Method of Grouping Subjects and Objects in Information Systems

Anastasiya Bondareva, Ilya Shilov
ITMO University
Saint Petersburg, Russian Federation
{bondareva.ad, ilia.shilov}@yandex.ru

Abstract–**The paper considers the problem of dividing users and information systems into groups in organizations of an arbitrary scale. Modern methods do not consider specifics of the organization, business priorities and actual attacking techniques. Two feature sets for subjects and information systems are presented. The features are selected by analysis of dispersion, correlation coefficients and linear regression models built on pairs of features. An evaluation of clustering algorithms applicability to the problem of dividing users and information systems into groups is performed. An algorithm applying the results to real world organizations is constructed. The output of the algorithm can be used for network information security evaluation, access rights management and for designing requirements for network segmentation.**

I. Introduction

In the presence of constant information security threats and increasing number of network attacks the necessity to increase security of information infrastructure is growing. Often this can be achieved with network segmentation which implies division of information network into parts and setting up routing between such segments [1], [2].

In a most simple approach, each department or a group of departments is assigned a subnet which has several allowed and forbidden connections with other subnets. However, such approach does not consider that business processes blur the borders of the segments. The information systems and services are provided individually and rarely follow the organizational hierarchy. Moreover, the information systems and services are also divided into groups depending on their purpose and set of functionalities. Their settings and properties are usually not considered.

Thus, the problem of dividing subjects and objects into groups arises. It must consider not only basic characteristics, but a set of features important for such grouping. The paper observes this problem and provides a method for implementing such grouping. The method considers a set of pre-selected features and is based on clustering methods.

II. State of art analysis

Until now, the problem under consideration has been solved in various ways, both in the organizational and in the technical sphere. Existing access control systems merely implement the already established access rules and do not introduce automation into the division of subjects into groups [3]. The decision to place a machine in a certain network segment is made based on the expert opinion of the system administrator or less frequently security administrator.

Now, the regulatory framework of the Russian Federation has several different classifications of subjects and objects. Usually, system users are considered as an internal intruder, and the systems themselves are taken as a set of personnel and a set of automation tools for their activities (or specific information processing information systems).

For example, in the “Concept of protecting computer equipment and automated systems from unauthorized access to information” [4] the subjects are divided into four groups depending on their capabilities and access rights provided by standard controls within a separate automated system (AS). This approach considers only the permissions granted to users of the system and not the information processed by a specific user. With this approach, employees of different departments with different fields of activity are usually assigned to a single group.

Also in this document, the objects (i.e., AS), are proposed to be classified based on the value of information and the conditions of its processing, the access rights of users and the consequences of improper functioning. However, the exact classification is given in the document "Classification of Automated Systems and Requirements for Information Protection"[5], where AS are divided into classes depending on the number of users and levels of information confidentiality. This approach makes it possible to determine the necessary protection means, however, it does not allow to design network segmentation, since several AS belonging to the same group and even class can operate in completely different areas, include assets of different value, and cause different degrees of damage consequences in case of violation of its functioning.

In the “Methodology for determining current threats to information systems for processing personal data” [6], the subjects are divided into internal and external offenders, which does not allow classification of users (all employees will have the same class - an internal offender). Information system (IS) classification is carried out according to a narrow list of common characteristics, for example, territorial distribution or interaction with other information systems. In this case, information systems are proposed to be divided into 3 groups to determine the correct set of protection measures established by the regulator.

ISSN 2305-7254

In the document establishing the requirements for the protection of information in state IS [7], the IS classes are established depending on their scale and the level of significance of the information being processed.

In the "Decree of the Government of the Russian Federation No. 1119", it is proposed to divide the IS into four classes, depending on the type and amount of processed information (category of personal data) and the type of threats relevant to the system [8]. This approach can be considered more correct than the previous one, since information systems assigned to the same class are likely to function in adjacent areas. Then the same information processing rules will be established for them, which means they can be placed into one network segment for the convenience of building a similar protection system.

A common practice at enterprises of various scale is to divide users into groups depending on departments or work groups. The division of services is performed depending on the purpose and the number of users who work with them. Thus, usually, similar services are placed in same subnets, but this approach does not consider information security requirements. For example, as the corporate mail system and the electronic document management system are used by all employees of the company, they will be placed the same network segment for ease of administration and access. But this approach is dangerous as the systems process information of different criticality and are to be separated at the network level [9], [10].

