



# Network User Behaviour Analysis by Machine Learning Methods

*Michal Turčaník*

*Armed Forces Academy, Liptovský Mikuláš, Slovak Republic  
<http://www.aos.sk>*

## ABSTRACT:

Cyber security is one of the prominent global challenges due to the significant increase in the number of cyberattacks over the last few decades. The amount of transferred data is growing, and a quick reaction to cyber incidents is needed. The paper is a contribution to this effort. There is a possibility to save time and resources by concentrating only on a subgroup of potential threats caused by a specific group of users. The main source of information about a selected group of users is the web access log file, where all the necessary data is stored. The contribution also presents the concept of preprocessing data from the log files to a form useful for clustering. In the next step, a density-based spatial clustering algorithm is applied to create the clusters. Clustering algorithms have been applied to many fields (marketing, business, etc.), but not for the purposes of cyber defence. The created clusters were analysed according to our definition of risky behaviour. After analysis of the clustering results, it was possible to select a potentially dangerous group of users in the specific cluster. The presented method has potential use in different areas of cyber defence and other applications where intelligent classification is required.

## ARTICLE INFO:

RECEIVED: 20 JUNE 2021

REVISED: 10 AUG 2021

ONLINE: 07 SEP 2021

## KEYWORDS:

cybersecurity, web users analysis, machine learning, clustering algorithm, web page categorisation



Creative Commons BY-NC 4.0

## **Introduction**

Today's world and our society heavily depend on communication. People are connected through various IT technologies, and they are "online" almost all the time. Together with the evolution of communication, there is a growing need for the secure transfer of information. Transferred data between users can also contain an unwanted piece of code.<sup>1</sup> By setting some security levels for ongoing communications, we can achieve the security of users on the internet and also organisations. Nowadays, cyber attackers can target not only the government and big companies but also ordinary users.<sup>2</sup> The most used commercial communication systems and applications that can be bought without restrictions can be attacked with a high probability if they are not updated and periodically checked. Attackers can use, for example, back doors and other weaknesses like old firmware. Without a massive investment in time or money, we can find many ways and schemes to attack a communication infrastructure,<sup>3</sup> an operating system and an application.<sup>4</sup> From a military point of view, coordinated cyberattacks against friendly units during operations are very dangerous. Therefore, we must achieve a high-security level for communication networks to successfully fulfil the tasks of military and non-military operations.

As the world's population grows, so does the volume of data and data generated. Collecting, sorting, or analysing this data in real-time is relatively time-consuming and, in some cases, almost impossible. However, everything is changing due to the development of machine learning technology. Today, this technology has been used in several areas for several years and affects and changes our daily lives. The analysis of web communication is an excellent source of information about possible threats to defended networks. The lowest level of analysis is at the packet level.<sup>5</sup> We can check IP addresses, services and protocols. Another possibility is to take into account the knowledge of which pages were viewed by the specific user. From the point of view of service providers, it is a helpful technology called cookies, which are inserted into users' equipment to collect particular information. However, this approach has many setbacks (users do not allow cookies, for example). Another good source of information about user activity on the internet is the list of web pages visited by users. So, to analyse users, we need to dig in and gain information from web log files. Based on this data collected in the log file, we can use automated tools to understand and visualise the browsing behaviour of selected or all users in our network.

Mining of web user behaviour and patterns from the log files saved on the route of packets from source to destination has been an attractive field for research and also for business applications. Existing methods for gaining information about user behaviour include statistical analysis, rules-based association, patterns based on sequential actions, classification by neural networks and clustering.<sup>6</sup> The clustering methods, as one of the approaches, define clusters of users based on similar behaviour. Essential is to analyse the characteristic resulting clusters to understand the connections between users which could not be seen from the first view.

Gained information about the behaviour of users of a computer network, which is processed by different methods, can be used to identify possible threats and risky activities. To prove the ability to use machine learning in cyber defence, a scenario was created. The goal is to find a description of the user with a high potential to pose a cyber threat. The probability of downloading malware to the user's computer is increasing by browsing risky pages. We will define our categories of risky pages: peer-to-peer networks, adult material, downloads, spam sites, and gambling.

