number of correct indications (hits),
- TN (true negative) – number of correct rejections,
- FP (false positive) – number of incorrect indications (Type I errors – 'false alarms'),
- FN (false negative) – number of incorrect rejections (Type II errors, misses – 'false peace'),

6 The similarity function symbol is put in brackets.

These indicators are often set into an error/confusion table/matrix [69][7]. Precision of indications and other indicators that allow for referencing the results obtained (e.g. to compare two methods) are also important in KWS tasks. These include derived indicators. The following indicators were selected for the research:

- **precision**, marked PPV,
- **accuracy**, marked ACC,
- **recall**, **true positive rate**, marked TPR,
- **specificity**, **true negative rate**, marked TNR,
- **F-measure**, **F_1Score**, marked F_1S) [9], [65] and
- **Youden's J statistic**, marked YJS [73].

Based on the PPV it can be assessed whether a given method (using a given similarity function) gives repeatable results, characterized by a small spread. The ACC value makes it possible to assess whether a given method always gives results close to true (real) results. The TPR indicates the ability of the method to correctly detect (indicate a result) where the value sought actually exists. On the other hand, the TNR specifies the ability of the method to correctly reject results (the so-called selectivity). F_1Score is used to assess the method *reliability*, i.e. a feature demonstrating the authenticity of the results obtained (both indications and rejections). However, the YJS is used to assess the method effectiveness[8] and to select the best method parameters in the ROC analysis (cf. Chapter 5.1).

2.3.2. Vector assessment scalarization

It is assumed in the paper that the vector assessment of the similarity function will be made using six derivative indicators listed above. It should be noted that the indicators described above have the same range of values. It is a number range $[0,1]$, where an indicator value of *one* characterises a good method (which is the most precise, most accurate, etc.).

A scalar assessment was made by adding the best results of each quality indicator to arrange the vector assessments in order and at the same time select the best function. The above assumptions result from the author's observation that these results strictly depend on the experiment conditions. In particular, in

[7] Based on: https://en.wikipedia.org/wiki/Confusion_matrix (visited: 19.08.2019).

[8] The method effectiveness, shown by the YJS, indicates its sensitivity when false results exist in the set of results obtained by the given method.

the case of high variability of the tested material, there is insufficient justification for statistical quality assessment, e.g. the number of slots significantly depends on the detected word. Therefore, the 'competition' method was adopted. It consists in assessing the tested function through the best result obtained (in the whole research series).

3. Research experiment

The research consisted in using the method presented in paper [42]. This method is aimed at the use of patterns derived from the TTS synthesizer; such patterns were the main focus of interest. Research was conducted for the Polish language, the CLARIN-PL Mobile Corpus (EMU) [35], to the extent and as per the procedure described in paper [44]. Table 2 shows the values of the method parameters unchanged in relation to [44] and changed values adopted for the similarity functions not tested in paper [44].

For comparative purposes, additional tests were carried out using patterns from real speech recordings. They are marked in the results as *real*.

4. Results

4.1. Basic quality indicators

The results of 120 tests are presented as charts and tables. The main results are the number indicators obtained directly from the experiment: TP, TN, FP, FN. They were the basis for determining the derived indicators described above.

Table 3 presents sample test results when the similarity function was based on the Bhattacharyya distance. The values in the table, in the following lines, present the results for the query extracted from the real speech recording (*real*) and the synthesized textual query (*TTS*). The number of analysis slots, designated as *Slots*, is the number of all units the method extracted in the analysed speech signal. The number depends on the query length, hence its difference in test for the same session. The slot is not an analysis window, but the length of the pattern sought (cf. Table 2).

Other test results (for other similarity functions) are presented in a cumulative manner in Figures 1 and 2.

Tab. 2. Parameters of the KWS method used in the described tests

	Parameter name	Parameter values											
Unchanged values	Number of FFTs	8192											
	Analysis window size	1024											
	Overlap percentage	33%											
	Number of HFCCs	15											
	Signal frequency range	[300, 3400]											
	Query length rate	1.5											
	Query match rate	0.5											
	Path threshold value	0.6											
Changed values	Similarity measurement method[9]	*bha*	*che*	*cor*	*cos*	*euc*	*hel*	*skl*	*man*	*mah*	*min*	*seu*	*spr*
	Normalisation method[10]	-	HE	HE	HE	HE	-	HE	HE	HE	HE	HE	HE
	Sequence threshold value (real/TTS)[11]	89/78	80/70	77/76	65/54	73/65	82/85	85/60	75/55	97/97	78/68	68/67	75/72
	Other[12]	NAN=1	NAN=1	NAN=0	NAN=0	NAN=1	NAN=1	NAN=0	NAN=1	ABS, NAN=1	NAN=1	NAN=1	NAN=1

Both charts show cumulative values for all selected sessions used in speech corpus research. The charts give the opportunity to compare the results for different similarity functions. They also show that despite the lack of proper method calibration, in each case, the method results are useful, i.e. true results (TP and TN) are always in total in the majority (i.e. more than 50% of all results). Undesirable false results (FP and FN) are partly the result of the said lack of calibration, although they also show the imperfection of the method, which depends on the dependence on the data itself (i.e. recordings), as mentioned in [42]. More information on the results can be found in the derivative indicator values presented in the next section.

[9] Designations as in Tab. 01.

[10] HE – normalisation by means of histogram equalization.

[11] In papers: [42], [43] and [44] the value is defined as the recognition quality threshold. It is used after marking detected sequences as suspicious, i.e. after applying the *path threshold*, which is clearly shown in paper [42].

[12] NAN – interpretation of non-numeric values, ABS – absolute value.

Tab. 3. Test results for ten selected recording sessions using the speech corpus. The similarity function is based on the Bhattacharyya distance

		1	2	3	4	5	6	7	8	9	10
Real	**Slots**	80	56	88	56	53	40	50	55	48	77
	TP	22	10	25	12	16	12	10	26	12	26
	FP	17	14	32	29	3	13	30	13	29	37
	TN	36	28	28	15	29	14	10	14	7	14
	FN	5	4	3	0	5	1	0	2	0	0
TTS	**Slots**	43	26	39	26	38	26	36	36	29	53
	TP	21	12	22	10	21	14	10	24	10	27
	FP	6	4	7	10	6	7	14	9	18	23
	TN	13	9	7	5	9	3	2	3	1	3
	FN	3	1	3	1	2	2	0	0	0	0

4.2. Quality indicators obtained

Table 4 shows an example of indicator values for the results obtained in the tests for the Hellinger distance-based similarity function. The row for the sensitivity rate (TPR) is marked in the table. It shows the ability of the method to detect (indicate a result) where the value sought actually exists. Values close to one demonstrate the high sensitivity of the classifier. In the presented case, there were sessions for which virtually all searched words were found with a small percentage of false rejections (TN).

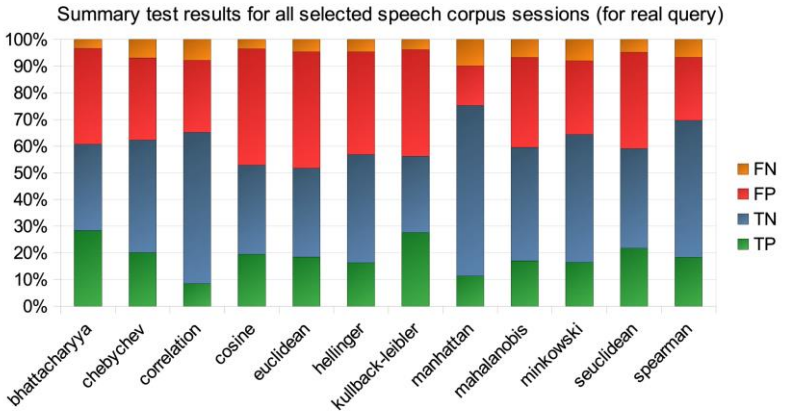


Fig. 1. Results for the real query – percentage value

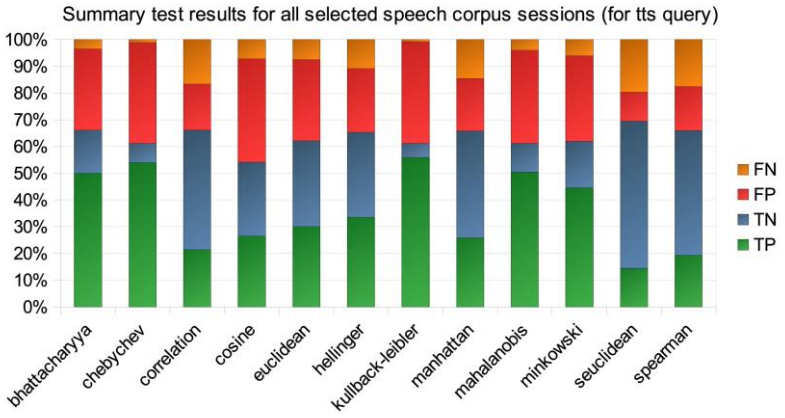


Fig. 2. Results for the TTS query – percentage value

Mean values: $\overline{TPR_{real}} = 0{,}74$, $\overline{TPR_{TTS}} = 0{,}75$, i.e. for the so-called *average case*, show that this similarity function can be successfully used in a situation where the researcher is primarily interested in maximizing the number of detections (true indications, TP), completely ignoring false positive (FP) values.

Tab. 4. Quality indicators for the method using the Hellinger distance-based similarity function. The results of 10 test sessions are presented

		1	2	3	4	5	6	7	8	9	10
Real	**PPV**	0.55	0.10	0.43	0.14	0.38	0.33	0.08	0.54	0.13	0.45
	ACC	0.71	0.60	0.66	0.41	0.62	0.66	0.38	0.66	0.43	0.59
	TPR	**0.89**	**0.57**	**0.80**	**0.73**	**0.69**	**0.77**	**0.57**	**0.85**	**0.75**	**0.77**
	TNR	0.62	0.60	0.61	0.36	0.60	0.64	0.36	0.55	0.39	0.49
	F1S	0.68	0.17	0.56	0.24	0.49	0.47	0.14	0.66	0.22	0.57
	YJS	0.51	0.17	0.41	0.09	0.29	0.41	-0.07	0.39	0.14	0.26
TTS	**PPV**	0.67	0.88	0.94	0.47	0.83	0.64	0.36	0.70	0.32	0.51
	ACC	0.70	0.81	0.82	0.59	0.82	0.62	0.42	0.72	0.38	0.57
	TPR	**0.56**	**0.64**	**0.74**	**0.70**	**0.67**	**0.64**	**0.89**	**0.84**	**0.89**	**0.96**
	TNR	0.80	0.93	0.94	0.53	0.91	0.58	0.18	0.59	0.15	0.24
	F1S	0.61	0.74	0.83	0.56	0.74	0.64	0.52	0.76	0.47	0.67
	YJS	0.36	0.57	0.68	0.23	0.58	0.23	0.07	0.43	0.04	0.20

For the other functions, the calculated values of indicators are presented graphically. The first summary shows PPVs and ACCs (Fig. 3). Four similarity functions were selected, for which the mean indicators were the highest. They should be analysed simultaneously, as then they can indicate the possible direction of the detection method calibration. Based on these results, it can be stated that the KWS method applied is accurate, as the ACC obtained quite high values, and at the same time they are characterised by a low spread (which can be seen in charts c and d). At the same time, the method is not very precise, i.e. for some of the analysed recordings it does not detect the fragments it should detect (low PPV), and detects it for others (PPV close to one) – charts a and b).

Figure 3 b) shows that the PPV set using the Bhattacharyya distance-based similarity function is not characterised by such a big difference in value in subsequent tests (for other data) than *better* function based on Spearman's correlation at some points. This demonstrates that the first similarity function is less dependent on the specific data used in the test, and therefore the robustness of the whole spotting method is higher.

The level of *reliability* to the applied detection method can be concluded based on the second summary (Fig. 4). The *TTS* synthesised query was used in the tests. In this case, a *reliable* method is understood as one that does not maximise the number of false results, but detects and rejects what it should, according to the facts.

The third summary (Fig. 5) shows the calculated Youden's J statistics for the average and maximum cases. The results obtained are presented in an orderly manner relative to the mean value. The best similarity functions, as per the indicator, are those based on Spearman, Bhattacharyya and Manhattan distances.

4.3. Qualitative assessment of similarity function

The similarity function ranking shown below in Table 5 is a summary of the tests described in the article. It was based on qualitative assessment for all test samples, as per the method described in item 2.3.2. The final result presented in the table was obtained through the previously described scalarization. The test results for *real* query are also included for comparison.

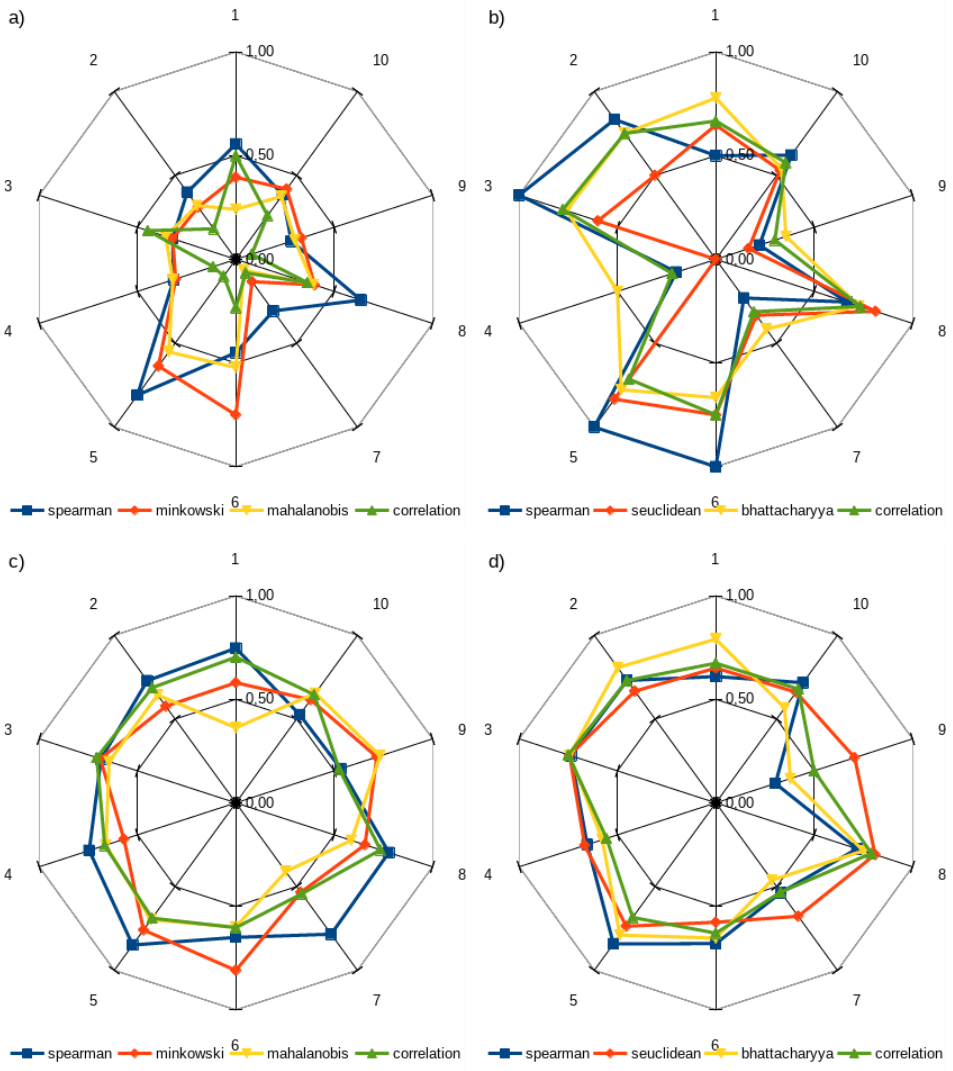


Fig. 3. Summary of PPV and ACC indicators for selected similarity functions: a) PPV for *real* query, b) PPV for *TTS* query, c) ACC for real query, d) ACC for *TTS* query; the results were obtained in subsequent test sessions (1 to 10)

Fig. 4. Summary of indicators demonstrating the *reliability of* the detection method. The *TTS* synthesised query was used in the tests