Existing solutions in the field of information technology provide the ability to configure various groups, roles for subjects and objects of access with different rights and powers. However, they do not have the functionality to automate the definition of groups and develop access matrices. These are only a means of implementing the necessary rules for differentiating access. This approach can impose restrictions on the number of groups or increase the complexity and time spent managing the created groups.

Same can be mentioned about special software and hardware used for automated segmentation. Usually, such technologies provide the system administrator with a way to split the network into logical segments without the need to change it physically. However, these still do not provide the functionality to determine what are the rules to be applied.

Thus, the proposed methods for classifying subjects and objects are rather arbitrary. They do not allow to carry out a full-fledged division of the network into groups according to the selected classes. On the other hand, the observed documents do not consider the real conditions of IS functioning at a particular enterprise.

The purpose of this work is to solve the problem of dividing subjects and objects of information systems into groups and determining their correct place in a network topology. The task is to develop an appropriate method for grouping users and services when performing network segmentation according to information security requirements.

The developed method allows obtaining roles for subjects and security labels for objects automatically, considering the specifics of the organization, characteristics of information systems and already built functional processes and information flows between departments.

III. The Method of Grouping Subjects and Objects

A. General method overview

The construction of a method for grouping users and services in a network infrastructure is carried out in several stages:

- Formation of feature sets for both subjects and objects and selection of informative features.
- Analysis of various clustering algorithms when used for grouping subjects and objects.
- Comparing the proposed method to the existing alternatives.

The selection of features is carried out using various methods of mathematical statistics aimed at identifying the dependencies between features [11-12]. In this work, a sequence of methods is applied consistently. Feature dispersion analysis is used to cut off features with a standard deviation value below the threshold. Then, using the correlation coefficients of Pearson and Spearman, correlated features are sequentially truncated [13]. Finally, regression analysis was used to cut off linearly dependent features. Training and validation take place on the same dataset. For all pairs of features, one feature is considered as the dependent variable, and the second as the target value of the function. Then, for the same two features, the average error value is calculated when the constructed model is applied. The smallest error will be obtained for linearly dependent features.

Clustering algorithms are often used as a basis for conducting primary exploration of a dataset, as well as for reducing the feature space. In addition, such algorithms are often used to search for dependencies in a dataset, considering the unknown initial number of groups (clusters) and (or) the correspondence of objects to these groups. This work analyzes the operation of the following algorithms when grouping subjects and objects of information systems:

- DBSCAN.
- Affinity Propagation.
- Hierarchical clustering.
- Spectral clustering.

The choice of these algorithms is determined by their peculiarities. These algorithms do not impose restrictions on the metrics used, which is useful in a feature space that includes many binary and ordinal features. An important feature of the clustering task is its essential dependence on the priorities of a particular organization for which the analysis is performed. Therefore, the distance function is used, which considers the significance of the features:

$$d\left(x_1, x_2\right) = \sum_{i=0}^{n} w_i \times \left[f_i^1 \neq f_i^2 \right] \quad (1)$$

where d is the distance function between objects, x_1, x_2 are the objects of the sample, w_i is the weighting coefficient of feature i, f_i^j is the value of feature i for object j, n is the number of features.

For methods based on affinity, it is assumed to use affinity matrices or functions-kernels (by analogy with the Gaussian kernel):

$$a(x_1, x_2) = e^{-d(x_q, x_2)} \quad (2)$$

B. Feature generation

As noted earlier, the feature space was formed in an expert way. Then, with the help of various methods for identifying information content and dependencies between groups of signs, the feature space was reduced. For subjects the number of features has been decreased by 40%, for objects - by more than a half. The characteristics selected as informative for the subjects are given in Table I, for objects – in Table II.

When selecting features, the following model parameters were used:

- $\varepsilon = 0.1$ is the threshold variance.
- $\gamma = 0.75$ is the threshold value for recognizing the correlation dependence as strong.

In practice, the set of features can be enlarged. Then a similar approach must be used to cut off dependent and non-informative features.