In the next step, we will use a specific machine learning method to analyse input data. As a machine learning method, we will apply a clustering algorithm. The selected clustering algorithm will create several groups of users with common behaviour. After that, we will analyse each group of users with the goal of finding characteristics of the potentially risky user.

### Machine learning in cyber defence

Although machine learning has been mentioned more frequently in recent periods, its origins date back to the 1950s. In the field of cyber security, its use was introduced in practice before the year 2000. In the context of cyber protection, machine learning algorithms are used mainly for sorting and analysing samples, identifying similarities, and determining the probability value for the processed object, which is then included in one of three main categories:

- harmful;
- potentially unwanted;
- safe object.

This helps to correctly mark the incoming sample as clean, potentially unwanted or harmful. However, if the goal is to achieve the best possible results, it is necessary to use human expertise and technology to train on a large set of correctly identified harmful and safe samples, based on which the algorithm will learn to distinguish samples. This way of learning is called "supervised learning." During this process, the algorithm learns how to analyse and identify most potential threats and how to actively respond to eliminate them.

Algorithms that are not trained on a predetermined set of sorted data fall into the category of so-called "Unsupervised learning." This approach is more suitable, for example, for looking for similarities and anomalies in the amount of data that might otherwise escape the human eye. At the same time, however, the algorithm does not necessarily learn to separate the good from the bad, or rather, harmless samples from harmful ones. These algorithms find use in working with large volumes of labelled samples, where they help to divide the data into groups so that smaller training sets can be created for other algorithms.

Another possibility is a combination of machine learning with and without a teacher. In this case, only partially marked data is used in the training of the algorithms and the results are then checked and fine-tuned by experts until the

required level of accuracy is reached. Such an approach is used because creating a training set with completely marked data is often time-consuming and expensive. In addition, for some problems, it is not even possible to create a set of completely and correctly marked data.<sup>7</sup>

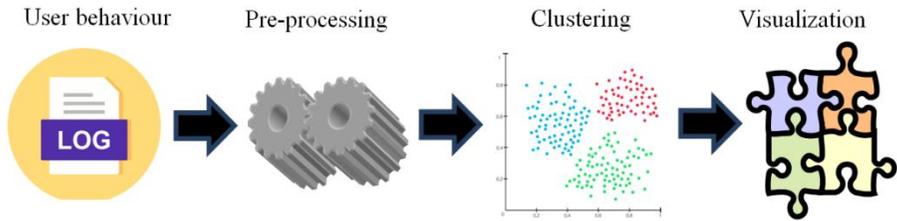
For our scenario, we will choose the machine learning method from the category “Unsupervised learning.” Clustering algorithms are good candidates to solve our problem of dividing users on the basis of their behaviour. The amount of data in the log file, at least for a short time and for small networks with few users, is so huge that any human is not able to find any ties and relations between them in real-time. Very good results can be achieved with the K-means algorithm,<sup>8</sup> but in this scenario, we will use a density-based spatial clustering algorithm.

### Web Users’ Behaviour

The web access log file is one of the possible sources for collecting information about web user behaviour. Data mining approaches can be applied to automatically dig up data with the goal of finding out some habits and tendencies.

Clustering web users is one of the most widely applied machine learning methods for extracting information about users’ online activity. The web access log file is one of the possible sources for collecting information about web user behaviour. Data mining approaches can be applied to automatically dig information with the goal to find out some habits and tendencies.<sup>9,10,11</sup> In that case, the web access log file is analysed in order to divide users on the basis of a selected algorithm into several clusters. Users within the same cluster should have similar properties from the point of view of their activity in cyberspace. The main reason for the application of machine learning algorithms is the amount of data stored in the log files, which is impossible to process by a human operator in real-time.<sup>12</sup>

The main objective of this contribution is to analyse a selected set of access log files. The analysis will be done on the basis of their web activity for a selected time period, and the result will be the creation of clusters of users with similar activity. The resulting clusters can be methodically examined from different points of view. The approach to the behaviour analysis of web users by clustering is shown in Fig 1. The input data is first preprocessed to obtain specific information which is needed for the next step. Pre-processing has been divided into several operations over input data to remove unusable parts. Preprocessed data is divided into clusters with specific properties on the basis of the selected method. To understand the results of clustering, we need to use different methods of visualisation to comprehend the connections between them.