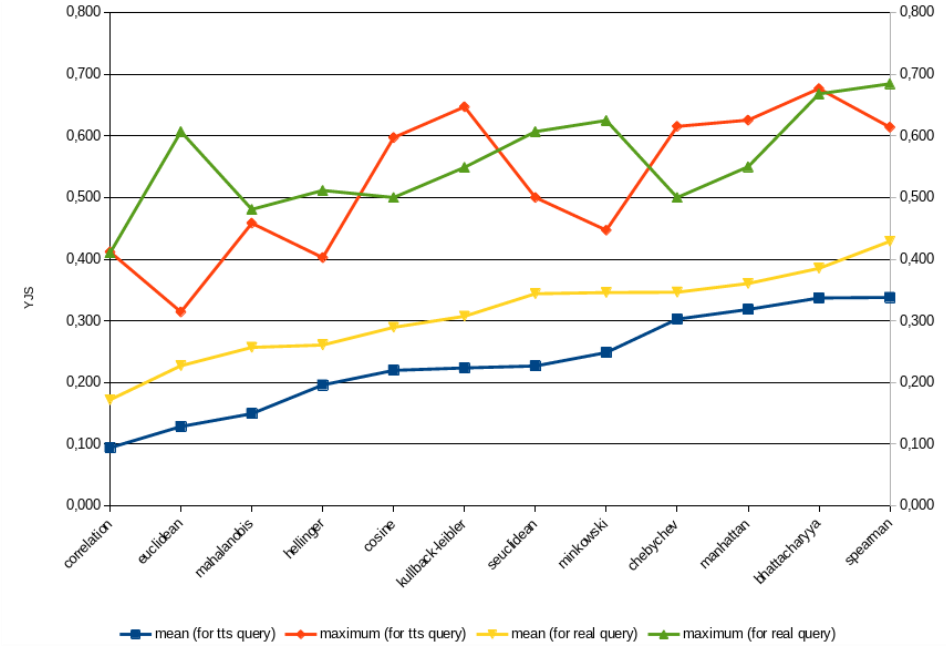


Fig. 5. List of Youden's J statistics (YJS). Blue squares and yellow triangles show average cases. Red diamonds and green triangles show best cases

5. Additional tests

5.1. ROC curve analysis

The numerical indicators used to select the signal similarity function describe only a certain momentary state of test. To learn how the keyword spotting method behaves in a wider range, a Receiver Operating Characteristic curve analysis was conducted [14], [60]. The analysis was carried out only for the selected (best) Spearman similarity function. The ROC curve is made as a set of indicating TPR and FPR values, obtained for several tests and repeated at different threshold values (see Fig. 6). Where:

$$FPR = 1 - TNR \tag{1}$$

is a fallout, false positive rate.

Tab. 5. Similarity function ranking

	TTS		real	
No.	**Similarity function**	**Indicator rating**	**Similarity function**	**Indicator rating**
1	K_{spr} (Spearman)	5.175	K_{bha} (Bhattacharyya)	5.066
2	K_{hel} (Hellinger)	5.167	K_{spr} (Spearman)	5.028
3	K_{man} (Manhattan)	5.154	K_{min} (Minkowski)	4.869
4	K_{cor} (correlation)	4.880	K_{man} (Manhattan)	4.782
5	K_{bha} (Bhattacharyya)	4.735	K_{seu} (standardized Euclidean)	4.670
6	K_{euc} (Euclidean)	4.726	K_{euc} (Euclidean)	4.537
7	K_{seu} (standardized Euclidean)	4.685	K_{che}(Chebyshev)	4.439
8	K_{min} (Minkowski)	4.556	K_{mah} (Mahalanobis)	4.046
9	K_{mah} (Mahalanobis)	4.370	K_{skl} (symmetrical Kullback-Leibler)	4.031
10	K_{skl} (symmetrical Kullback-Leibler)	4.176	K_{hel} (Hellinger)	3.971
11	K_{cos} (cosine)	4.023	K_{cos} (cosine)	3.957
12	K_{che}(Chebyshev)	3.957	K_{cor} (correlation)	3.808

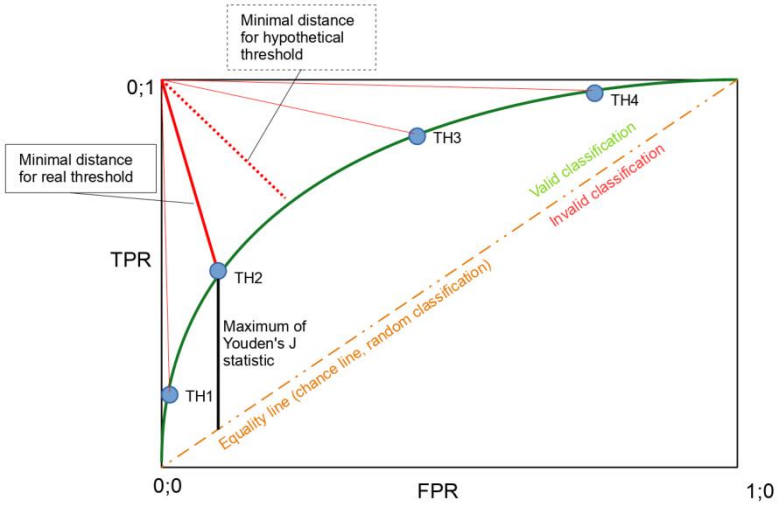


Fig. 6. Schematic representation of the ROC curve and the method of determining the threshold value. Thresholds TH 1 to 4 are applied in place of actual TPR and FPR values resulting from the measurement. TH can have any value range depending on the method. Youden's J statistic is marked similarly and its value is higher for the TH2 test than for the tests with other THs. The chart also shows hypothetically best TH value, which can be determined graphically, for example by comparing two adjacent threshold values (in this case TH2 and TH3)

The tests were conducted for the *TTS* query case. A total of 250 tests were conducted for the method presented in paper [42]. Method parameters adopted are the same as in Tab. 2; only the threshold value of the sequence is changed in the range of 50 to 98 with step 2 (i.e. for 25 values of this threshold). The tests were carried out for all selected recording sessions using the analysed speech corpus. TPR and FPR values were based on the results obtained and included on charts. The charts below show:

- Fig. 7: detailed analysis of the ROC curve for the selected session, including the method of selecting the threshold value that maximises TPR and minimises FPR,
- Fig. 8: curve analyses conducted for the remaining sessions with indicated best threshold value.

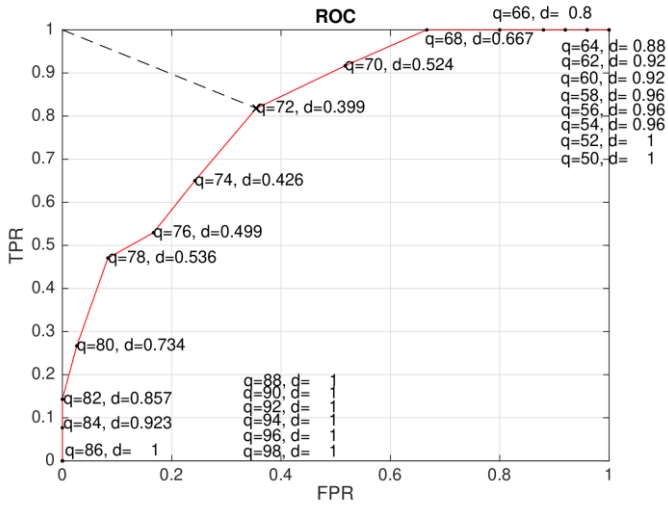


Fig. 7. ROC curve analysis for the selected session. The measurement points for the threshold value (q) are included on the chart with the determined distance value (d)

5.2. Matthews correlation coefficient

Subsequent tests were aimed to assess the impact of the selected similarity function on the random prediction of the detection method. The random prediction of the method means that it produces equally true and false results (cf. Fig 6). This is a very undesirable feature of the method, which is associated with its imperfection or lack of calibration. The Matthews correlation coefficient was

used to achieve this goal [49]. This indicator takes into account the values of all four basic indicators (cf. formula 2), and its values are interpreted as follows [8]:

- '1' perfect prediction (zero false detections and rejections),
- '-1' total disagreement (zero true values),
- '0' random prediction.

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP+FP) \cdot (TP+FN) \cdot (TN+FP) \cdot (TN+FN)}} \tag{2}$$

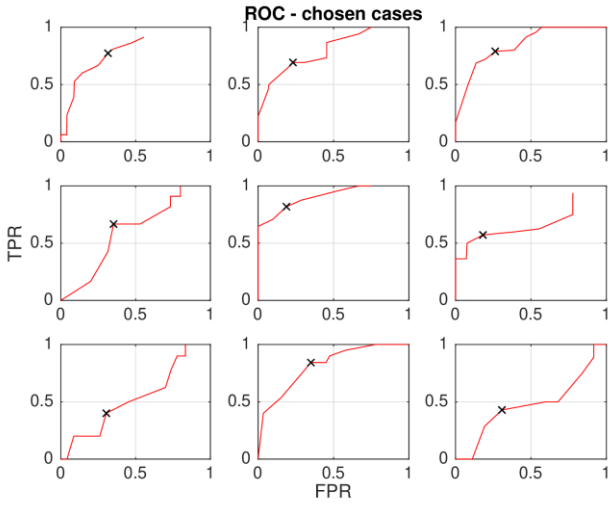


Fig. 8. Selected cases of ROC curve analysis for various sessions

Calculated values of the MCC are shown in Fig. 9. The values for other similarity functions are also included for comparison. It should be noted that the presented matrices are not correlation matrices. The MCC applies to mutual relation between true (TP, TN) and false (FP, FN) values of the method.

The test results confirmed the lack of random prediction for the detection method that uses similarity function K_{bha} and partly for the method that uses function K_{spr}.

6. Experiment conclusions

In the task of word spotting in speech signal, the choice of the signal similarity function is not obvious. The main aspect is the dependence of the

similarity function on data, i.e. recordings of speech signal and its representation. This relationship translates into the quality of detection, as observed by comparing differences in results for *real* and *TTS* queries. The selection of the similarity function may come down to indicating the function which will be the most robust to data change. In the tests conducted, such a similarity function was based on the Spearman distance (K_{spr}).

real	*bha*	*che*	*cor*	*cos*	*euc*	*hel*	*skl*	*man*	*mah*	*min*	*seu*	*spr*
1	0,47	0,26	0,38	0,47	0,02	0,49	0,20	0,18	-0,15	0,20	0,17	0,34
2	0,33	0,24	0,22	0,12	0,11	0,09	0,20	0,41	0,25	0,17	0,16	0,35
3	0,35	0,43	0,35	0,31	0,36	0,36	0,22	0,07	0,30	0,20	0,22	0,20
4	0,32	0,30	-0,01	0,14	0,12	0,06	0,34	0,42	0,33	0,38	0,29	0,41
5	0,68	0,29	-0,09	0,10	0,60	0,25	0,30	0,59	0,44	0,38	0,58	0,68
6	0,43	0,50	0,17	0,46	0,21	0,32	0,18	0,36	0,23	0,62	0,26	0,33
7	0,25	0,19	0,02	0,14	0,06	-0,04	0,36	0,22	-0,15	0,17	0,24	0,35
8	0,49	0,30	0,15	0,37	0,36	0,39	0,40	0,29	0,21	0,17	0,25	0,59
9	0,24	0,37	0,01	0,13	0,00	0,09	0,32	0,33	0,35	0,42	0,48	0,35
10	0,34	0,23	0,17	0,30	0,31	0,25	0,34	0,54	0,39	0,30	0,33	0,19

tts	*bha*	*che*	*cor*	*cos*	*euc*	*hel*	*skl*	*man*	*mah*	*min*	*seu*	*spr*
1	0,57	0,00	0,18	0,36	0,26	0,37	0,06	0,54	0,07	0,26	0,29	0,05
2	0,63	-0,13	0,36	0,13	0,16	0,61	-0,13	0,41	-0,01	0,11	0,19	0,45
3	0,42	0,43	0,45	0,42	0,20	0,67	0,35	0,34	0,20	0,41	0,27	0,35
4	0,28	0,00	-0,06	0,23	0,33	0,22	0,00	0,03	0,25	0,17	-0,17	-0,04
5	0,55	0,24	0,31	0,13	0,62	0,61	0,24	0,38	0,23	0,34	0,42	0,71
6	0,22	0,19	0,26	0,24	0,56	0,23	-0,12	0,27	0,43	0,43	0,21	0,36
7	0,23	0,18	0,11	-0,13	0,10	0,09	0,27	0,09	0,06	0,14	0,28	0,04
8	0,43	0,42	0,61	0,42	0,37	0,45	0,55	0,63	0,04	0,40	0,61	0,46
9	0,14	0,33	0,10	-0,13	0,10	0,05	0,24	0,43	0,43	0,30	-0,01	-0,10
10	0,25	0,27	0,42	0,29	0,51	0,28	0,00	0,23	0,26	0,26	0,24	0,46

Fig. 9. Summary of Matthews correlation coefficient (MCC) values. The left side of the figure shows the *real* query, the right side shows the *TTS* query

The method of choosing the best similarity function proposed in the paper is based on six quality indicators. Therefore, the selected similarity function is not assessed unilaterally.

The analysis of the ROC curve conducted as part of the additional tests showed that the detection quality can be significantly impacted by the selection of the appropriate threshold value (marked *q* in Fig. 7). It should be noted that completely bad results (i.e. more false detections and rejections than true results), using similarity function K_{spr}.

It is worth noting that the differences in the values of quality indicators obtained for different similarity functions are small. Choosing a similarity function based only on a single quality indicator value can be deceptive. Therefore, when choosing the similarity function, it is justified to carry out at least several tests for different data. The analysis of quality indicators for such tests gives more complete knowledge and it can be then expected that the chosen similarity function will give correct results for different data.

Literature

[1] AMGOUD L., DAVID V., DODER D., *Similarity Measures Between Arguments Revisited*. In: Kern-Isberner G., Ognjanović Z. (eds) Symbolic and Quantitative Approaches to Reasoning with Uncertainty, ECSQARU 2019, Lecture Notes in Computer Science, Vol. 11726, pp. 98-107, DOI 10.1007/978-3-030-29765-7_1

[2] BHATTACHARYYA A., *On a measure of divergence between two statistical populations defined by their probability distributions.* Bulletin of the Calcutta Mathematical Society, Vol. 35, 1943, pp. 99-109.

[3] BASENER W., FLYNN M., *Microscene evaluation using the Bhattacharyya distance.* Proc. of SPIE 10780, Honolulu, 2018, DOI 10.1117/12.2327004

[4] BOYTSOV L., *Indexing methods for approximate dictionary searching: Comparative analysis.* Journal of Experimental Algorithmics, Vol. 16, Article 1.1, May 2011, pp. 1-91, DOI 10.1145/1963190.1963191

[5] CHANG H.Y., *An SVM Kernel With GMM-Supervector Based on the Bhattacharyya Distance for Speaker Recognition.* IEEE Signal Processing Letters, 2009, Vol. 16, Issue 1, pp. 49-52, DOI 10.1109/LSP.2008.2006711

[6] CHEN B., WANG H.-M., CHIEN L.-F. LEE L.-S., *A*-Admissible Key-Phrase Spotting With Sub-Syllable Level Utterance Verification.* The 5th International Conference on Spoken Language Processing, Incorporating The 7th Australian International Speech Science and Technology Conference, Sydney, Australia, 1998, pp. 783-786.