TABLE I. FEATURES OF SUBJECTS

Features of subjects	
1. General	1.1. Department. 1.2. Position. 1.3. A key figure for the business. 1.4. The need to work in constant access mode (24/7). 1.5. Interaction with external IS. 1.6. Electronic exchange of information with the outside world. 1.7. Potential qualifications.
2. Projects	2.1. The number of projects in which the user participates. 2.2. Project 1; 2.3. Project 2, etc.
3. PC Information	3.1. Elevated privileges on the local PC. 3.2. Elevated privileges in the domain. 3.3. Administrator of IS. 3.4. Presence of development and administration tools. 3.5. Operating system. 3.6. Availability anti-virus software. 3.7. Availability of information about infrastructure and security measures. 3.8. Presence of Internet connection. 3.9. High performance applications with network requirements. 3.10. Permission for self-installation of software.
4. Types of processed information	4.1. Processing of publicly available information. 4.2. Trade secret processing. 4.3. Personal data processing. 4.4. Information processing "for official use". 4.5. The maximum value of the criticality of the information asset.
5. Categories of information	5.1. Scientific and technical information. 5.2. Technological information. 5.3. Production information. 5.4. Management information. 5.5. Financial information. 5.6. Economic information (market information). 5.7. Price information. 5.8. Information planning. 5.9. Meeting information. 5.10. Information about partners and contractors. 5.11. Information about the organization's security management system. 5.12. Information about tenders and auctions.

TABLE II. FEATURES OF OBJECTS

Features of objects	
1. The criticality of the contained information assets	1.1. Maximum. 1.2. Minimum.
2. Categories of information contained	2.1. Publicly available information. 2.2. Personal Information. 2.3. Trade secret.
3. Degree of influence on business processes in case of violations	3.1. confidentiality. 3.2. availability. 3.3. integrity.
4. Technical information	4.1. Operating system. 4.2. Number of open ports. 4.3. Internet connection. 4.4. Number of IS users. 4.5. The share of IS users from the total number of all subjects of access. 4.6. The degree of load on the IC. 4.7. Domain authorization. 4.8. The ability to configure two-factor authentication. 4.9. Availability of an internal access control system. 4.10. Availability of technical support. 4.11. The presence of identified unpatched vulnerabilities. 4.12. Licensed software.
5. Type of connection to IS	5.1. Thin client. 5.2. ssh. 5.3. RDP. 5.4. Web interface. 5.5. Disk mounting.
6. Actual types of attacks from MITER ATT & CK	6.1. Brute-force passwords. 6.2. File leaks on file storages 6.3. Active Directory attacks. 6.4. Attacks on web applications.
7. Other	7.1. Providing resources to third parties. 7.2. Number of dependent ISs. 7.3. Number of influencing ISs. 7.4. Maximum criticality of assets in dependent IS. 7.5. Maximum criticality of assets in influencing IS. 7.6. The number of ongoing projects in the system.

C. Clustering formation

During the analysis, several metrics were used. These must be considered when making decision on which method to use for grouping subjects and objects of information systems:

1) Number of clusters. It determines the number of groups that are created by each algorithm. An increase in the number of clusters makes it possible to achieve a more accurate division of subjects and objects into groups.

2) The number of single-element clusters. It characterizes the number of generated emission-elements; a significant number of such clusters testifies to the low quality of the division into groups. Most often, either entities that have access to a specific set of information resources, or objects that are critical for business sustainability should be considered as single-element clusters.

3) The size of the largest cluster (relative). It characterizes the quality of division into groups. This characteristic allows to assess the quality of the division into groups, since the presence of large clusters does not allow to configure the differentiation of access to resources.

4) Average cluster size. It characterizes the ability of the algorithm to divide subjects and objects into groups of similar size. In fact, this value is equivalent to the number of clusters, therefore, it is not given in the tables below.

5) Average intra-cluster distance. Characterizes the degree of affinity in each cluster. This characteristic is the most significant, as it allows to evaluate the degree of difference between subjects and objects assigned to the same cluster. The calculation of this value implies calculating the average distance between all pairs of objects for each cluster. Then the average value of all the obtained values is found:

$$D = \left(\sum_k \frac{\sum_i \sum_j d(x_i, x_j)}{N_k^2} \right) / N \qquad (3)$$

where N is the number of clusters, N_k is a size of a cluster with index k, d is a distance function.

To analyze the applicability of the algorithms, a dataset obtained from one of the organizations working in the field of aviation instrumentation was used. The data was anonymized in order to comply with the trade secret regime. Table III presents the results for the entities and entities of the Organization's information systems. The dataset (properly depersonalized) is available in [14].