**Figure 1: Behaviour analysis of web users by clustering.**

### Log File Data

The activity of a user can be stored on different devices through the route of data from the source of information to the user. Some data can be stored on the active devices of the computer network (switches, routers, firewalls, etc.) or on dedicated servers (proxy servers). Let's consider data about the internet activity of the user that is stored on the proxy server in the web access log file. A typical file about the web activity of network users can contain the following information: time, duration, a client IP address, a result code, bytes, a method of request, URL, a user ID, a hierarchy code, and a type.<sup>13</sup>

We can analyse when some request created by a user was sent and when an answer to this request was sent back. Each request has its own time for processing, and this is called duration. The requesting user is identified by his IP address. For each request, we have stored its status and the result. The essential information is the amount of transferred data. A good indicator of dangerous activity can be the huge transfer of data from our users to an unknown destination. Another unusual activity can be described by a user repeatedly sending or receiving the same length of content responses. One explanation could be an application during an actualisation. The second explanation of this could be malware communication with control servers. A request method can define the way of transferring data between client and server (GET, POST, PUT, CONNECT, etc.). The URL address of the destination of the communication of our users can also help us to identify possible threats to our network.<sup>14</sup>

For our scenario of behaviour analysis of users, we will use the following subset of presented information from the record of the web access log file:

- Source and destination IP addresses.
- Number of bytes of the transferred data.
- URL address of destination.

For better understanding of the behaviour analysis, a small group of parameters were chosen, but the selected subset may be changed by adding other items from the complete web access log file. By removing unimportant and duplicated information, a new file will be created. Typical communication can be done by the exchange of huge amounts of packets but with the same source and destination IP addresses. These records with identical addresses must be

counted, and the resulting file will consist of only one record for all these occurrences. To the newly created record will be added two new parameters: the number of occurrences (the number of connections between a specific source and destination) and the number of transferred data.

It is impossible to create for every destination IP address a category or cluster. We can create groups of destination IP addresses that have something in common. Behind the IP address is a web page. In the next step, we assign every web page to one or more predefined labels which belong to the particular category. In general, we can classify web pages on the basis of different aspects. It can be on the basis of a subject, a function, or a sentiment. For example, if we decide whether the page topic is about “sport,” “weather forecast,” or “marketing,” we will realise the classification of the subject. In some cases, the function of the web page can be more important. We can classify pages on the basis of the tasks for which that page is used. Typical examples can be a private page, a page with course materials and presentations, or an e-shop page. The last classification is based on the viewer’s feelings about the ideas behind that page. We can also use many other ways to classify web pages: on the basis of the age of the target group of users, the spatial location of users, detection of unwanted ads, and others.<sup>15</sup>

The number of classes in the problem can define the type of classification: binary classification and multiclass classification. Binary classification categorises instances into exactly one of two classes, and multiclass classification distinguishes items into more than two classes. Based on the number of classes that can be assigned to an instance, classification can be divided into single-label classification and multi-label classification. In single-label classification, one and only one class label is to be assigned to each instance. In multi-label classification, more than one class can be assigned to an instance.

This paper will apply to the classification of the subject of the destination web page. To cover nearly all possibilities, a special list of subject categories was created with sixty-four records. To make comparison possible, a used list of categories is identical to that used at work.

After the definition of the list of categories, we can assign to every line in the input file only one category from that list. Some records can be assigned to more than one category. In that case, we use only one label with the highest value of membership in a specific category.

## Density-Based Spatial Clustering

Clustering is very time-consuming for large amounts of data, especially if the search space has several dimensions. Furthermore, the effectiveness of clustering is determined by the chosen method of distance measurement. As a rule, there is no generally correct result. This is due to the fact that the results are drastically different depending on the different methods with different parameters. This makes it difficult to interpret the result and check its correctness. Possibilities would be to carry out this examination, to compare the results of the clustering with known classifications or to evaluate them by a human ex-

pert. However, both variants are only conditionally suitable, since clustering would not be necessary if there were already known classifications and human judgments are always very subjective.