[7] CHEN Y.-I., WU CH.-H., YAN G.-L., *Utterance Verification Using Prosodic Information for Mandarin Telephone Speech.* 1999 IEEE International Conference on Acoustics, Speech and Signal Processing. Keyword Spotting Proceedings, ICASSP '99, Vol. 2, Phoenix, AZ, USA, pp. 697-700, DOI 10.1109/ICASSP.1999.759762

[8] CHICCO D., *Ten quick tips for machine learning in computational biology.* BioData Mining, Vol. 10, No. 35, 2017, pp. 1-17, DOI 10.1186/s13040-017-0155-3

[9] CHINCHOR N., *MUC-4 Evaluation Metrics*. In Proceedings of the Fourth Message Understanding Conference, 1992, pp. 22-29, http://www.aclweb.org/anthology-new/M/M92/M92-1002.pdf

[10] DEB K., *Introduction to Evolutionary Multiobjective Optimization.* In: Branke J., Deb K., Miettinen K., Słowiński R. (eds) Multiobjective Optimization. Lecture Notes in Computer Science, Vol. 5252, 2008, Springer, Berlin, Heidelberg, pp. 59-96, DOI 10.1007/978-3-540-88908-3_3

[11] DUIN R.P.W., PĘKALSKA E., *The Dissimilarity Representation for Structural Pattern Recognition.* Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications, 2011, pp. 1-24, DOI 10.1007/978-3-642-25085-9_1

[12] DUIN R.P.W., PĘKALSKA E., *Non-euclidean dissimilarities: Causes and informativeness.* In proc. Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR), 2010, LNCS, Vol. 6218, Springer, Heidelberg, pp. 324-333, DOI 10.1007/978-3-642-14980-1_31

[13] DUBUISSON M.P., JAIN A.K., *A Modified Hausdorff distance for object matching.* In ICPR94, Jerusalem, Israel, 1994, pp. 566-568.

[14] FAWCETT T., *An Introduction to ROC Analysis.* Pattern Recognition Letters, Vol. 27, No. 8, 2006, pp. 861-874, DOI 10.1016/j.patrec.2005.10.010

[15] FOOTE J., *An Overview of Audio Information Retrieval.* ACM Multimedia Systems, Vol. 7, 1998, pp. 2-10, DOI 10.1.1.39.6339

[16] FUKUNAGA K., *Introduction to Statistical Pattern Recognition.* 2nd Edition, Elsevier Inc, 1990, DOI 10.1016/C2009-0-27872-X

[17] GÜNDOĞDU B., *Keyword Search for Low Resource Languages.* PhD Thesis, Bogazici Universit, 2017.

[18] GÜNDOĞDU B., SARAÇLAR M., *Distance metric learning for posteriorgram based keyword search.* 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, 2017, pp. 5660-5664, DOI 10.1109/ICASSP.2017.7953240

[19] GUPTA K., GUPTA D., *An analysis on LPC, RASTA and MFCC techniques in Automatic Speech Recognition.* 2016 6th International Conference – Cloud System and Big Data Engineering System (Confluence), Noida, 2016, pp. 493-497, DOI 10.1109/CONFLUENCE.2016.7508170

[20] GUPTA P., PUROHIT G.N., RATHORE M., *Number Plate Extraction using Template Matching Technique.* International Journal of Computer Applications, Vol. 88, No. 3, 2014, pp. 40-44, DOI 10.5120/15336-3670

[21] HAASDONK B., BAHLMANN C., *Learning with distance substitution kernels.* In Pattern Recognition – Proc. of the 26th DAGM Symposium, 2004, pp. 220-227, DOI 10.1007/978-3-540-28649-3_27

[22] HAFEN R.P., HENRY M.J., *Speech information retrieval: a review.* Multimedia Systems, Vol. 18, No. 6, 2012, pp. 499-518.

[23] HELLINGER E., (in German) *Neue Begründung der Theorie quadratischer Formen von unendlichvielen Veränderlichen.* Journal für die reine und angewandte Mathematik, Vol. 136, 1909, pp. 210–271, DOI 10.1515/crll.1909.136.210

[24] HENRIKSON J., *Completeness and total boundedness of the Hausdorff metric.* MIT Undergraduate Journal of Mathematics, 1999, pp. 69-80.

[25] HIGGINS A., WOHLFORD R., *Keyword recognition using template concatenation.* IEEE International Conference on Acoustics, Speech and Signal Processing,

ICASSP 1985, Tampa, FL, USA, 1985, pp. 1233-1236, DOI 10.1109/ICASSP.1985.1168253

[26] HOBSON A., CHENG B-K., *A comparison of the Shannon and Kullback information measures.* Journal of Statistical Physics, Vol. 7, No. 4, 1973, pp. 301–310, DOI: 10.1007/BF01014906

[27] HOLYOAK K.J., THAGARD P., *Mental Leaps: Analogy in Creative Thought.* A Bradford Book series, MIT Press, 1996.

[28] JANSEN A., DURME VAN B., *Efficient Spoken Term Discovery Using Randomized Algorithms*. 2011 IEEE Workshop on Automatic Speech Recognition & Understanding, Waikoloa, HI, 2011, pp. 401-406, DOI 10.1109/ASRU.2011.6163965

[29] JANSEN B., RIEH S.Y., *The Seventeen Theoretical Constructs of Information Searching and Information Retrieval.* In Journal of the American Society for Information Science and Technology, Vol. 61, No. 8, 2010, pp. 1517-1534, DOI 10.1002/asi.21358

[30] JENSEN J.H., ELLIS D.P.W., CHRISTENSEN M.G., JENSEN S.H., *Evaluation of Distance Measures Between Gaussian Mixture Models of MFCCs.* Proceedings of the 8th International Conference on Music Information Retrieval, ISMIR 2007, Vienna, 2007, pp. 107-108.

[31] KAILATH T., *The Divergence and Bhattacharyya Distance Measures in Signal Selection.* IEEE Transactions on Communication Technology, 1967, Vol. 15, No. 1, pp. 52-60, DOI 10.1109/TCOM.1967.1089532

[32] KAMIŃSKA D., SAPIŃSKI T., ANBARJAFARI G., *Efficiency of chosen speech descriptors in relation to emotion recognition.* EURASIP Journal on Audio, Speech, and Music Processing (2017), Vol. 3, pp. 1-9, DOI 10.1186/s13636-017-0100-x

[33] KASSAMBARA A., *Practical Guide to Cluster Analysis in R: Unsupervised Machine Learning (Multivariate Analysis)*, Vol. 1, CreateSpace Independent Publishing Platform, 2017.

[34] KESHET J., GRANGIER D., BENGIO S.A., *Discriminative keyword spotting.* Speech Communication, 2009, Vol. 51, No. 4, pp. 317-329, DOI 10.1016/j.specom.2008.10.002

[35] KORŽINEK D., MARASEK K., BROCKI Ł., WOŁK K., *Polish Read Speech Corpus for Speech Tools and Services.* Selected papers from the CLARIN Annual Conference 2016, Aix-en-Provence, 26-28 October 2016, CLARIN Common Language Resources and Technology Infrastructure, No. 136, Linköping University Electronic Press, Linköpings universitet, 2017, pp. 54-62.

[36] KULLBACK S., LEIBLER R.A., *On information and sufficiency.* Annals of Mathematical Statistics, Vol. 22, No. 1, 195, pp. 79-86, DOI 10.1214/aoms/1177729694

[37] KULLBACK S., *Information theory and statistics.* Dover Books on Mathematics, New Edition, 1997.

[38] KWIATKOWSKI W., (in Polish) *Klasyfikacja metodą grupowania cech z uwzględnieniem ich wzajemnej korelacji.* Biuletyn Instytutu Automatyki i Robotyki, Nr 14, 2000, pp. 139-146.

[39] KWIATKOWSKI W., (in Polish) *Metody automatycznego rozpoznawania wzorców.* Instytut Automatyki i Robotyki, WAT, Wydanie I, Warszawa, 2001.

[40] KWIATKOWSKI W., (in Polish) *Wykrywanie anomalii bazujące na wskazanych przykładach*. Przegląd Teleinformatyczny, Nr 1-2, 2018, pp. 3-21.

[41] KWIATKOWSKI W., (in Polish) *Wstęp do cyfrowego przetwarzania sygnałów. BEL Studio*, WAT, Warszawa, 2003.

[42] LASZKO Ł., *Word detection in recorded speech using textual queries*. Proceedings of the 2015 Federated Conference on Computer Science and Information Systems, M. Ganzha, L. Maciaszek, M. Paprzycki (eds). ACSIS, Vol. 5, pp. 849-853, DOI 10.15439/2015F341

[43] LASZKO Ł., *Using formant frequencies to word detection in recorded speech*. Proceedings of the 2016 Federated Conference on Computer Science and Information Systems, M. Ganzha, L. Maciaszek, M. Paprzycki (eds). ACSIS, Vol. 8, pp. 797-801, DOI 10.15439/2016F518

[44] LASZKO Ł., *Developing keyword spotting method for the Polish language*. Communication Papers of the 2018 Federated Conference on Computer Science and Information Systems, M. Ganzha, L. Maciaszek, M. Paprzycki (eds). ACSIS, Vol. 17, pp. 123-127, DOI 10.15439/2018F178

[45] LEBRET R., COLLOBERT R., *Word Embeddings through Hellinger PCA*. 14th Conference of the European Chapter of the Association for Computational Linguistics, EACL, 2014, pp. 482-490, DOI 10.3115/v1/E14-1051

[46] LI H., HAN J., ZHENG T., ZHENG G., *Mandarin keyword spotting using syllable based confidence features and SVM*. 2nd International Conference on Intelligent Control and Information Processing, Harbin, 2011, pp. 256-259, DOI 10.1109/ICICIP.2011.6008243

[47] LI W., BILLARD A., BOURLARD H., *Keyword Detection for Spontaneous Speech*. 2nd International Congress on Image and Signal Processing, Tianjin, 2009, pp. 1-5, DOI 10.1109/CISP.2009.5303824

[48] LIU D., CHO S., SUN D., QIU Z., *A Spearman correlation coefficient ranking for matching-score fusion on speaker recognition.* TENCON 2010 - 2010 IEEE Region 10 Conference, Fukuoka, 2010, pp. 736-741, DOI 10.1109/TENCON.2010.5686608

[49] MATTHEWS B.W., *Comparison of the predicted and observed secondary structure of T4 phage lysozyme*. Biochimica et Biophysica Acta (BBA) – Protein Structure, Vol. 405, No. 2, 1975, pp. 442-451, DOI 10.1016/0005-2795(75)90109-9

[50] MANNING CH.D., RAGHAVAN P., SCHÜTZE H., *Introduction to Information Retrieval*, Cambridge University Press, 2008.

[51] MIETTINEN K., *Introduction to Multiobjective Optimization: Noninteractive Approaches*. In: Branke J., Deb K., Miettinen K., Słowiński R. (eds) Multiobjective Optimization. Lecture Notes in Computer Science, Vol. 5252, 2008, Springer, Berlin, Heidelberg, pp. 1-26, DOI 10.1007/978-3-540-88908-3_1

[52] MIETTINEN K., RUIZ F., WIERZBICKI A.P., *Introduction to Multiobjective Optimization: Interactive Approaches*. In: Branke J., Deb K., Miettinen K., Słowiński R. (eds) Multiobjective Optimization. Lecture Notes in Computer Science, Vol. 5252, 2008, Springer, Berlin, Heidelberg, pp. 27-57, DOI 10.1007/978-3-540-88908-3_2

[53] MITRA V., HAUT VAN J., FRANCO H., VERGYRI D., *Feature Fusion for High-Accuracy Keyword Spotting.* Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference Lei Y., et al. on, 2014, pp. 7143-7147.

[54] MOHAMED S.S., ABDALLA A., JOHN R.I., *New Entropy-Based Similarity Measure between Interval-Valued Intuitionstic Fuzzy Sets.* Axioms, Vol. 8, No. 2, 2019, Article-Number 73, DOI 10.3390/axioms8020073

[55] MUSCARIELLO A., GRAVIER G., BIMBOT F., *Audio keyword extraction by unsupervised word discovery.* In Proceedings of the Interspeech, 2009, pp. 2843-2847.

[56] MÜLLER M., *Information Retrieval for Music and Motion*, Springer, 2007.

[57] NIELSEN F., *A generalization of the Jensen divergence: The chord gap divergence.* arXiv preprint, 2017, pp. 1-13, https://arxiv.org/abs/1709.10498

[58] PARDO L., *Statistical Inference Based on Divergence Measures.* Statistics: A Series of Textbooks and Monographs, 1st Edition, Chapman and Hall/CRC, 2006.

[59] PARK A.S., GLAS J.R., *Unsupervised pattern discovery in speech.* IEEE Trans. on Audio, Speech and Language Processing, 2008, Vol. 16, No. 1, pp. 186-197.

[60] PONTIUS R.G., KANGPING S., *The total operating characteristic to measure diagnostic ability for multiple thresholds.* International Journal of Geographical Information Science, Vol. 28, No. 3, 2014, pp. 570-583, DOI 10.1080/13658816.2013.862623

[61] POWERS D.M.W., *Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation.* Journal of Machine Learning Technologies, Vol. 2, No. 1, 2007, pp. 37-63.

[62] QIAO Y., MINEMATSU N., *A Study on Invariance of f-Divergence and Its Application to Speech Recognition.* IEEE Transactions on Signal Processing, 2010, Vol. 58, No. 7, pp. 3884-3890, DOI 10.1109/TSP.2010.2047340

[63] RAIELI R., *Introducing Multimedia Information Retrieval to libraries.* Italian Journal of Library, Archives, and Information Science, Vol. 7, No. 3, 2016, pp. 9-42, DOI 10.4403/jlis.it-11530

[64] SAMMUT C., WEBB G.I. (eds.), *Encyclopedia of Machine Learning and Data Mining.* 2nd Edition, Springer, 2017.

[65] SASAKI Y., *The truth of the F-measure.* 2007, 5 pages, Web resource available at https://www.toyota-ti.ac.jp/Lab/Denshi/COIN/people/yutaka.sasaki/F-measure-YS-26Oct07.pdf

[66] SCHÖLKOPF B., *The Kernel Trick for Distances.* Advances in neural information processing systems, Vol. 13, 2000, pp. 301-307.

[67] SINGH A., YADAV A., RANA A., *K-means with Three different Distance Metrics.* International Journal of Computer Applications, Vol. 67, No. 10, 2013, pp. 13-17, DOI 10.5120/11430-6785

[68] SINGHAL A., *Modern Information Retrieval: A Brief Overview.* Bulletin of the IEEE Computer Society Technical Committee on Data Engineering, Vol. 24, No. 4, 2001, pp. 35-43.

[69] STEHMAN S.V., *Selecting and interpreting measures of thematic classification accuracy*. Remote Sensing of Environment, Vol. 62, No. 1, 1997, pp. 77-89, DOI 10.1016/S0034-4257(97)00083-7

[70] TABIBIAN S., AKBARI A., NASERSHARIF B., *Improved dynamic match phone lattice search for Persian spoken term detection system in online and offline applications*. International Journal of Speech Technology, March 2019, Vol. 22, Issue 1, pp. 205-217, DOI 10.1007/s10772-019-09594-w

[71] TÜSKEA Z., NOLDEN D., SCHLÜTERA R., NEY H., *Multilingual MRASTA features for low-resource keyword search and speech recognition systems*. 2014 IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP), 2014, pp. 7349-7353.

[72] WILPON J.G., RABINER L.R., LEE C., GOLDMAN E.R., *Automatic recognition of keywords in unconstrained speech using hidden Markov.* IEEE Transactions on Acoustics, Speech and Signal Processing, 1990, Vol. 38, No. 11, pp. 1870-1878, DOI 10.1109/29.103088

[73] YOUDEN W.J., *Index for rating diagnostic tests.* Cancer, Vol. 3, 1950, pp. 32-35, DOI 10.1002/1097-0142(1950)3:1<32::AID-CNCR2820030106>3.0.CO;2-3

[74] ZEDDELMANN VON D., KURTH F., MÜLLER M., *Perceptual audio features for unsupervised key-phrase detection.* Proc. ICASSP2010, 2010, pp. 257-260, DOI 10.1109/ICASSP.2010.5495974

[75] ZHANG Y., *Unsupervised Speech Processing with Applications to Query-by-Example Spoken Term Detection.* PhD thesis, Massachusetts Institute of Technology, 2013.

[76] ZHANG Y., GLASS J.R., *Unsupervised spoken keyword spotting via segmental DTW on Gaussian posteriorgrams.* 2009 IEEE Workshop on Automatic Speech Recognition & Understanding, Merano, 2009, pp. 398-403, DOI 10.1109/ASRU.2009.5372931

[77] ZHU X., PENN G., RUDZICZ F., *Summarizing multiple spoken documents: finding evidence from untranscribed audio.* Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP, Vol. 2, 2009, pp. 549-557.