TABLE III. ALGORITHM COMPARISON

Method	Number of clusters	Number of single-element clusters	Maximum cluster size	Average intra-cluster distance
Subjects				
DBSCAN	53	0	200	1,02
Affinity propagation method	144	3	12	0,82
Spectral clustering	5	0	559	2,47
	10	0	308	2,09
	25	0	94	1,6
	49	0	44	1,35
	64	0	41	1,18
	121	1	14	0,93
Hierarchical clustering	47	9	143	1,15
	48	15	317	0,95
	20	6	334	1,29
Objects				
DBSCAN	9	0	23	49,89
Affinity propagation method	11	0	8	88,63
Spectral clustering	5	0	15	132,94
	10	0	8	108,93
	15	2	6	71,74
	20	4	4	71,28
Hierarchical clustering	9	0	11	141,63
	17	8	11	55,79

1) Analysis for subjects. When dividing into groups, it is important to consider that dividing into too many or too few clusters does not allow to highlight the general characteristics of the subjects. Therefore, the algorithm should not generate a lot of single-element clusters or clusters with more than half of the elements. In this case, the average intra-cluster distance should be minimal.

When using the DBSCAN algorithm, many clusters reflect the structure of departments, and production and non-production departments are correctly separated. Outside the task of grouping, this method can be used to find the closest users in terms of functions and privileges.

The method of affinity propagation is characterized by a significantly lower quality of separation, since many users with similar responsibilities and affiliation to departments were assigned into different clusters (for example, managers and their assistants).

When using hierarchical clustering, a lot of single-element clusters were created, which indicates the presence of outliers in the data. However, this algorithm is characterized by the allocation of one or two large clusters, which does not meet the goals of grouping users into groups. In some cases, it can be used to identify situations in which the provision of two or more terminals or access points to the user is redundant.

Spectral clustering was performed for a several target number of clusters, since this value is a parameter of the model. In the case of 5 and 10 clusters, a large cluster is formed, including more than half of the organization employees. With a larger number of clusters, the algorithm is good at distinguishing small groups operating with the same information. The algorithm mixes subdivisions that roughly coincide in functionality, which allows to avoid the creation of many small clusters. With a parameter of 64 clusters, the division is carried out almost in accordance with the departments, and the algorithm still has a good combination of similar departments. This algorithm makes it possible to form clusters with practically the same users in the case when a group of users is characterized by work with two or more PCs (that is, regardless of the operating mode, similar users are combined into clusters). If the required number of clusters significantly exceeds the number of departments, the algorithm starts to work worse and unreasonably divide users within one department.

In accordance with the results obtained, the most suitable algorithm for use as a basis for the user grouping method is spectral clustering when the number of clusters is slightly larger than the number of departments.

2) Analysis for subjects. When grouping objects (services and technologies), other criteria for choosing a clustering algorithm are used. Although the condition for the absence of a cluster that is too large in terms of the number of elements remains, the most critical condition is the smallest intra-cluster distance as a parameter reflecting the maximum similarity of both the purpose of services and their technical characteristics.

When clustering objects, the DBSCAN algorithm correctly separates objects considering the purpose of IS and the logic of their placement in separate subnets. At the same time, a lot of unallocated objects are present. The affinity propagation

method performs almost perfect clustering, however, based on the information obtained, it can be noted that for some ISs (for example, for a domain controller), additional criteria are required to separate them from other IS into separate groups. Hierarchical clustering is characterized by the erroneous assignment of some ISs to groups that are not typical for them. Therefore, this algorithm is also of little use, given the metrics and feature spaces used. Spectral clustering, as in the case of user clustering, shows fairly good results in the case of a sufficient target number of clusters.

Based on this, the method of affinity propagation turns out to be suitable.

D. Grouping method

Consider the proposed method for grouping subjects and objects of information systems:

1) Form the necessary features, namely the most essential characteristics of subjects and objects that were not considered in proposed set of features but are important for the specifics of a particular organization.
2) Select informative features using correlation and dispersion analysis and application of regression.
3) Combine the feature space with the one presented earlier and form the feature description of objects: adapt the set of features proposed in this paper for a specific organization and supplement the adapted set with non-correlated features obtained at stage 2.
4) If necessary, get rid of outliers: for subjects - using hierarchical clustering, for objects - using spectral clustering.
5) Perform clustering: for subjects - using spectral clustering, for objects - using the proximity propagation method.
6) Analyze the results and evaluate for the presence of obvious errors.