When the preprocessing phase is done, the required data is ready for the next step, which is clustering. The structure of selected data objects and a few examples of real records from web access logs are presented in Table I. A user activity (one line in the table) is saved in a single data entity which is described by the following parameters: source IP address, destination IP address, and the category of the destination. You can find only several lines in Table I, but the full table for a selected time window and group of users may consist of more than 200,000 records.

**Table 1. The user activity.**

Data objects		
Source IP address	Destination IP address	Column B (h)
172.16.25.82	104.31.7.221	News
172.16.25.82	23.25.245.221	Military
172.16.25.17	27.127.251.179	Business & Economy
172.16.17.12	214.151.215.197	Games
172.16.34.45	49.31.67.21	Real estate
172.16.84.130	186.63.157.241	Advertisement

The Density-Based Spatial Clustering Algorithm (DBSCA) is a density-based clustering algorithm similar to mean-shift but with a couple of notable advantages. DSBCA refers to unsupervised learning methods that can find distinctive groups/clusters in the data. DSBCA is based on the idea that a cluster in data space is a contiguous region of high point density. These regions are separated from other clusters by contiguous regions of low point density. DBSCA can find clusters of districts of different shapes and sizes from a huge amount of data. This data can contain noise and outliers. The DBSCAN algorithm uses two parameters: the minimum number of points (a threshold), which are clustered together (mPts), and the distance ( $\epsilon$ ). The distance is used to measure the neighbourhood of analysed points.<sup>16</sup>

DSBCA starts with an arbitrary starting point. The neighbourhoods of this point will be analysed using distance  $\epsilon$ . Points, which are within a distance  $\epsilon$ , are neighbourhood points. If there are an adequate number of points inside a cluster (mPts) then the clustering begins again. The selected data point will be the first point in the new cluster. Otherwise, the point will be labelled as noise (later, this noisy point might become part of the cluster). In both cases, that point is marked as "visited." For every first point in the newly created cluster, all the

points in the neighbourhood with a distance of less than  $\epsilon$  will be part of the new cluster. This process is repeated for all of the new points that have been added to the cluster group. This process is repeated until all points in the cluster are analysed (so all points within the  $\epsilon$  distance of the cluster are labelled). Once the current cluster is finished, a new unvisited point is checked and processed with the goal of stating if it will be another cluster or noise. The algorithm iterates until all points are checked and analysed. Each point is marked as visited and belongs to a specific cluster or becomes a noise.

DBSCA has two main advantages over other clustering algorithms. At first, DBSC does not require you to know the number of clusters at the start. DBSC identifies outliers as noises. The mean-shift clustering algorithm puts these outliers into a cluster even if the data points are very different. DBSC can find arbitrarily sized and arbitrarily shaped clusters.

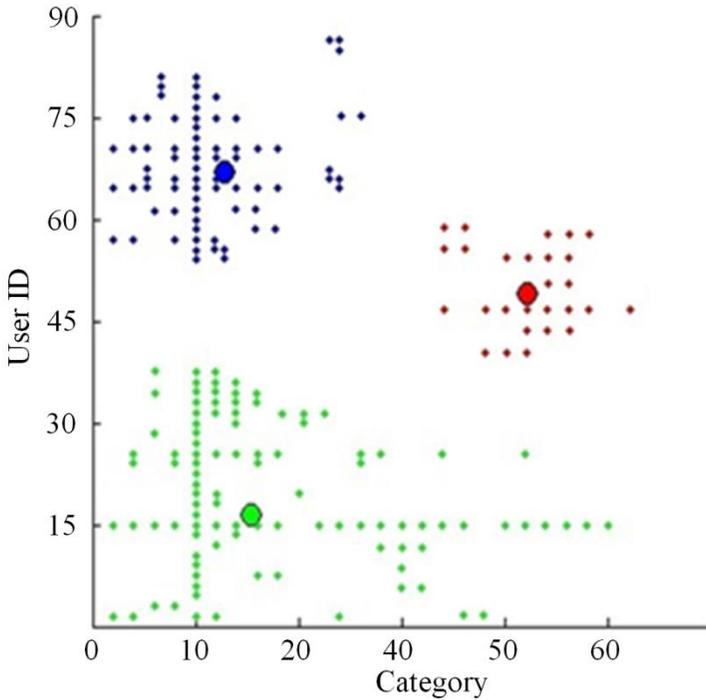
The main disadvantage of DBSCA is that performance is very low for clusters with varying densities. The reason is the different values of mPts and  $\epsilon$  from cluster to cluster when the density varies. Very important and sometimes difficult, can be the estimation of  $\epsilon$  value for high-dimensional data.