[78] ZIELIŃSKI T.P., (in Polish) *Cyfrowe przetwarzanie sygnałów od teorii do zastosowań.* Wydawnictwa Komunikacji i Łączności, Warszawa, 2005.

[79] ZIÓŁKO B., GAŁKA J., SKURZOK D., JADCZYK T., *Modified Weighted Levenshtein Distance in Automatic Speech Recognition.* Krajowa Konferencja Zastosowań Matematyki w Biologii i Medycynie, Krynica, 2010, s. 116-120.

Eksperymentalne badanie wpływu wyboru funkcji podobieństwa na jakość wykrywania słów w sygnale mowy

STRESZCZENIE: W pracy przedstawiono ocenę zastosowania wybranych funkcji podobieństwa w zadaniu wykrywania słów kluczowych. Przeprowadzono eksperymenty dla języka polskiego. Wyniki badań można wykorzystać do ulepszenia już istniejących metod wykrywania słów kluczowych lub do opracowania nowych.

SŁOWA KLUCZOWE: wykrywanie słów kluczowych, podobieństwo sygnałów, wskaźniki jakości wykrycia, odkształcanie skali czasu, kwerenda tekstowa

Received by the editorial staff on: 27.11.2019

PRZEGLĄD TELEINFORMATYCZNY
NR 3-4, 2019

Eksperymentalne badanie wpływu wyboru funkcji podobieństwa na jakość wykrywania słów w sygnale mowy

Łukasz LASZKO

Instytut Teleinformatyki i Cyberbezpieczeństwa, Wydział Cybernetyki, WAT
ul. gen. Sylwestra Kaliskiego 2, 00-908 Warszawa,
lukasz.laszko@wat.edu.pl

STRESZCZENIE: W pracy przedstawiono ocenę zastosowania wybranych funkcji podobieństwa w zadaniu wykrywania słów kluczowych. Przeprowadzono eksperymenty dla języka polskiego. Wyniki badań można wykorzystać do ulepszenia już istniejących metod wykrywania słów kluczowych lub do opracowania nowych.

SŁOWA KLUCZOWE: wykrywanie słów kluczowych, podobieństwo sygnałów, wskaźniki jakości wykrycia, odkształcanie skali czasu, kwerenda tekstowa

1. Wprowadzenie

Zadanie wykrywania słów w sygnale mowy (ang. *keyword spotting*, KWS) polega na wykryciu zadanych słów[1] (ang. *query-by-example*) w zarejestrowanym sygnale mowy spontanicznej. Cel tego zadania jest realizowany przez wskazanie miejsc w sygnale mowy, w których zadane słowo występuje. Zwykle wskazania te powinny minimalizować prawdopodobieństwo fałszywego spokoju oraz fałszywego ataku [22].

[1] Spotykane są też sformułowania: *wykrywanie słów kluczowych* (ang. *keyword* lub *key-word*), *wykrywanie fraz* (ang. *key-phrase*) [74] lub *detekcja słów ze słownika* (ang. *spoken term detection*) [59].

Zadanie KWS należy do dziedziny określanej jako *wyszukiwanie informacji* (ang. *information retrieval*) [50][2]. W tej dziedzinie jest ono określone w następujący sposób:

a) sygnał mowy jest z założenia generowany przez różnych mówców,
b) poszukiwane słowo jest zadane w postaci tekstowej,
c) sygnał wzorcowy pozyskiwany jest metodą przekształcenia tekstu na sygnał mowy (ang. *text-to-speech*), przez wykorzystanie nagrań mówców naturalnych bądź wykorzystanie syntezatorów mowy,
d) wyszukiwanie wzorca w sygnale mowy jest realizowane na podstawie porównywania badanego sygnału z sygnałem wzorcowym,
e) porównanie odbywa się w przestrzeni sygnałów a nie tekstu (ciągu symboli fonetycznych).

Jednym z najistotniejszych problemów do rozwiązania jest wyznaczenie podobieństwa pomiędzy modelami dwóch sygnałów: badanej wypowiedzi (ang. *utterance*) i wzorcowego (tzw. kwerendy) [17]. Analiza publikacji z okresu ostatnich dwudziestu lat pozwoliła autorowi zaobserwować, że zwykle podobieństwo to jest wyznaczane w metrycznej przestrzeni cech sygnału mowy R^N. Stosowanymi cechami są współczynniki akustyczne, takie jak mel-frequency cepstral coefficients (MFCC). Ocena podobieństwa pomiędzy modelami sygnałów dokonywana jest na podstawie odległości między nimi w R^N, przy czym odległość mniejsza oznacza większe podobieństwo. Najczęściej stosowana metryka w zadaniach KWS to metryka kosinusowa [28], [77], [68].

Wybór metryki jest najczęściej arbitralny i w publikacjach nie dyskutowany przez badaczy. Jak zauważono w pracy [17], może to wynikać z własności samej metryki. Ale istotne różnice interpretacyjne występują np. dla metryk euklidesowej i kosinusowej. Wpłynęło to na ukierunkowanie celu badań opisywanych w tym artykule.

Celem badań autora było określenie wpływu wyboru funkcji podobieństwa na jakość wykrywania słów w sygnale mowy. W niniejszym artykule opisano uzyskane przez autora wyniki badań porównawczych dla przypadku zastosowania metody wykrywania słów wprowadzonej w pracy [42]. Badania wykonano dla języka polskiego analogicznie do badań raportowanych w pracy [44], wykorzystując między innymi ten sam korpus mowy polskiej [35].

[2] Konkretnie w dziedzinie dźwięku spotkać można umiejscowienie KWS w ramach *Audio IR* [15], *Multimedia IR* [63], [56]. Jeszcze inne spojrzenie przedstawia [29].

2. Podobieństwo słów w sygnale mowy

2.1. Metody oceny podobieństwa

Można wyróżnić następujące podejścia do określania podobieństwa dwu sygnałów mowy [64], [27][3]:

Podobieństwo kategorialne (ontologiczne) – polega na dokonaniu oceny na podstawie klasyfikacji, bazującej na znanych kategoriach pojęciowych (np. głoska dźwięczna).

Podobieństwo atrybutów – polega na posiadaniu przez analizowane słowa identycznych lub podobnych cech (właściwości), a wartości liczbowe cech wykazują niewielkie różnice (są zbliżone), np. częstotliwości formantowe.

Podobieństwo relacji – polega na tym, że pomiędzy analizowanymi słowami zachodzą identyczne lub podobne relacje, np. proporcji.

Podobieństwo związków przyczynowych (semantyczne) – polega na tym, że analizowane słowa mają takie same (zbliżone) konteksty, np. dane słowa określają ten sam podmiot w zdaniu.

W przypadku zadań wykrywania słów w sygnale mowy najczęściej wyznacza się podobieństwo ze względu na atrybuty sygnału mowy (czyli podobieństwo atrybutów). Tymi atrybutami (cechami sygnału mowy) są najczęściej współczynniki akustyczne, takie jak: MFCC [55], human-factor cepstral coefficients (HFCC) [74], relative spectral-perceptual linear prediction (RASTA-PLP) [71], [32], [19] oraz inne, o których pisze się np. w pracy [53]. Problem wyboru funkcji podobieństwa może zależeć od przyjętych cech reprezentujących porównywane sygnały.

2.2. Ocena podobieństwa

Do rozwiązania zadania KWS można podejść w dwojaki sposób: wykorzystując metody rozpoznawania mowy [72] lub metody przetwarzania mowy [59].

[33] Opracowanie własne na podstawie [64] pp. 190-193, [27] pp. 22-37. Inną klasyfikację podejść przedstawiono np. w pracy [1].

W wyniku zastosowania metod rozpoznawania mowy właściwe wykrywanie słów odbywa się w przestrzeni tekstu (ciągu symboli fonetycznych), który uzyskano, analizując słowa z nagrania. Wyznaczenie podobieństwa słów sprowadza się wtedy do obliczenia odległości pomiędzy ciągami symboli, bazując np. na odległości Levenshteina, jak w pracy [79]. W takim przypadku dokonuje się wskazania słowa, którego odległość Levenshteina od kwerendy tekstowej jest najmniejsza.

W miejsce odległości Levenshteina stosuje się również inne, takie jak:

- Damerau–Levenshteina [4],
- Jaro–Winklera [70],
- Hamminga [75] oraz
- LCS (ang. *longest common subsequence*) [42].

W przypadku zastosowania metod przetwarzania mowy wykrywanie słów odbywa się w przestrzeni sygnału. Sygnał mowy dla podanej kwerendy tekstowej uzyskuje się drogą syntezy *text-to-speech*. Uzyskany wektor próbek sygnału przekształca się na wektor cech. Dalej, zależnie od przyjętego modelu sygnału, wyróżnia się następujące podejścia do oceny podobieństwa słów:

1) Jeśli reprezentacja sygnału jest pojedynczym wektorem (np. MFCC), oceny podobieństwa dokonuje się na podstawie:
   a) odległości między wektorami, typowo jest to odległość kosinusowa, choć stosowane są również inne odległości, takie jak:
      - euklidesowa [34], [25],
      - kosinusowo-euklidesowa [22],
      - logarytmiczno-kosinusowa [18],
      - Manhattan [20],
      - sigma [18],
   b) współczynnika korelacji (przy czym wartość zerowa oznacza brak podobieństwa), typowo jest to korelacja Pearsona choć stosowane są również: korelacja Kendalla lub Spearmana[4] [33], [48], [39].

2) Jeśli model sygnału jest grupą (klastrem) wektorów (np. zbiorem cech grupy ramek), to wnioskowanie o podobieństwie dwu sygnałów wymaga zdefiniowania podobieństwa między klastrami. Oceny

[4] Zwana również korelacją rangową. Rangi są numerami kolejnych obserwacji w uporządkowanej próbie statystycznej.

podobieństwa dokonuje się na podstawie odległości między klastrami, przy czym tak rozumiana *odległość* zazwyczaj nie spełnia aksjomatów metryki[5]. W tym przypadku można wyróżnić następujące podejścia:

a) określenie odległości na podstawie elementów klastra (np. pomiędzy elementami centralnymi klastrów), do czego zwykle jest stosowana odległość euklidesowa lub inne odległości bazujące na odległości Minkowskiego [67],
b) określenie na podstawie odległości rozkładu elementów w klastrze, w tym modelu probabilistycznego, do czego często stosowaną odległością jest odległość Kullbacka–Leiblera [26], [30], choć stosowane są również inne, takie jak:
    - Bhattacharyyi [1], [16], [5], [3], [31],
    - Mahalanobisa [3], [38],
    - Hellingera [45], [23], [31], [58] oraz
    - dywergencje: *f-divergence*, Jensena itp. [57], [62].

W tym artykule opisuje się badania, które obejmują swoim zakresem podejście drugie, to jest do rozwiązania zadania KWS wykorzystywane są metody przetwarzania mowy (por. [42]), a zadaniem badawczym jest wybór funkcji podobieństwa.

2.3. Ocena funkcji podobieństwa

W tabeli 1 znajduje się wykaz funkcji podobieństwa wykorzystanych w opisywanych badaniach. Funkcja podobieństwa należy do istotnych składników metod stosowanych w zadaniach KWS i ma bezpośredni wpływ na jakość wykrywania słów. Celowe jest więc używanie tej funkcji podobieństwa, która zastosowana w konkretnej metodzie spowoduje uzyskanie wyników o najwyższej jakości.

2.3.1. Wskaźniki jakości wykrywania słów w zadaniach KWS

Jakość wykrywania można mierzyć za pomocą podstawowych wskaźników, związanych bezpośrednio z liczebnością osiągniętych wyników [61]. Do nich zalicza się:

[5] Por. np. [78] s. 39.

Tab. 1. Wykaz badanych funkcji podobieństwa

Lp.	Podstawa definicji funkcji podobieństwa; odległość[6]:
1	Bhattacharyyi (K_{bha})
2	Chebyszewa (K_{che})
3	korelacyjna (K_{cor})
4	kosinusowa (K_{cos})
5	euklidesowa (K_{euc})
6	Hellingera (K_{hel})
7	symetryczna Kulbacka–Leiblera (K_{skl})
8	Manhattan (K_{man})
9	Mahalanobisa (K_{mah})
10	Minkowskiego (K_{min})
11	euklidesowa standaryzowana (K_{seu})
12	Spearmana (K_{spr})

- TP (ang. *True Positive*) – liczba prawidłowych wskazań (trafień),
- TN (ang. *True Negative*) – liczba prawidłowych odrzuceń,
- FP (ang. *False Positive*) – liczba nieprawidłowych wskazań (błędów I rodzaju, „fałszywych alarmów"),
- FN (ang. *False Negative*) – liczba nieprawidłowych odrzuceń (błędów II rodzaju, chybień, „fałszywych spokojów").

Wskaźniki te często zestawia się w tablicę błędów (ang. *error/confusion table/matrix*) [69][7]. W zadaniach KWS istotna jest także precyzja wskazań oraz inne wskaźniki, które dają możliwość odniesienia uzyskanych wyników (np. w celu porównania dwóch metod). Do nich należą wskaźniki pochodne. W przeprowadzonych badaniach wybrano następujące wskaźniki:

- **precyzja** (ang. *Precision*, ozn. PPV),
- **dokładność** (ang. *Accuracy*, ozn. ACC),
- **czułość** (ang. *Recall*, *True positive rate*, ozn. TPR),

[6] W nawiasie umieszczono zastosowane oznaczenie funkcji podobieństwa.

[7] Za: https://en.wikipedia.org/wiki/Confusion_matrix (visited: 19.08.2019).

- **swoistość** (ang. *Specificity*, *True negative rate*, ozn. TNR),
- **wskaźnik F_1** (*F-measure*, *F_1Score*, ozn. F_1S) [9], [65] oraz
- **indeks Youdena** (*Youden's J statistic*, ozn. YJS) [73].

Na podstawie wartości wskaźnika PPV można ocenić, czy dana metoda (przy wykorzystaniu danej funkcji podobieństwa) daje powtarzalne wyniki, charakteryzujące się małym rozrzutem. Po wartości wskaźnika ACC można ocenić, czy dana metoda daje zawsze wyniki zbliżone do prawdziwych (rzeczywistych). Wskaźnik TPR określa zdolność metody do prawidłowego wykrycia (wskazania wyniku), tam, gdzie faktycznie poszukiwana wartość występuje. Wskaźnik TNR natomiast określa zdolność metody do prawidłowego odrzucania wyników (czyli tzw. selektywności). Za pomocą wskaźnika F_1S ocenia się *wiarygodność* metody, czyli cechę świadczącą o autentyczności uzyskiwanych nią wyników (zarówno wskazań, jak i odrzuceń). Wskaźnik YJS jest natomiast używany do oceny efektywności metody[8] oraz do wyboru najlepszych parametrów metody w analizie ROC (por. rozdz. 5.1).

2.3.2. Skalaryzacja oceny wektorowej

W pracy przyjęto, że ocena wektorowa funkcji podobieństwa dokonana zostanie za pomocą sześciu wskaźników pochodnych, wymienionych wyżej. Warto zwrócić uwagę, że opisane wyżej wskaźniki mają ten sam zakres wartości. Jest nim przedział liczbowy $[0,1]$, przy czym wartość wskaźnika równa *jeden* charakteryzuje dobrą metodę (np. najbardziej precyzyjną, najbardziej dokładną itp.).

W celu uszeregowania ocen wektorowych i zarazem wyboru najlepszej funkcji dokonano oceny skalarnej, poprzez sumę najlepszych wyników każdego wskaźnika jakości. Powyższe założenia wynikają z obserwacji autora, że wyniki te ściśle zależą od warunków eksperymentu. W szczególności, w warunkach dużej zmienności badanego materiału nie ma wystarczającego uzasadnienia do statystycznej oceny jakości, np. liczba slotów istotnie zależy od wykrywanego słowa. Przyjęto więc metodę „konkursu". Polega ona na ocenie badanej funkcji za pomocą najlepszego uzyskanego wyniku (w całej serii badawczej).