The proposed method can be used to solve several types of problems. First, with groups of subjects and objects, it is possible to design a segmented network, in which subjects and objects with similar access rights and functions are placed in the same segments, which reduces the likelihood of successful lateral movement for the attacker. Secondly, the grouping of subjects allows to distinguish roles, in accordance with which the access control rules are formed and special tools for performing software segmentation of the network are configured. Third, the grouping of objects allows to identify services for which the same set of protection measures can be applied, considering similar principles of operation and similar categories of information being processed, which reduces the expenses on information protection tools.

The segmentation of network is possibly the most significant application of the proposed method. As the algorithm unites similar subjects and similar objects into groups placing such subjects and objects into same or adjacent subnets should provide the administrator with less work when setting the network up. Several occasions might happen:

- The users work with same information, which simplifies access control management.
- The services handle same information categories which simplifies designing protecting measures.
- Services with same interfaces (for example, web) fall into same groups which simplifies construction of DMZ.
- The users have similar capabilities which allows to localize threats.

In general, if the groups are used as a basis for network segments the number of firewall rules should decrease. Also managing such network structure is easier as all the new users and services can be added to the network according to their cluster (i.e., their placement is pre-defined). Finally, the number of defensive measures and routing hardware decreases as the number of clusters is not large.

IV. Comparison to existing alternatives

Let us consider the proposed method in comparison with the previously observed methods of grouping subjects and objects, established normatively and existing in practice. Table IV presents the characteristics obtained for the methods of grouping subjects, in Table V - similar characteristics for grouping objects.

TABLE IV. Comparison to existing grouping methods for subjects

Method	Number of clusters	Number of single-element clusters	Maximum cluster size	Average intra-cluster distance
Proposed method	64	0	41	0,0006
Guidance document "Classification of the AS»	4	0	473	5,64
Resolution of the Government of the Russian Federation No. 1119	1	0	669	5,8
Organization’s practice	53	4	65	3,34

TABLE V. Comparison to existing grouping methods for objects

Method	Number of clusters	Number of single-element clusters	Maximum cluster size	Average intra-cluster distance
Proposed method	11	0	8	88,63
Guidance document "Classification of the AS»	2	0	35	138,02
Methodology for determining current threats to the security of personal data	2	0	39	164,79
FSTEC order of 11.02.2013 N 17	2	0	31	199,2
Organization’s practice	14	3	6	72,33

Existing methods consider access subjects only as an internal intruder with several levels of capabilities. At the same time, in the organization, users can be divided into four groups: an ordinary user, a programmer, a local administrator, a network administrator. With this approach, the correct configuration of access control is excluded.

A significant advantage of the proposed method is the ability to create larger number of clusters, which is impossible when using departmental methods. In addition, the created subject groups reflect the practice of differentiating access to resources used in the Organization.

The classification proposed in the Decree of the Government of the Russian Federation No. 1119 is not applicable for IS that do not process personal data and cannot be compared in this study. It also means its inapplicability for the separation of services in enterprises. Other methods proposed in regulatory documents conditionally divide information systems into two types, which indicates the impossibility of setting priorities for the protection of systems, and, accordingly, it is impossible to reason about the economic feasibility of protecting individual resources. Production-critical information systems fall into one segment regardless of different purposes, which does not allow preventively localizing information security incidents and preventing their spread.

Thus, the proposed method reflects the network distribution of network resources that is closest to the real conditions of the Organization's functioning. This indicates the possibility of automating the network segmentation process with minimal influence of the human factor.

IV. Conclusion

The paper deals with the problem of grouping objects and subjects of informatization. Feature descriptions are formed for subjects (users) and objects (information systems). Based on the data set characterizing the real organization of average size, a selection of features was made. Then clustering was performed using various machine learning methods. Based on the results obtained, conclusions were drawn about the degree of success of the clustering methods with the proposed feature set and distance function. A method was developed for dividing users and objects into groups.

The results of the work can be used in the design and modification of the network infrastructure in organizations of any size, as well as for drawing up access matrices and setting up access control systems. At the same time, the formed metric also considers the priorities of the organization's management.

It should be noted that the proposed method has disadvantages, such as: a small number of features, rather large intra-cluster distances, as well as the still existing need for a final expert assessment of the results obtained. Also, in the current work weight coefficients are not used as no assumptions on the business targets and priorities are made. Further analysis is to be performed with the usage of these coefficients.