To allow clustering of users on the basis of the category of the web they access, we need to assign an ID to each category, which was presented in chapter III. For example, a group of pages which are connected to “Advertisements” will be assigned an ID = 1, and a group of pages which are connected to “Adult materials” will be assigned an ID = 2. This way, all categories will be assigned a number.

## Results

The real meaning of the results of clustering and the understanding of connections between data points is impossible to achieve without visualisation. The technique of visual representation allows us to very easily and very quickly understand and comprehend internal connections between input data points, which represent users of the targeted communication network (with any device).

DBSCA divides all users from the log file into three clusters (Fig 2). The graph shows the distribution of users into individual clusters based on the category of pages they accessed. Every user is assigned an ID (an integer number). In our log file for analysis, we have 89 users. Also, every category is transformed into an integer number from the interval of 1 to 64 (we have defined 64 categories). To understand the results, we will need to analyse the inside of all of these three clusters. The analysis will show their digital communication goals and common behaviour with other users. The members of the first cluster are represented by green, the second cluster by red and the third cluster by blue. In the next step, each cluster will be analysed. Our goal was to find a description of the behaviour of the computer user who is browsing risky pages. Our risky pages belong to the following categories: peer-to-peer networks, adult material, downloads, spam sites, and gambling.



**Figure 2: The results of DBSCA – three clusters of web users.**

**Table 2. The user activity for the 1<sup>st</sup> cluster.**

Nr.	The category	Percentage share
1	Arts	21,68%
2	Entertainment	19,51%
3	Shopping	14,35%
4	Advertisements & Pop-Ups	9,30%
5	Sports	8,20%
6	Leisure & Recreation	4,76%
7	Health & Medicine	4,23%
8	Fashion & Beauty	3,70%
9	Forums & Newsgroups	3,44%
10	Finance	1,85%

The first cluster accommodates more than 20 categories. Only the first 10 of them are presented in table 2. The users in this cluster were the most attracted by pages from the categories “Arts,” “Entertainment,” and “Shopping.” A relatively high is the value of interest in the categories “Advertisements & Pop-Ups” and “Sports.” Around 5% of users were heading to pages from the categories “Leisure & Recreation” and “Health & Medicine.” Less than 4% have “Fashion & Beauty” and “Forums & Newsgroups” categories. The rest of the categories have a percentage value of less than 2%.

The second cluster is smaller than the first one and accommodates more than 13 categories. Only the first 10 of them are presented in table 3. The users in this cluster were the most attracted by pages from the categories “Computers & Technology” and “Search Engines & Portals.” A relatively valuable are the categories “Streaming Media & Downloads,” “Games,” and “Military.” Around 3% of users were heading to pages from the categories “Business,” “Information Security,” “Fashion & Beauty,” and more than 2 % of “News.” Four categories have a percentage value of less than 1% (“Non-profits & NGOs” and the rest of the categories).

**Table 3. The user activity for the 2<sup>nd</sup> cluster.**

Nr.	The category	Percentage share
1	Computers & Technology	39,00%
2	Search Engines & Portals	29,00%
3	Streaming Media & Downloads	8,00%
4	Games	6,00%
5	Military	5,00%
6	Business	3,00%
7	Information Security	3,00%
8	Fashion & Beauty	3,00%
9	News	2,00%
10	Non-profits & NGOs	1,00%

The last cluster includes around 11 categories. Only the first 10 of them are presented in Table 4. The users in this cluster were the most attracted by pages from the categories “Image Sharing,” “Travel,” and “Advertisements & Pop-Ups.” More than 10 % of users were heading to pages from the category “Sports.” Around 7 % of users connect to pages in the “Streaming Media & Downloads” and “Gambling” categories. The category “General” has a percentage value of less than 5 %, and “Dating & Personals” – more than 2 %. “Spam Sites” and “Real Estate” have around 1 %.