[8] Efektywność metody, którą pokazuje wskaźnik YJS, dotyczy wynikowej jej czułości w przypadku, gdy w zbiorze uzyskanych daną metodą wyników znajdują się wyniki fałszywe.

3. Eksperyment badawczy

Przeprowadzone badania polegały na wykorzystaniu metody przedstawionej w pracy [42]. Metoda ta ukierunkowana jest na wykorzystanie wzorców pochodzących z syntezatora TTS i takie wzorce były głównym obiektem zainteresowania. Badania prowadzono dla języka polskiego dla korpusu mowy CLARIN-PL Mobile Corpus (EMU) [35], w zakresie i zgodnie z procedurą opisaną w pracy [44]. W tab. 2 przedstawiono wartości parametrów metody niezmienione w stosunku do [44] oraz wartości zmienione, przyjęte dla funkcji podobieństwa niebadanych w pracy [44].

Tab. 2. Parametry metody KWS użytej w opisanych badaniach

	Nazwa parametru	**Wartości parametrów**											
Wartości niezmienione	Liczba współczynników FFT	8192											
	Rozmiar okna analizy	1024											
	Procent nakładania	33%											
	Liczba współczynników HFCC	15											
	Zakres częstotliwości sygnału	[300, 3400]											
	Współczynnik długości kwerendy	1,5											
	Współczynnik dopasowania kwerendy	0,5											
	Wartość progowa ścieżki	0,6											
Wartości zmienione	Sposób pomiaru podobieństwa[9]	*bha*	*che*	*cor*	*cos*	*euc*	*hel*	*skl*	*man*	*mah*	*min*	*seu*	*spr*
	Sposób normalizacji[10]	-	HE	HE	HE	HE	-	HE	HE	HE	HE	HE	HE
	Wartość progowa sekwencji (real/TTS)[11]	89/78	80/70	77/76	65/54	73/65	82/85	85/60	75/55	97/97	78/68	68/67	75/72
	Pozostałe[12]	NAN=1	NAN=1	NAN=0	NAN=0	NAN=1	NAN=1	NAN=0	NAN=1	ABS, NAN=1	NAN=1	NAN=1	NAN=1

W celach porównawczych wykonano dodatkowo badania, wykorzystując wzorce wyodrębnione z nagrań rzeczywistej mowy. Oznaczono je w wynikach jako *real*.

9 Oznaczenia jak w tab. 01.

10 HE – normalizacja metodą wyrównywania histogramu (ang. *histogram equalization*).

11 Wartość ta w pracach: [42], [43] i [44] jest określana jako próg jakości rozpoznania. Jest ona stosowana po oznaczeniu wykrywanych sekwencji jako podejrzane, czyli po zastosowaniu *wartości progowej ścieżki*, co dobitnie przedstawiono w pracy [42].

12 NAN – sposób interpretacji wartości nieliczbowych, ABS – wartość bezwzględna.

4. Wyniki

4.1. Podstawowe wskaźniki jakości

Przedstawiono wyniki 120 badań w postaci wykresów i tabel. Zasadnicze wyniki to uzyskane bezpośrednio z eksperymentu wskaźniki liczebności: TP, TN, FP, FN. Na ich podstawie wyznaczono opisywane wyżej wskaźniki pochodne.

W tabeli 3 przedstawiono przykładowe wyniki badań, gdy zastosowana funkcja podobieństwa bazowała na odległości Bhattacharyyi. Wartości zawarte w tabeli, w kolejnych wierszach przedstawiają wyniki dla kwerendy wyodrębnionej z nagrania rzeczywistej mowy (*real*) oraz kwerendy tekstowej zsyntezowanej (*TTS*). Liczba slotów analizy, ozn. jako *Seg*, jest liczbą wszystkich jednostek, które metoda wyodrębniła w analizowanym sygnale mowy. Liczba ta jest zależna od długości kwerendy, stąd wynika jej różnica w badaniu dla tej samej sesji. Slot ten nie jest oknem analizy, ale długością poszukiwanego wzorca (por. tabela 2).

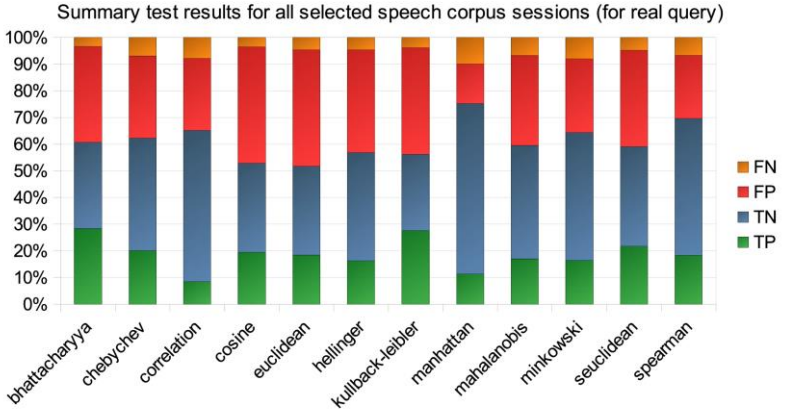
Tab. 3. Wyniki badań dla wybranych dziesięciu sesji nagrań z korpusu mowy. Funkcja podobieństwa bazuje na odległości Bhattacharyyi

		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Real	**Slots**	80	56	88	56	53	40	50	55	48	77
	TP	22	10	25	12	16	12	10	26	12	26
	FP	17	14	32	29	3	13	30	13	29	37
	TN	36	28	28	15	29	14	10	14	7	14
	FN	5	4	3	0	5	1	0	2	0	0
TTS	**Slots**	43	26	39	26	38	26	36	36	29	53
	TP	21	12	22	10	21	14	10	24	10	27
	FP	6	4	7	10	6	7	14	9	18	23
	TN	13	9	7	5	9	3	2	3	1	3
	FN	3	1	3	1	2	2	0	0	0	0

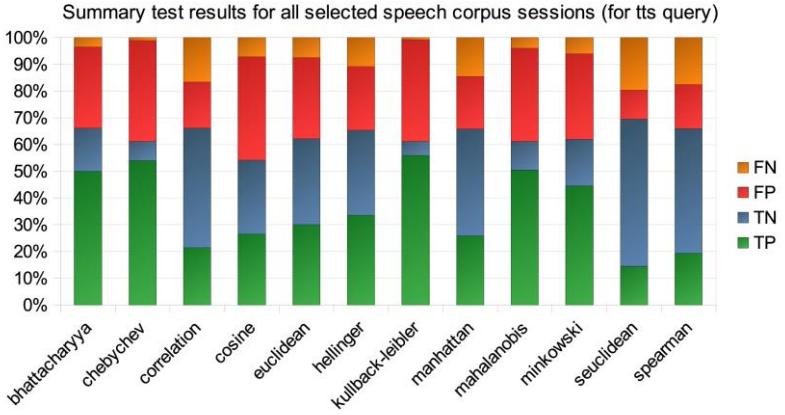
Pozostałe wyniki badań (dla pozostałych funkcji podobieństwa) przedstawiono w sposób skumulowany na rysunkach 1 i 2.

Oba wykresy przedstawiają skumulowane wartości dla wszystkich wybranych sesji wykorzystanego w badaniach korpusu mowy. Wykresy dają

możliwość porównania wyników dla różnych funkcji podobieństwa. Pokazują również, że mimo braku stosownej kalibracji metody, w każdym przypadku wyniki metody były użyteczne, to znaczy wyniki prawdziwe (TP i TN) w sumie zawsze znajdowały się w większości (czyli miały powyżej 50% wszystkich wyników). Niepożądane wyniki fałszywe (FP i FN) są po części wynikiem wspomnianego braku kalibracji, choć pokazują też niedoskonałość metody, która polega na zależności od samych danych (czyli nagrań), o czym wspomniano w pracy [42]. Więcej informacji o wynikach dają przedstawione w punkcie następnym wartości wskaźników pochodnych.



Rys. 1. Zobrazowanie wyników dla kwerendy real w ujęciu procentowym



Rys. 2. Zobrazowanie wyników dla kwerendy TTS w ujęciu procentowym

4.2. Uzyskane wskaźniki jakości

W tabeli 4 przedstawiono przykład wartości wskaźników dla wyników uzyskanych w badaniach dla funkcji podobieństwa bazującej na odległości Hellingera. W tabeli zaznaczono wiersz dla wskaźnika czułości (TPR). Wskaźnik ten pokazuje zdolność danej metody do wykrycia (wskazania wyniku) tam, gdzie faktycznie poszukiwana wartość występuje. Wartości bliskie jedynce świadczą o wysokiej czułości użytego klasyfikatora. W przedstawionym przypadku występowały sesje, dla których znaleziono praktycznie wszystkie poszukiwane słowa przy jednoczesnym małym odsetku fałszywych odrzuceń (TN).

Uśrednione wartości: $\overline{TPR_{real}} = 0{,}74$, $\overline{TPR_{TTS}} = 0{,}75$, czyli dla tzw. *średniego przypadku* pokazują, że tę funkcję podobieństwa można z powodzeniem stosować w sytuacji, kiedy badaczowi zależy przede wszystkim na maksymalizacji liczby wykryć (wskazań prawdziwych, TP), nie dbając zupełnie o wartości fałszywie pozytywne (FP).

Tab. 4. Wskaźniki jakości, dla metody wykorzystującej funkcję podobieństwa, bazującej na odległości Hellingera. Przedstawiono wyniki 10 sesji badawczych.

		1	2	3	4	5	6	7	8	9	10
Real	**PPV**	0,55	0,10	0,43	0,14	0,38	0,33	0,08	0,54	0,13	0,45
	ACC	0,71	0,60	0,66	0,41	0,62	0,66	0,38	0,66	0,43	0,59
	TPR	**0,89**	**0,57**	**0,80**	**0,73**	**0,69**	**0,77**	**0,57**	**0,85**	**0,75**	**0,77**
	TNR	0,62	0,60	0,61	0,36	0,60	0,64	0,36	0,55	0,39	0,49
	F1S	0,68	0,17	0,56	0,24	0,49	0,47	0,14	0,66	0,22	0,57
	YJS	0,51	0,17	0,41	0,09	0,29	0,41	-0,07	0,39	0,14	0,26
TTS	**PPV**	0,67	0,88	0,94	0,47	0,83	0,64	0,36	0,70	0,32	0,51
	ACC	0,70	0,81	0,82	0,59	0,82	0,62	0,42	0,72	0,38	0,57
	TPR	**0,56**	**0,64**	**0,74**	**0,70**	**0,67**	**0,64**	**0,89**	**0,84**	**0,89**	**0,96**
	TNR	0,80	0,93	0,94	0,53	0,91	0,58	0,18	0,59	0,15	0,24
	F1S	0,61	0,74	0,83	0,56	0,74	0,64	0,52	0,76	0,47	0,67
	YJS	0,36	0,57	0,68	0,23	0,58	0,23	0,07	0,43	0,04	0,20

Dla pozostałych funkcji obliczone wartości wskaźników przedstawiono w sposób graficzny. Pierwsze zestawienie pokazuje wskaźniki PPV i ACC (rys. 3). Wybrano po cztery funkcje podobieństwa, dla których wskaźniki te po uśrednieniu były najwyższe. Te wskaźniki dobrze jest analizować równocześnie,

gdyż tak analizowane mogą wskazać możliwy kierunek kalibracji metody wykrywania. Na podstawie tych wyników można stwierdzić, uogólniając, że wykorzystana metoda KWS jest dokładna, gdyż wskaźnik ACC uzyskał dość wysokie wartości, a przy tym wartości te cechują się małym rozrzutem (co widać na wykresach c i d). Jednocześnie metoda ta jest mało precyzyjna, to jest dla jednych analizowanych nagrań nie wykrywa tych fragmentów, które powinna wykryć (mała wartość PPV), a dla innych wykrywa (PPV bliskie jedynce) – wykresy a i b).

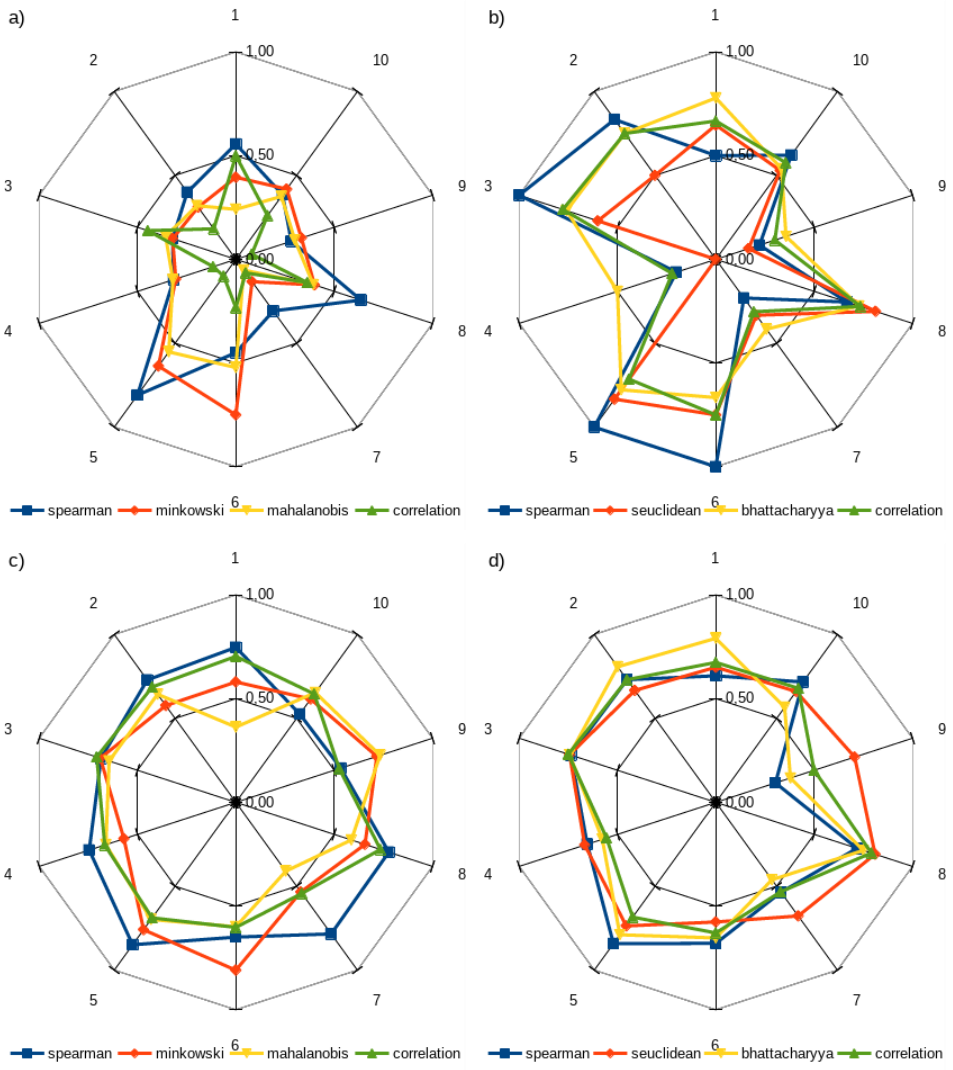
Na rysunku 3 b) widać także, że PPV wyznaczony w przypadku stosowania funkcji podobieństwa bazującej na odległości Bhattacharyyi nie ma tak dużej różnicy wartości w kolejnych badaniach (dla innych danych) niż miejscami *lepsza* funkcja bazująca na korelacji Spearmana. Świadczy to o mniejszej zależności tej pierwszej funkcji podobieństwa od konkretnych danych użytych w badaniach, a zatem i o większej odporności (ang. *robustness*) całej metody wykrywania.

Drugie zestawienie (rys. 4) pozwala wnioskować o stopniu *wiarygodności* do zastosowanej metody wykrywania. W badaniach wykorzystano kwerendę zsyntezowaną *TTS*. Metoda *wiarygodna*, w tym przypadku rozumiana jest jako taka, która nie maksymalizuje liczby fałszywych wyników, a wykrywa i odrzuca to, co powinna według stanu faktycznego.