In the future, it is assumed that the feature space will become more complex to improve the quality of clustering, as well as the formation of new metrics and an assessment of the quality of their work both with the already considered and not mentioned in this work clustering algorithms. The method of forming groups of users and IS will be used in the problems of assessing network security, as well as in the construction of network segmentation methods based on various machine learning algorithms and optimization methods.

V. References

[1] B. Genge, C. Siaterlis. “An Experimental Study on the Impact of Network Segmentation to the Resilience of Physical Processes”, NETWORKING 2012. Lecture Notes in Computer Science, vol.12, 2012, pp. 121-134.

[2] A. Grusho. “Data Mining and Information Security”, Computer Network Security. MMM-ACNS 2017. Lecture Notes in Computer Science, vol.10446, 2017, pp. 28-33.

[3] A. Konev, A. Shelupanov, N. Egoshin. “Functional Scheme of the Process of Access Control”, 3rd Russian-Pacific Conference on Computer Technology and Applications (RPC), 2018, pp. 1-7.

[4] Guidance document. The concept of protecting computer equipment and automated systems from unauthorized access to information. The decision of the Chairman of the State Technical Commission of Russia dated March 30, 1992, Web: https://fstec.ru/tekhnicheskaya-zashchita-informatsii/dokumenty/114-spetsialnye-normativnye-dokumenty/387-rukovodyashchij-dokument-reshenie-predsedatelya-gostekhkomissii-rossii-ot-30-marta-1992-g4

[5] Guidance document. Automated systems. Protection against unauthorized access to information. Classification of automated systems and requirements for information protection. The decision of the Chairman of the State Technical Commission of Russia dated March 30, 1992, Web: https://fstec.ru/tekhnicheskaya-zashchita-informatsii/dokumenty/114-spetsialnye-normativnye-dokumenty/384-rukovodyashchij-dokument-reshenie-predsedatelya-gostekhkomissii-rossii-ot-30-marta-1992-g

[6] Methodology for determining current threats to the security of personal data during their processing in personal data information systems, Web: https://fstec.ru/tekhnicheskaya-zashchita-informatsii/dokumenty/114-spetsialnye-normativnye-dokumenty/380-metodika-opredeleniya-aktualnykh-ugroz-bezopasnosti-personalnykh-dannykh-pri-ikh-obrabotke-v-informatsionnykh-sistemakh-personalnykh-dannykh-fstek-rossii-2008-god

[7] FSTEC order of 11.02.2013 N 17 "On approval of requirements for the protection of information that does not constitute a state secret contained in state information systems", Web: https://fstec.ru/normotvorcheskaya/akty/53-prikazy/702-prikaz-fstek-rossii-ot-11-fevralya-2013-g-n-17

[8] Decree of the Government of the Russian Federation of 01.11.2012 N 1119 "On approval of requirements for the protection of personal data during their processing in personal data information systems", Web: http://www.consultant.ru/document/cons_doc_LAW_137356/

[9] M. Jouini, L. Ben Arfa Rabai, A. Ben Aissa. “Classification of Security Threats in Information Systems”, Procedia Computer Science, vol.32, 2014, pp. 489-496.

[10] L. Wang, S. Jajodia, A. Singhal, S. Noel. “k-Zero Day Safety: Measuring the Security Risk of Networks against Unknown Attacks”, Computer Security – ESORICS 2010. ESORICS 2010. Lecture Notes in Computer Science, vol.6345, 2010, pp. 573-587.

[11] B. Ghojogh B, M. Crowley. “Principal Sample Analysis for Data Reduction”, 2018 IEEE International Conference on Big Knowledge (ICBK), 2018, pp. 350-357.

[12] J. Cai, J. Luo, S. Wang, S. Yang. “Feature selection in machine learning: A new perspective”, Neurocomputing, vol.300, 2018, pp. 70-79.

[13] I. Zikratov, V. Korzhuk, I. Shilov, A. Gvozdev. “Formalization of the feature space for detection of attacks on wireless sensor networks”, 2017 20th Conference of Open Innovations Association (FRUCT), 2017, pp. 526-533.

[14] Dataset for research on services and users grouping, Web: https://github.com/Shtrikh17/datasets/tree/main/services_and_users_groups