**Table 4. The user activity for the 3<sup>rd</sup> cluster.**

Nr.	The category	Percentage share
1	Image Sharing	26,40%
2	Travel	21,30%
3	Advertisements & Pop-Ups	17,00%
4	Sports	10,45%
5	Streaming Media & Downloads	7,50%
6	Gambling	7,10%
7	General	4,35%
8	Dating & Personals	2,30%
9	Spam Sites	1,40%
10	Real Estate	1,20%

The first cluster does not consist of any risky pages from our list. The second cluster contains only one category from our list – „Downloads.” In our case, the only category is not sufficient to assign a given cluster as threatening. Three risky categories of web pages (downloads, spam sites, and gambling) are included in the last cluster, and users from the last cluster are potentially dangerous for a given network. We can create in the future a classifier on the basis of the description of the third cluster. The classifier will be used to detect potentially dangerous users in advance.

The analysis of results of clustering by DBSCA can be realised for a particular user as a single element of the given cluster. The properties of the communication network can be set according to obtained information. Some groups of users could be punished by decreasing the speed of the connection. Other users and their devices (desktops, laptops, and mobiles) will be checked more often on the basis of their risky browsing behaviour. Another possibility is to create groups of users with a specific designation, like spammers, gamers, victims, and attackers.

## Conclusion

The most important reason for creating clustering applications is a special need to answer a particular question. An example can be the question of how to understand and optimise the use of internet communication. When we receive answers, we need to investigate the properties of clusters and their similarities and differences.

One possible way to better understanding the results is through their visualisation. Visualisation can help to accelerate understanding of the current situation on the computer network. Cyber defence must very often solve the problem of why an executed cyber attack was successful. Prevention is always better

and much cheaper than late reaction, which must be done to ensure the resilience of destroyed or disabled resources.

The presented method for web user's behaviour analysis using DBSCA clustering can be applied, and the results can be used to create a classifier of users on the computer network. The main task was to create a concept for processing of huge amounts of information (analysis was done over web access log files) for prevention and identification of offensive cyber activities in our area of responsibility. One obstacle to the designed methodology implementation is the time-consuming process of adjustment and adaptation to a particular target architecture.

Challenges for today's cyber security lie on different levels than at the packet level. However, understanding low levels can put together all the picked evidence for forensic analysis, which must be done after a successful cyber-attack.

Anomaly detection in computer networks is an important topic, which solution is required for the achievement of a better security level. However, to solve this issue, we need to find a huge number of unknown dependencies. The speed of today's communication networks will need the fastest solutions and used algorithms will be created to be executed in an optimal way without redundancies and inefficiencies.

## Acknowledgements

This research was supported by the research and development program granted by the Ministry of Defence of the Slovak Republic within project no. VV-8 "Artificial intelligence and its influence on the development of defence capabilities."

## References

- <sup>1</sup> Jozef Kostelanský and Ľubomír Dederá, "An evaluation of output from current Java bytecode decompilers: Is it Android which is responsible for such quality boost?" *2017 Communication and Information Technologies (KIT)*, Vysoke Tatry, Slovakia, 4-6 October 2017, pp. 1-6, <https://doi.org/10.23919/KIT.2017.8109451>.
- <sup>2</sup> Miroslav Dulík, "Network attack using TCP protocol for performing DoS and DDoS attacks," *2019 Communication and Information Technologies (KIT)*, Vysoke Tatry, Slovakia, 9-11 October 2019, pp. 1-6, <https://doi.org/10.23919/KIT.2019.8883481>.
- <sup>3</sup> Július Baráth, "Optimizing Windows 10 logging to detect network security threats," *2017 Communication and Information Technologies (KIT)*, Vysoke Tatry, Slovakia, 4-6 Oct. 2017, pp. 1-4, <https://doi.org/10.23919/KIT.2019.8883481>.
- <sup>4</sup> Július Baráth, "Network behavior analysis of selected operating systems," *2019 Communication and Information Technologies (KIT)*, Vysoke Tatry, Slovakia, 9-11 October 2019, pp. 1-5, <https://doi.org/10.23919/KIT.2019.8883302>.
- <sup>5</sup> Michal Turčaník. "Packet filtering by artificial neural network," *ICMT 2015: International Conference on Military Technologies 2015*, Brno, Czech Republic, University of Defense, 2015, pp. 415-418, <https://doi.org/10.1109/MILTECHS.2015.7153739>.
- <sup>6</sup> Arun K. Pujari, *Data Mining Techniques* (Orient Longman Private Limited, University Press, 2001), 288.