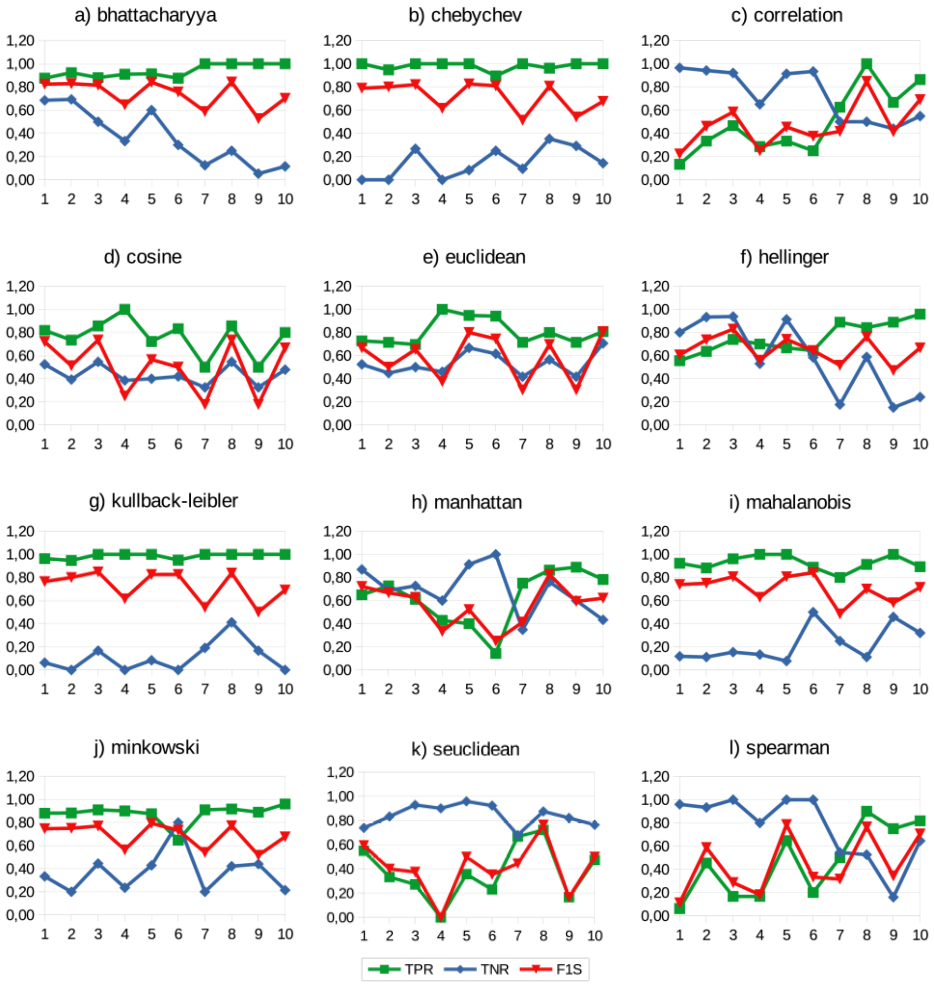
Trzecie zestawienie (rys. 5) pokazuje obliczone indeksy Youdena dla przypadków: średniego i maksymalnego. Uzyskane rezultaty przedstawiono w uporządkowany sposób względem wartości średniej. Najlepsze funkcje podobieństwa, według tego wskaźnika, to te bazujące na odległości Spearmana, Bhattacharyyi i Manhattan.

4.3. Ocena jakościowa funkcji podobieństwa

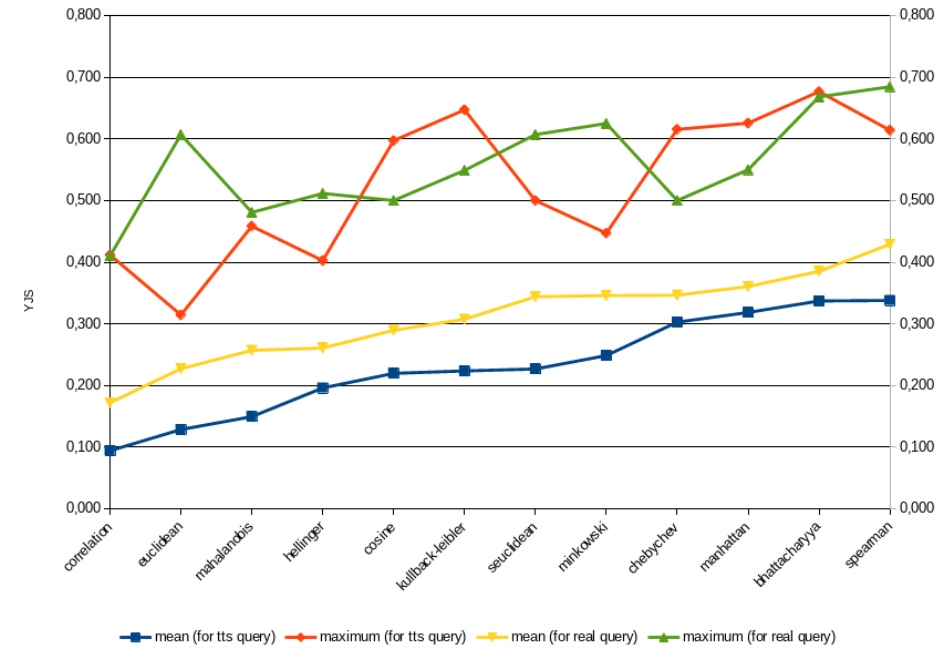
Przedstawiony poniżej w tabeli 85 ranking funkcji podobieństwa jest podsumowaniem wykonanych badań celowych opisywanych w artykule. Został on wykonany na podstawie oceny jakościowej dokonanej dla wszystkich prób badawczych, zgodnie ze sposobem opisanym w pkt. 2.3.2. Wynik końcowy przedstawiony w tabeli otrzymano drogą wcześniej opisanej skalaryzacji. Dla porównania umieszczono również wyniki badań dla kwerendy *real*.



Rys. 3. Zestawienie wartości wskaźników PPV i ACC, dla wybranych funkcji podobieństwa: a) PPV dla kwerendy *real*, b) PPV dla kwerendy *TTS*, c) ACC dla kwerendy real, d) ACC dla kwerendy *TTS*; Wyniki uzyskano w kolejnych sesjach badawczych (od 1 do 10)



Rys. 4. Zestawienie wskaźników świadczących o *wiarygodności* metody wykrywania. W badaniach wykorzystano kwerendę syntezowaną *TTS*



Rys. 5. Zestawienie indeksów Youdena (YJS). Niebieskie kwadraty i żółte trójkąty pokazują przypadki średnie. Czerwone romby i zielone trójkąty pokazują przypadki najlepsze

5. Badania dodatkowe

5.1. Analiza krzywej ROC

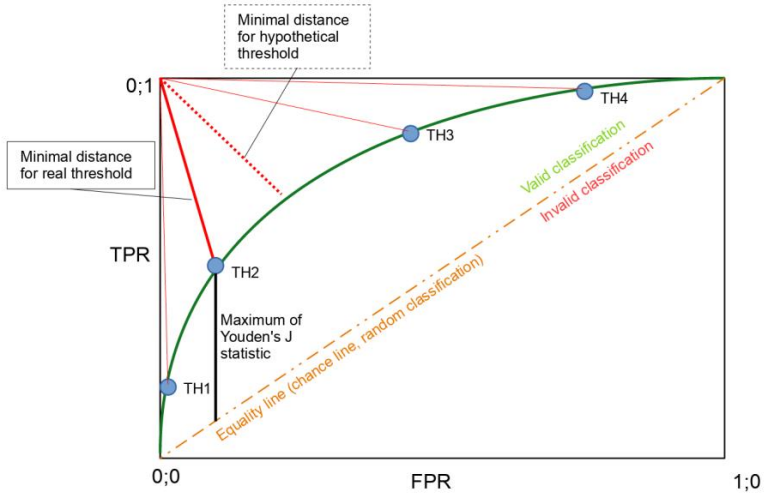
Wskaźniki liczbowe, które posłużyły do wyboru funkcji podobieństwa sygnałów, opisują tylko pewien chwilowy stan badawczy. W celu poznania, jak zachowuje się metoda wykrywania słów w szerszym zakresie, przeprowadzono analizę krzywej ROC (ang. *Receiver Operating Characteristic curve*) [14], [60]. Analizę przeprowadzono tylko dla wybranej (najlepszej) funkcji podobieństwa Spearmana. Krzywa ROC powstaje jako zestawienie wartości wskaźników TPR i FPR uzyskanych dla kilku powtórzeń badania przy różnych wartościach progowych (patrz rys. 6). Przy czym:

$$FPR = 1 - TNR \tag{1}$$

jest wskaźnikiem odrzucenia (ang. *fallout*, *False Positive Rate*).

Tab. 5. Ranking funkcji podobieństwa

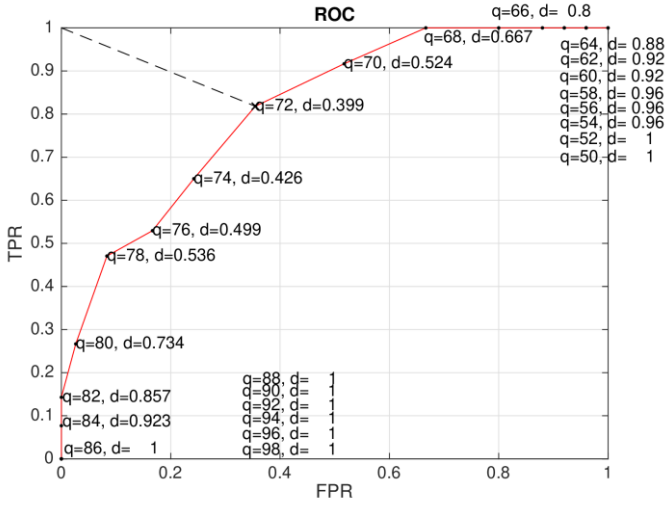
	TTS		real	
Lp.	**Funkcja podobieństwa**	**Ocena wskaźnikowa**	**Funkcja podobieństwa**	**Ocena wskaźnikowa**
1	K_{spr} (Spearmana)	5,175	K_{bha} (Bhattacharyyi)	5,066
2	K_{hel} (Hellingera)	5,167	K_{spr} (Spearmana)	5,028
3	K_{man} (Manhattan)	5,154	K_{min} (Minkowskiego)	4,869
4	K_{cor} (korelacyjna)	4,880	K_{man} (Manhattan)	4,782
5	K_{bha} (Bhattacharyyi)	4,735	K_{seu} (euklidesowa standaryzowana)	4,670
6	K_{euc} (euklidesowa)	4,726	K_{euc} (euklidesowa)	4,537
7	K_{seu} (euklidesowa standaryzowana)	4,685	K_{che}(Chebyszewa)	4,439
8	K_{min} (Minkowskiego)	4,556	K_{mah} (Mahalanobisa)	4,046
9	K_{mah} (Mahalanobisa)	4,370	K_{skl} (symetryczna Kulbacka-Leiblera)	4,031
10	K_{skl} (symetryczna Kulbacka–Leiblera)	4,176	K_{hel} (Hellingera)	3,971
11	K_{cos} (kosinusowa)	4,023	K_{cos} (kosinusowa)	3,957
12	K_{che}(Chebyszewa)	3,957	K_{cor} (korelacyjna)	3,808



Rys. 6. Schematyczne przedstawienie krzywej ROC i sposobu wyznaczania najlepszej wartości progowej. Wartości progowe TH od 1 do 4 nałożono w miejsce rzeczywistych wartości wskaźników TPR i FPR wynikających z pomiaru. TH mogą mieć dowolny zakres wartości, zależny od metody. Analogicznie oznaczono wskaźnik Youdena, który dla badania z TH2 posiadał wartość większą niż dla badań z pozostałymi TH. Na wykresie oznaczono także hipotetycznie najlepszą wartość TH, którą można wyznaczyć metodą graficzną, porównując np. dwa sąsiednie wartości progowe (w tym przypadku TH2 i TH3)

Te badania przeprowadzono dla przypadku kwerendy. Wykonano w sumie 250 badań nad metodą przedstawioną w pracy [42]. Przyjęto parametry metody takie jak w tab. 2, zmieniając jedynie wartość progową sekwencji w przedziale od 50 do 98 z krokiem 2 (czyli dla 25 wartości tego progu). Badania przeprowadzono dla wszystkich wybranych sesji nagrań z analizowanego korpusu mowy. Na podstawie uzyskanych wyników wyznaczono wartości TPR i FPR i umieszczono je na wykresach. Poniższe wykresy przedstawiają:

- rysunek 7: szczegółową analizę krzywej ROC dla wybranej sesji, wraz ze sposobem wyboru wartości progowej maksymalizującej TPR i minimalizującej FPR,
- rysunek 8: przeprowadzone analizy krzywej dla pozostałych sesji z zaznaczoną najlepszą wartością progową.

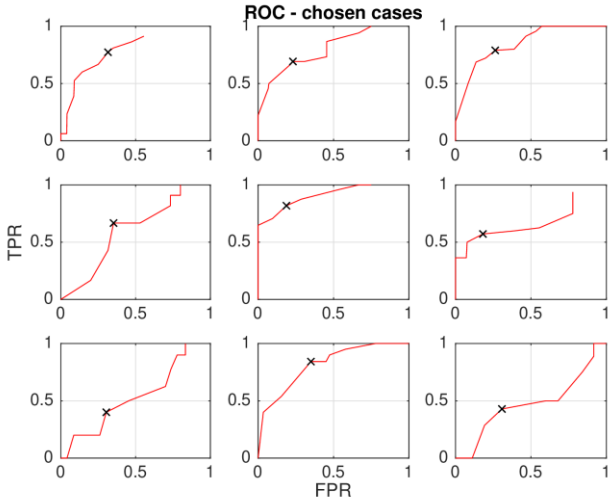


Rys. 7. Analiza krzywej ROC dla wybranej sesji. Na wykresie naniesiono punkty pomiarowe dla wartości progowej (q), wraz z wyznaczoną wartością odległości (d)

5.2. Współczynnik korelacji Matthews'a

Kolejne badania miały na celu ocenę wpływu wybranej funkcji podobieństwa na losowe działanie metody wykrywania. O losowym działaniu metody mówi się w przypadku, kiedy daje ona wyniki w równym stopniu prawdziwe i fałszywe (por. rys. 6). Jest to bardzo niepożądana cecha metody,

która jest związana z jej niedoskonałością lub nieskalibrowaniem. Do realizacji tego celu wykorzystano współczynnik korelacji Matthews'a [49]. Wskaźnik ten uwzględnia wartości wszystkich czterech wskaźników podstawowych (por. wzór 2), a jego wartości posiadają następującą interpretację [8]:



Rys. 8. Wybrane przypadki analizy krzywej ROC dla różnych sesji

- '1' doskonałe działanie (zero fałszywych wykryć i odrzuceń),
- '-1' całkowicie złe działanie (zero wartości prawdziwych),
- '0' działanie losowe.

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP+FP) \cdot (TP+FN) \cdot (TN+FP) \cdot (TN+FN)}} \tag{2}$$

Obliczone wartości współczynnika MCC przedstawiono na rysunku 9. Dla porównania umieszczono także wartości dla pozostałych funkcji podobieństwa. Należy zauważyć, że przedstawione macierze nie są macierzami korelacji. Współczynnik MCC dotyczy bowiem wzajemnej zależności pomiędzy wartościami prawdziwymi (TP, TN) i fałszywymi (FP, FN) metody.

Wyniki tych badań potwierdziły brak losowego działania dla metody wykrywania, która wykorzystuje funkcję podobieństwa K_{bha} oraz częściowo dla metody, która wykorzystuje funkcję K_{spr}.

real	bha	che	cor	cos	euc	hel	skl	man	mah	min	seu	spr
1	0,47	0,26	0,38	0,47	0,02	0,49	0,20	0,18	-0,15	0,20	0,17	0,34
2	0,33	0,24	0,22	0,12	0,11	0,09	0,20	0,41	0,25	0,17	0,16	0,35
3	0,35	0,43	0,35	0,31	0,36	0,36	0,22	0,07	0,30	0,20	0,22	0,20
4	0,32	0,30	-0,01	0,14	0,12	0,06	0,34	0,42	0,33	0,38	0,29	0,41
5	0,68	0,29	-0,09	0,10	0,60	0,25	0,30	0,59	0,44	0,38	0,58	0,68
6	0,43	0,50	0,17	0,46	0,21	0,32	0,18	0,36	0,23	0,62	0,26	0,33
7	0,25	0,19	0,02	0,14	0,06	-0,04	0,36	0,22	-0,15	0,17	0,24	0,35
8	0,49	0,30	0,15	0,37	0,36	0,39	0,40	0,29	0,21	0,17	0,25	0,59
9	0,24	0,37	0,01	0,13	0,00	0,09	0,32	0,33	0,35	0,42	0,48	0,35
10	0,34	0,23	0,17	0,30	0,31	0,25	0,34	0,54	0,39	0,30	0,33	0,19

tts	bha	che	cor	cos	euc	hel	skl	man	mah	min	seu	spr
1	0,57	0,00	0,18	0,36	0,26	0,37	0,06	0,54	0,07	0,26	0,29	0,05
2	0,63	-0,13	0,36	0,13	0,16	0,61	-0,13	0,41	-0,01	0,11	0,19	0,45
3	0,42	0,43	0,45	0,42	0,20	0,67	0,35	0,34	0,20	0,41	0,27	0,35
4	0,28	0,00	-0,06	0,23	0,33	0,22	0,00	0,03	0,25	0,17	-0,17	-0,04
5	0,55	0,24	0,31	0,13	0,62	0,61	0,24	0,38	0,23	0,34	0,42	0,71
6	0,22	0,19	0,26	0,24	0,56	0,23	-0,12	0,27	0,43	0,43	0,21	0,36
7	0,23	0,18	0,11	-0,13	0,10	0,09	0,27	0,09	0,06	0,14	0,28	0,04
8	0,43	0,42	0,61	0,42	0,37	0,45	0,55	0,63	0,04	0,40	0,61	0,46
9	0,14	0,33	0,10	-0,13	0,10	0,05	0,24	0,43	0,43	0,30	-0,01	-0,10
10	0,25	0,27	0,42	0,29	0,51	0,28	0,00	0,23	0,26	0,26	0,24	0,46

Rys. 9. Zestawienie wartości współczynnika korelacji Matthews'a (MCC). Lewa strona rysunku dotyczy kwerendy *real*, prawa natomiast kwerendy *TTS*

6. Wnioski z eksperymentu

W zadaniach wykrywania słów w sygnale mowy wybór funkcji podobieństwa sygnałów nie jest sprawą oczywistą. Na pierwszy plan wychodzi zależność działania funkcji podobieństwa od danych, czyli nagrań sygnału mowy i jego reprezentacji. Zależność ta przekłada się na jakość wykrycia, co można zaobserwować, porównując różnice w wynikach dla kwerend *real* i *TTS*. Dobór funkcji podobieństwa sprowadzać się może do wskazania funkcji, która będzie najbardziej odporna (ang. *robust*) na zmianę danych. W przeprowadzonych badaniach taką funkcją podobieństwa była bazująca na odległości Spearmana (K_{spr}).

Zaproponowany w pracy sposób wyboru najlepszej funkcji podobieństwa bazował na uwzględnieniu sześciu wskaźników jakości. Dzięki temu wybrana funkcja podobieństwa nie była oceniana jednostronnie.

Wykonana w ramach badań dodatkowych analiza krzywej ROC pokazała, że dobierając odpowiednią wartość progową (ozn. q na rys. 7), można znacząco wpłynąć na jakość wykrycia. Warto przy tym zwrócić uwagę, że w ani jednym przypadku nie uzyskano całkowicie złych wyników (to jest przewagi fałszywych wykryć i odrzuceń nad prawdziwymi), stosując funkcję podobieństwa K_{spr}.

Godny zauważenia jest fakt, że różnice wartości wskaźników jakości, uzyskane dla różnych funkcji podobieństwa są niewielkie. Wybór funkcji podobieństwa, bazujący tylko na pojedynczej wartości wskaźnika jakości może być złudny. Przy wyborze funkcji podobieństwa uzasadnione jest zatem

przeprowadzenie co najmniej kilku badań dla różnych danych. Analiza wskaźników jakości dla takich badań daje pełniejszą wiedzę i pozwala oczekiwać, że wybrana funkcja podobieństwa będzie dawała prawidłowe wyniki dla różnych danych.

Literatura

[1] AMGOUD L., DAVID V., DODER D., *Similarity Measures Between Arguments Revisited*. In: Kern-Isberner G., Ognjanović Z. (eds). Symbolic and Quantitative Approaches to Reasoning with Uncertainty, ECSQARU 2019, Lecture Notes in Computer Science, Vol. 11726, pp. 98-107, DOI 10.1007/978-3-030-29765-7_1

[2] BHATTACHARYYA A., *On a measure of divergence between two statistical populations defined by their probability distributions*. Bulletin of the Calcutta Mathematical Society, Vol. 35, 1943, pp. 99-109.

[3] BASENER W., FLYNN M., *Microscene evaluation using the Bhattacharyya distance*. Proc. of SPIE 10780, Honolulu, 2018, DOI 10.1117/12.2327004

[4] BOYTSOV L., *Indexing methods for approximate dictionary searching: Comparative analysis.* Journal of Experimental Algorithmics, Vol. 16, Article 1.1, May 2011, pp. 1-91, DOI 10.1145/1963190.1963191

[5] CHANG H.Y., *An SVM Kernel With GMM-Supervector Based on the Bhattacharyya Distance for Speaker Recognition*. IEEE Signal Processing Letters, 2009, Vol. 16, Issue 1, pp. 49-52, DOI 10.1109/LSP.2008.2006711

[6] CHEN B., WANG H.-M., CHIEN L.-F. LEE L.-S., *A*-Admissible Key-Phrase Spotting With Sub-Syllable Level Utterance Verification*. The 5th International Conference on Spoken Language Processing, Incorporating The 7th Australian International Speech Science and Technology Conference, Sydney, Australia, 1998, pp. 783-786.

[7] CHEN Y.-I., WU CH.-H., YAN G.-L., *Utterance Verification Using Prosodic Information for Mandarin Telephone Speech*. 1999 IEEE International Conference on Acoustics, Speech and Signal Processing. Keyword Spotting Proceedings, ICASSP'99, Vol. 2, Phoenix, AZ, USA, pp. 697-700, DOI 10.1109/ICASSP.1999.759762

[8] CHICCO D., *Ten quick tips for machine learning in computational biology*. BioData Mining, Vol. 10, No. 35, 2017, pp. 1-17, DOI 10.1186/s13040-017-0155-3

[9] CHINCHOR N., *MUC-4 Evaluation Metrics*, In Proceedings of the Fourth Message Understanding Conference, 1992, pp. 22-29, http://www.aclweb.org/anthology-new/M/M92/M92-1002.pdf

[10] DEB K., *Introduction to Evolutionary Multiobjective Optimization*. In: Branke J., Deb K., Miettinen K., Słowiński R. (eds) Multiobjective Optimization. Lecture

Notes in Computer Science, Vol. 5252, 2008, Springer, Berlin, Heidelberg, pp. 59-96, DOI 10.1007/978-3-540-88908-3_3

[11] DUIN R.P.W., PĘKALSKA E., *The Dissimilarity Representation for Structural Pattern Recognition*. Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications, 2011, pp. 1-24, DOI 10.1007/978-3-642-25085-9_1

[12] DUIN R.P.W., PĘKALSKA E., *Non-euclidean dissimilarities: Causes and informativeness*. In proc. Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR), 2010, LNCS, Vol. 6218, Springer, Heidelberg, pp. 324-333, DOI 10.1007/978-3-642-14980-1_31

[13] DUBUISSON M.P., JAIN A.K., *A Modified Hausdorff distance for object matching*. In ICPR94, Jerusalem, Israel, 1994, pp. 566-568.

[14] FAWCETT T., *An Introduction to ROC Analysis*. Pattern Recognition Letters, Vol. 27, No. 8, 2006, pp. 861-874, DOI 10.1016/j.patrec.2005.10.010

[15] FOOTE J., *An Overview of Audio Information Retrieval*. ACM Multimedia Systems, Vol. 7, 1998, pp. 2-10, DOI 10.1.1.39.6339

[16] FUKUNAGA K., *Introduction to Statistical Pattern Recognition*. 2nd Edition, Elsevier Inc, 1990, DOI 10.1016/C2009-0-27872-X

[17] GÜNDOĞDU B., *Keyword Search for Low Resource Languages*. PhD Thesis, Bogazici Universit, 2017.

[18] GÜNDOĞDU B., SARAÇLAR M., *Distance metric learning for posteriorgram based keyword search*. 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, 2017, pp. 5660-5664, DOI 10.1109/ICASSP.2017.7953240

[19] GUPTA K., GUPTA D., *An analysis on LPC, RASTA and MFCC techniques in Automatic Speech Recognition*. 2016 6th International Conference – Cloud System and Big Data Engineering System (Confluence), Noida, 2016, pp. 493-497, DOI 10.1109/CONFLUENCE.2016.7508170

[20] GUPTA P., PUROHIT G. N., RATHORE M., *Number Plate Extraction using Template Matching Technique*. International Journal of Computer Applications, Vol. 88, No. 3, 2014, pp. 40-44, DOI 10.5120/15336-3670

[21] HAASDONK B., BAHLMANN C., *Learning with distance substitution kernels*. In Pattern Recognition – Proc. of the 26th DAGM Symposium, 2004, pp. 220-227, DOI 10.1007/978-3-540-28649-3_27

[22] HAFEN R.P., HENRY M.J., *Speech information retrieval: a review*. Multimedia Systems, Vol. 18, No. 6, 2012, pp. 499-518.

[23] HELLINGER E., (in German) *Neue Begründung der Theorie quadratischer Formen von unendlichvielen Veränderlichen*. Journal für die reine und angewandte Mathematik, Vol. 136, 1909, pp. 210-271, DOI 10.1515/crll.1909.136.210

[24] HENRIKSON J., *Completeness and total boundedness of the Hausdorff metric*. MIT Undergraduate Journal of Mathematics, 1999, pp. 69-80.

[25] HIGGINS A., WOHLFORD R., *Keyword recognition using template concatenation.* IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP'85, Tampa, FL, USA, 1985, pp. 1233-1236, DOI 10.1109/ICASSP.1985.1168253

[26] HOBSON A., CHENG B.-K., *A comparison of the Shannon and Kullback information measures.* Journal of Statistical Physics, Vol. 7, No. 4, 1973, pp. 301-310, DOI: 10.1007/BF01014906

[27] HOLYOAK K.J., THAGARD P., *Mental Leaps: Analogy in Creative Thought*. A Bradford Book series, MIT Press, 1996.

[28] JANSEN A., DURME VAN B., *Efficient Spoken Term Discovery Using Randomized Algorithms*. 2011 IEEE Workshop on Automatic Speech Recognition & Understanding, Waikoloa, HI, 2011, pp. 401-406, DOI 10.1109/ASRU.2011.6163965

[29] JANSEN B., RIEH S.Y., *The Seventeen Theoretical Constructs of Information Searching and Information Retrieval*. In Journal of the American Society for Information Science and Technology, Vol. 61, No. 8., 2010, pp. 1517-1534, DOI 10.1002/asi.21358

[30] JENSEN J.H., ELLIS D.P. W., CHRISTENSEN M.G., JENSEN S.H., *Evaluation of Distance Measures Between Gaussian Mixture Models of MFCCs.* Proceedings of the 8th International Conference on Music Information Retrieval, ISMIR 2007, Vienna, 2007, pp. 107-108.

[31] KAILATH T., *The Divergence and Bhattacharyya Distance Measures in Signal Selection.* IEEE Transactions on Communication Technology, 1967, Vol. 15, No. 1, pp. 52-60, DOI 10.1109/TCOM.1967.1089532

[32] KAMIŃSKA D., SAPIŃSKI T., ANBARJAFARI G., *Efficiency of chosen speech descriptors in relation to emotion recognition*. EURASIP Journal on Audio, Speech, and Music Processing (2017), Vol. 3, pp. 1-9, DOI 10.1186/s13636-017-0100-x

[33] KASSAMBARA A., *Practical Guide to Cluster Analysis in R: Unsupervised Machine Learning (Multivariate Analysis).* Vol. 1, CreateSpace Independent Publishing Platform, 2017.

[34] KESHET J., GRANGIER D., BENGIO S.A., *Discriminative keyword spotting*. Speech Communication, 2009, Vol. 51, No. 4, pp. 317-329, DOI 10.1016/j.specom.2008.10.002

[35] KORŽINEK D., MARASEK K., BROCKI Ł., WOŁK K., *Polish Read Speech Corpus for Speech Tools and Services*. Selected papers from the CLARIN Annual Conference 2016, Aix-en-Provence, 26-28 October 2016, CLARIN Common Language Resources and Technology Infrastructure, No. 136, Linköping University Electronic Press, Linköpings universitet, 2017, pp. 54-62.

[36] KULLBACK S., LEIBLER R.A., *On information and sufficiency*. Annals of Mathematical Statistics, Vol. 22, No. 1, 195, pp. 79-86, DOI 10.1214/aoms/1177729694

[37] KULLBACK S., *Information theory and statistics*. Dover Books on Mathematics, New Edition, 1997.

[38] KWIATKOWSKI W., *Klasyfikacja metodą grupowania cech z uwzględnieniem ich wzajemnej korelacji*. Biuletyn Instytutu Automatyki i Robotyki, Nr 14, 2000, s. 139-146.

[39] KWIATKOWSKI W., *Metody automatycznego rozpoznawania wzorców*. Instytut Automatyki i Robotyki, WAT, Wydanie I, Warszawa, 2001.

[40] KWIATKOWSKI W., *Wykrywanie anomalii bazujące na wskazanych przykładach*. Przegląd Teleinformatyczny, Nr 1-2, 2018, s. 3-21.

[41] KWIATKOWSKI W., *Wstęp do cyfrowego przetwarzania sygnałów*. BEL Studio, WAT, Warszawa, 2003.

[42] LASZKO Ł., *Word detection in recorded speech using textual queries*. Proceedings of the 2015 Federated Conference on Computer Science and Information Systems, M. Ganzha, L. Maciaszek, M. Paprzycki (eds). ACSIS, Vol. 5, pp. 849-853, DOI 10.15439/2015F341

[43] LASZKO Ł., *Using formant frequencies to word detection in recorded speech*. Proceedings of the 2016 Federated Conference on Computer Science and Information Systems, M. Ganzha, L. Maciaszek, M. Paprzycki (eds). ACSIS, Vol. 8, pp. 797-801, DOI 10.15439/2016F518

[44] LASZKO Ł., *Developing keyword spotting method for the Polish language*. Communication Papers of the 2018 Federated Conference on Computer Science and Information Systems, M. Ganzha, L. Maciaszek, M. Paprzycki (eds). ACSIS, Vol. 17, pp. 123-127, DOI 10.15439/2018F178

[45] LEBRET R., COLLOBERT R., *Word Embeddings through Hellinger PCA*. 14th Conference of the European Chapter of the Association for Computational Linguistics, EACL, 2014, pp. 482-490, DOI 10.3115/v1/E14-1051

[46] LI H., HAN J., ZHENG T., ZHENG G., *Mandarin keyword spotting using syllable based confidence features and SVM*. 2nd International Conference on Intelligent Control and Information Processing, Harbin, 2011, pp. 256-259, DOI 10.1109/ICICIP.2011.6008243

[47] LI W., BILLARD A., BOURLARD H., *Keyword Detection for Spontaneous Speech*. 2nd International Congress on Image and Signal Processing, Tianjin, 2009, pp. 1-5, DOI 10.1109/CISP.2009.5303824

[48] LIU D., CHO S., SUN D., QIU Z., *A Spearman correlation coefficient ranking for matching-score fusion on speaker recognition*. TENCON 2010-2010 IEEE Region 10 Conference, Fukuoka, 2010, pp. 736-741, DOI 10.1109/TENCON.2010.5686608

[49] MATTHEWS B.W., *Comparison of the predicted and observed secondary structure of T4 phage lysozyme*. Biochimica et Biophysica Acta (BBA) – Protein Structure, Vol. 405, No. 2, 1975, pp. 442-451, DOI 10.1016/0005-2795(75)90109-9

[50] MANNING CH.D., RAGHAVAN P., SCHÜTZE H., *Introduction to Information Retrieval*. Cambridge University Press, 2008.