- <sup>7</sup> Charu Aggarwal, "An Introduction to Clustering Analysis," in *Data Clustering: Algorithms and Applications*, edited by Charu Aggarwal, Chandan Reddy, Chapter 1 (CRC Press, 2014).
- <sup>8</sup> Michal Turčaník, "Web Users Clustering by their Behaviour on the Network," *2020 New Trends in Signal Processing (NTSP)*, Demanovska dolina, Slovakia, 14-16 October 2020, pp. 1-5, <https://doi.org/10.1109/NTSP49686.2020.9229548>.
- <sup>9</sup> Maged Nasser, Naomie Salim, Hentabli Hamza, Faisal Saeed, "Clustering web users for reductions the internet traffic load and users access cost based on K-means algorithm," *International Journal of Engineering and Technology* 7, no. 4 (2018): 3162-3169.
- <sup>10</sup> Alejandro G. Martín, Alberto Fernández-Isabel, Isaac Martín de Diego, and Marta Beltrán, "A survey for user behavior analysis based on machine learning techniques: current models and applications," *Applied Intelligence* 51, no. 3 (2021): 6029–6055, <https://doi.org/10.1007/s10489-020-02160-x>.
- <sup>11</sup> P. Dhana Lakshmi, K. Ramani, and Eswara Reddy B., "Efficient Techniques for Clustering of Users on Web Log Data," in *Computational Intelligence in Data Mining, Advances in Intelligent Systems and Computing* 556 (Singapore: Springer, 2017), [https://doi.org/10.1007/978-981-10-3874-7\\_35](https://doi.org/10.1007/978-981-10-3874-7_35).
- <sup>12</sup> Vaclav Platenka, Antonin Mazalek, and Zuzana Vranova, "The transfer of hidden information in data in the AMR-WB codec," *2019 Communication and Information Technologies (KIT)*, Vysoke Tatry, Slovakia, 9-11 October 2019, pp. 1–5, <https://doi.org/10.23919/KIT.2019.8883461>.
- <sup>13</sup> Tadeusz Morzy, Marek Wojciechowski, and Maciej Zakrzewic, "Web Users Clustering," *Proc. of the 15th International Symposium on Computer and Information Sciences*, Istanbul, Turkey, 2000, pp. 374-382.
- <sup>14</sup> Andrei Kazarov, Giuseppe Avolio, Adrian Chitan, and Mikhail Mineev, "Experience with SPLUNK for archiving and visualisation of operational data in ATLAS TDAQ system," *Journal of Physics: Conference Series* 1085, no. 3 (2018): 1- 4.
- <sup>15</sup> Amit Sheth, *Semantic Web: Ontology and Knowledge Base Enabled Tools, Services, and Applications* (IGI Global,2013), <https://doi.org/10.4018/978-1-4666-3610-1>.
- <sup>16</sup> Kamran Khan, Saif Ur Rehman, Kamran Aziz, Simon Fong, and S. Sarasvady, "DBSCAN: Past, present and future," *The Fifth International Conference on the Applications of Digital Information and Web Technologies (ICADIWT 2014)*, Bangalore, India, 17-19 Feb. 2014, pp. 232-238, <https://doi.org/10.1109/ICADIWT.2014.6814687>.

## About the Author

Michal **Turčaník** is an associate professor at the Department of Informatics at the Armed Forces Academy in Liptovsky Mikulas. He has been teaching different subjects for more than 20 years. He is a Panel Member of the STO IST organization for the Slovak republic. His scientific research is focusing on reconfigurable logic, artificial intelligence and computer networks.