[51] MIETTINEN K., *Introduction to Multiobjective Optimization: Noninteractive Approaches*. In: Branke J., Deb K., Miettinen K., Słowiński R. (eds) Multiobjective Optimization. Lecture Notes in Computer Science, Vol. 5252, 2008, Springer, Berlin, Heidelberg, pp. 1-26, DOI 10.1007/978-3-540-88908-3_1

[52] MIETTINEN K., RUIZ F., WIERZBICKI A.P., *Introduction to Multiobjective Optimization: Interactive Approaches*. In: Branke J., Deb K., Miettinen K., Słowiński R. (eds), Multiobjective Optimization. Lecture Notes in Computer Science, Vol. 5252, 2008, Springer, Berlin, Heidelberg, pp. 27-57, DOI 10.1007/978-3-540-88908-3_2

[53] MITRA V., HAUT VAN J., FRANCO H., VERGYRI D., *Feature Fusion for High-Accuracy Keyword Spotting*. Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference Lei Y., et al. on, 2014, pp. 7143-7147.

[54] MOHAMED S.S., ABDALLA A., JOHN R.I., *New Entropy-Based Similarity Measure between Interval-Valued Intuitionstic Fuzzy Sets.* Axioms, Vol. 8, No. 2, 2019, Article-Number 73, DOI 10.3390/axioms8020073

[55] MUSCARIELLO A., GRAVIER G., BIMBOT F., *Audio keyword extraction by unsupervised word discovery*. In Proceedings of the Interspeech, 2009, pp. 2843-2847.

[56] MÜLLER M., *Information Retrieval for Music and Motion*. Springer, Berlin–Heidelberg–New York, 2007.

[57] NIELSEN F., *A generalization of the Jensen divergence: The chord gap divergence.* arXiv preprint, 2017, pp. 1-13, https://arxiv.org/abs/1709.10498

[58] PARDO L., *Statistical Inference Based on Divergence Measures*. Statistics: A Series of Textbooks and Monographs, 1st Edition, Chapman and Hall/CRC, 2006.

[59] PARK A.S., GLAS J.R. *Unsupervised pattern discovery in speech.* IEEE Trans. on Audio, Speech and Language Processing, 2008, Vol. 16, No. 1, pp. 186-197.

[60] PONTIUS R.G., KANGPING S., *The total operating characteristic to measure diagnostic ability for multiple thresholds*. International Journal of Geographical Information Science, Vol. 28, No. 3, 2014, pp. 570-583, DOI 10.1080/13658816.2013.862623

[61] POWERS D.M.W., *Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation*. Journal of Machine Learning Technologies, Vol. 2, No. 1, 2007, pp. 37-63.

[62] QIAO Y., MINEMATSU N., *A Study on Invariance of f-Divergence and Its Application to Speech Recognition*. IEEE Transactions on Signal Processing, 2010, Vol. 58, No. 7, pp. 3884-3890, DOI 10.1109/TSP.2010.2047340

[63] RAIELI R., *Introducing Multimedia Information Retrieval to libraries*. Italian Journal of Library, Archives, and Information Science, Vol. 7, No. 3, 2016, pp. 9-42, DOI 10.4403/jlis.it-11530

[64] SAMMUT C., WEBB G.I. (eds.), *Encyclopedia of Machine Learning and Data Mining*. 2nd Edition, Springer, 2017.

[65] SASAKI Y., *The truth of the F-measure*. 2007, 5 pages, Web resource available at https://www.toyota-ti.ac.jp/Lab/Denshi/COIN/people/yutaka.sasaki/F-measure-YS-26Oct07.pdf

[66] SCHÖLKOPF B., *The Kernel Trick for Distances*. Advances in neural information processing systems, Vol. 13, 2000, pp. 301-307.

[67] SINGH A., YADAV A., RANA A., *K-means with Three different Distance Metrics*. International Journal of Computer Applications, Vol. 67, No. 10, 2013, pp. 13-17, DOI 10.5120/11430-6785

[68] SINGHAL A., *Modern Information Retrieval: A Brief Overview*. Bulletin of the IEEE Computer Society Technical Committee on Data Engineering, Vol. 24, No. 4, 2001, pp. 35-43.

[69] STEHMAN S.V., *Selecting and interpreting measures of thematic classification accuracy*. Remote Sensing of Environment, Vol. 62, No. 1, 1997, pp. 77-89, DOI 10.1016/S0034-4257(97)00083-7

[70] TABIBIAN S., AKBARI A., NASERSHARIF B., *Improved dynamic match phone lattice search for Persian spoken term detection system in online and offline applications*. International Journal of Speech Technology, March 2019, Vol. 22, Issue 1, pp. 205-217, DOI 10.1007/s10772-019-09594-w

[71] TÜSKEA Z., NOLDEN D., SCHLÜTERA R., NEY H., *Multilingual MRASTA features for low-resource keyword search and speech recognition systems*. 2014 IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP), 2014, pp. 7349-7353.

[72] WILPON J.G., RABINER L.R., LEE C., GOLDMAN E.R., *Automatic recognition of keywords in unconstrained speech using hidden Markov*. IEEE Transactions on

Acoustics, Speech and Signal Processing, 1990, Vol. 38, No. 11, pp. 1870-1878, DOI 10.1109/29.103088

[73] YOUDEN W.J., *Index for rating diagnostic tests*. Cancer, Vol. 3, 1950, pp. 32-35, DOI 10.1002/1097-0142(1950)3:1<32::AID-CNCR2820030106>3.0.CO;2-3

[74] ZEDDELMANN VON D., KURTH F., MÜLLER M., *Perceptual audio features for unsupervised key-phrase detection*. Proc. ICASSP2010, 2010, pp. 257-260, DOI 10.1109/ICASSP.2010.5495974

[75] ZHANG Y., *Unsupervised Speech Processing with Applications to Query-by-Example Spoken Term Detection.* PhD thesis, Massachusetts Institute of Technology, 2013.

[76] ZHANG Y., GLASS J.R., *Unsupervised spoken keyword spotting via segmental DTW on Gaussian posteriorgrams*. 2009 IEEE Workshop on Automatic Speech Recognition & Understanding, Merano, 2009, pp. 398-403, DOI 10.1109/ASRU.2009.5372931

[77] ZHU X., PENN G., RUDZICZ F., *Summarizing multiple spoken documents: finding evidence from untranscribed audio*. Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP, Vol. 2, 2009, pp. 549-557.

[78] ZIELIŃSKI T.P., *Cyfrowe przetwarzanie sygnałów od teorii do zastosowań*. Wydawnictwa Komunikacji i Łączności, Warszawa, 2005.

[79] ZIÓŁKO B., GAŁKA J., SKURZOK D., JADCZYK T., *Modified Weighted Levenshtein Distance in Automatic Speech Recognition*. Krajowa Konferencja Zastosowań Matematyki w Biologii i Medycynie, Krynica, 2010, s. 116-120.

Experimental research on the impact of similarity function selection on the quality of keyword spotting

ABSTRACT: In the paper an evaluation of application of selected similarity functions in the task of keyword spotting is described. Experiments were carried out for the Polish language. The results of the research can be used to improve already existing keyword spotting methods or to develop new ones.

KEYWORDS: keyword spotting, signal similarity, quality of detection, dynamic time warping, textual query

Praca wpłynęła do redakcji: 27.11.2019 r.

Information for Authors – rules of papers preparation and reviewing for TELEINFORMATICS REVIEW

The *Teleinformatics Review* is devoted to the publication of original research results in fields of science including, but not limited to: computer science, telecommunication, signal processing, network systems, automation and robotics, etc., which have not been published elsewhere in their entirety or considerable part. If a submitted paper is a part of another published work, e.g. a doctoral dissertation, a postdoctoral thesis, etc., the source work should be included in the list of literature and the editorial office must be informed about it.

In order to publish a paper in the Teleinformatics Review it is necessary to submit it to the editorial office in an electronic form (and possibly its printed copy, one-sided, legible, on white A4 sheets) according to the given template. Only original works in English or Polish will be accepted. The text of the paper should be prepared in the format of Microsoft Word editor (versions 2003 or 2010 are suggested). Appropriate templates can be downloaded from website review.ita.wat.edu.pl (or przeglad.ita.wat.edu.pl). The electronic version submitted to the editorial office should contain a source file of the paper in DOC or DOCX format, with all figures and tables being inserted. The editorial office does not rewrite the text neither make drawings. In addition to the mentioned source file, all figures should be delivered in commonly used image formats (preferably as EPS, JPG, TIFF, or others).

Papers to be published in the Teleinformatics Review are subject to initial acceptance by the editorial office and then are subject to review by two external reviewers. Reviewers and authors do not know each other personal data. The content of the review will be available at the editorial office. If one review is negative (or imprecise) then a third reviewer may be appointed. If both reviews are negative the paper is rejected. If the review indicates a necessity of some corrections, the author must consider all of them and resubmit the improved paper by the determined deadline.

The volume of a submitted paper generally not exceed 20 pages of typescript A4. A deviation from this rule requires agreement of the editorial office. Except the last page, no more than 10% of any page within the paper can be left empty. Figures must be numbered and described below them as well as tables must be numbered and described at the top of them. The literature should hold the form given in the template.

The authors are obliged to submit a statement to the editorial office on the percentage contribution to the creation of the accepted paper, confirming the lack of prior publication of such a work, or a public speech on the subject at a conference or symposium.

The editorial board reserves rights to introduce minor editorial changes to the content of paper without consulting the author. The editorial office insists that no special formatting should be used, which would be inconsistent with the template.

Papers printed in the Teleinformatics Review and their abstracts are placed in the national database of Polish technical journals BazTech as well as on the INDEX COPERNICUS website. Additionally, the papers will be available in the electronic PDF form on website review.ita.wat.edu.pl.

Publication in the Teleinformatics Review does not involve any costs for authors. The editorial office does not charge for submitting, reviewing, preparing for publication and publishing the work. The publication of a paper in the Teleinformatics Review is tantamount to transfer of authors' property rights for publication to the publisher, i.e. the Military University of Technology. By submitting a paper for publication in the Teleinformatics Review, the author agrees, for publication purposes, to the processing by the editorial office the author's name, email address, affiliation, and other contact details.

The editorial office does not return received materials.

The editorial office does not pay fees for papers publishing.

The editor-in-chief may refuse to publish a paper in the following cases:

- if the content of the paper violates the law (principles of secrecy protection, press law, copyright law, etc.) or good manners;
- the author does not agree to introduce all necessary corrections proposed by the editorial board or reviewers;
- the text and illustrative material submitted by the author does not meet the technical requirements given in this document or the template.

Informacje dla autorów – zasady przygotowania tekstu i recenzowania artykułów do PRZEGLĄDU TELEINFORMATYCZNEGO

W Przeglądzie Teleinformatycznym zamieszczane są oryginalne artykuły z dziedzin: *informatyka, telekomunikacja, przetwarzanie sygnałów, systemy sieciowe, automatyka i robotyka* oraz pokrewnych, niepublikowane dotychczas w całości lub w znaczącej części. Jeśli nadesłana praca stanowi część innej opublikowanej pracy, np. pracy doktorskiej, habilitacji, etc., to źródło powinno być umieszczone w spisie literatury, a redakcja powinna być o tym poinformowana.

W celu opublikowania artykułu w *Przeglądzie* niezbędne jest dostarczenie do redakcji treści artykułu w postaci **elektronicznej** według podanego szablonu i ewentualnie jednego egzemplarza wydrukowanego (jednostronnie, czytelnie, na białym papierze formatu A4). Przyjmowane są tylko oryginalne prace w języku angielskim lub polskim. Tekst artykułu powinien być przygotowany w formacie edytora Microsoft Word (wersja 2003 lub 2010 jest zalecana). Szablony dla artykułów są dostępne w pliku na stronie przeglad.ita.wat.edu.pl (lub review.ita.wat.edu.pl). Przekazane do redakcji materiały powinny zawierać plik źródłowy w formacie DOC lub DOCX~~,~~ ze wstawionymi rysunkami. Redakcja nie przepisuje tekstów i nie wykonuje rysunków. Dodatkowo należy dostarczyć pliki źródłowe rysunków (najlepiej w formacie EPS, JPG, TIFF lub innym powszechnie używanym).

Artykuły przeznaczone do opublikowania w *Przeglądzie* podlegają wstępnej ocenie przez redaktora działu, a następnie podlegają recenzji przez dwóch zewnętrznych recenzentów. Recenzenci i autorzy nie znają swoich danych personalnych. Z treścią recenzji można zapoznać się w redakcji. Jeśli jedna z recenzji jest negatywna (lub nieprecyzyjna), może być powołany trzeci recenzent. Jeśli dwie recenzje są negatywne, artykuł jest odrzucany. Jeśli z recenzji wynika konieczność dokonania poprawek w treści artykułu, to autor jest zobowiązany do ich rozpatrzenia i dostarczenia do redakcji poprawionej wersji artykułu, w terminie ustalonym przez redakcję.

Objętość artykułu zasadniczo nie powinna przekroczyć 20 stron maszynopisu A4. Odstąpienie od tej zasady wymaga uzgodnień z redakcją *Przeglądu*. Na stronach tekstu artykułu nie może być pozostawione więcej niż 10% pustego miejsca, za wyjątkiem ostatniej strony. Rysunki należy numerować i opatrzyć (pod spodem) wyczerpującym podpisem. Tabele również muszą być numerowane (tytuł nad tabelą). Literatura może być uszeregowana alfabetycznie oraz powinna mieć postać jak w szablonie.

Autorzy są zobligowani do złożenia w redakcji oświadczenia autorskiego o wkładzie procentowym w powstanie artykułu, braku wcześniejszej publikacji artykułu w przedstawionej formie lub wystąpieniu publicznym na ten temat na konferencji lub sympozjum.

Redakcja zastrzega sobie prawo wprowadzenia niewielkich redakcyjnych zmian w treści artykułu bez konsultacji z autorem. Redakcja nalega, aby **nie stosować** żadnego specjalnego formatowania i **trzymać się ściśle** ustaleń zawartych w szablonie.

Streszczenia i pełne teksty artykułów drukowanych w *Przeglądzie* zamieszczane są w krajowej bazie danych o zawartości polskich czasopism technicznych BazTech oraz na platformie INDEX COPERNICUS. Opublikowane w *Przeglądzie* artykuły będą także w całości udostępnione w internetowej wersji (format PDF) czasopisma, pod adresem przeglad.ita.wat.edu.pl (lub review.ita.wat.edu.pl).

Publikacja w *Przeglądzie* nie wiąże się z żadnymi kosztami dla autorów. Redakcja nie pobiera opłat za zgłoszenie, przygotowanie do druku, recenzję czy publikację pracy. Przekazanie artykułu do publikacji w *Przeglądzie* jest równoznaczne z przekazaniem autorskich praw majątkowych do publikacji na rzecz wydawcy, tj. Wojskowej Akademii Technicznej. Przekazując artykuł do publikacji w *Przeglądzie*, autor zgadza się na przechowywanie i przetwarzanie przez redakcję, w celach publikacyjnych, imienia, nazwiska, adresu e-mail i afiliacji.

Redakcja nie zwraca materiałów dostarczonych do redakcji.

Redakcja nie przewiduje honorariów za opublikowanie artykułu.

Redaktor naczelny może odmówić opublikowania artykułu w przypadku, gdy:

- treści zawarte w materiałach naruszają prawo (zasady ochrony tajemnicy, prawo prasowe, prawo autorskie itp.) lub dobre obyczaje;
- autor nie zgadza się na wprowadzenie wszystkich koniecznych poprawek zaproponowanych przez redakcję lub recenzentów;
- tekst i materiał ilustracyjny złożony przez autora nie spełnia wymagań technicznych podanych w niniejszym dokumencie i szablonie.