Markov Chain-Based Modeling of IP Network Behavior

1st Smiešková Monika
Faculty of Management Science and Informatics
University of Zilina
Zilina, Slovakia
monika.smieskova@fri.uniza.sk

2nd Smieško Juraj
Faculty of Management Science and Informatics
University of Zilina
Zilina, Slovakia
juraj.smiesko@fri.uniza.sk

***Abstract*—The paper presents a basic methodology for constructing a Markov model of normal IP network traffic for the purpose of detecting anomalies that may indicate, for example, an attacker's lateral movement. Discrete-time Markov chains are employed to model the temporal sequence of events occurring within the network. The main stages of model construction are demonstrated using an example of a departmental network with four users, which facilitates a clearer understanding of the overall methodology. The paper thus reflects current research activities carried out at the department and demonstrates the transfer of research findings into teaching practice.**

***Index Terms*—Markov chain, IP network, anomaly detection, lateral movement, transition matrix, Sysmon, Windows Event Log, SIEM**

I. Introduction

One of the important roles of a university institution is to integrate teaching with current outcomes of scientific research. Knowledge gained from both basic and applied research, as well as from national and international research projects, is therefore continuously incorporated into specialized courses. The aim is to ensure that students acquire knowledge reflecting current developments in information and communication technologies and are able to apply it in practical and research-oriented tasks.

This approach is particularly important in the teaching of mathematical subjects, where the mathematical formalism itself may be overly abstract without a specific application. The research and teaching team therefore also incorporates the mathematical methods used in its research into the courses Discrete Probability, Process Analysis, [1], and Information Network Theory, [2]. In this way, students acquire theoretical knowledge together with its practical context.

An important aspect of this approach is to emphasize the universality of mathematical methods. From the perspective of the mathematical apparatus employed, it is often not essential whether the research problem concerns modeling IP network traffic [3], studying the topological properties of Grassmann manifolds by means of Stiefel-Whitney classes [4], or computational modeling of blood cell flow in a microchannel [5]. With an appropriate interpretation, the same mathematical models can be applied in substantively very different fields. This fact enables students to better understand the significance of mathematical methods and their applicability beyond the scope of a particular course.

With the continued development of digital technologies and the increasingly widespread use of network and cloud services, the security of IP networks is becoming increasingly important. Research results in this area are gradually being incorporated into the modernization of existing networking courses, as well as into the development of new courses focused on modern areas of cybersecurity.

Current research tasks focus on the analysis, modeling, and detection of advanced cyberattack strategies, with particular emphasis on lateral movement. Lateral movement is inherently sequential, as an attacker gradually moves between individual nodes, devices, and systems. When Markov chains are used for its detection, two basic approaches can be considered.

The first approach is based on modeling normal network traffic and subsequently evaluating deviations from typical sequences of states and transitions. The anomalies identified in this way do not necessarily represent lateral movement, but they may indicate suspicious behavior that requires further analysis.The second approach directly models lateral movement, with states and transitions representing individual steps taken by the attacker.

In this paper, we focus on the first approach, namely the construction of a Markov model of normal network traffic and the detection of anomalous sequences that may indicate potential lateral movement.

II. Related Works

Markov chains are used in the field of anomaly and intrusion detection in computer systems primarily to model sequential user behavior. The basic idea is to represent observed behavior as a sequence of discrete events and to estimate the transition probabilities between individual states from historical data.

Ju and Vardi [6] employed a higher-order Markov model to construct a user behavior profile based on sequences of UNIX commands. For each user, they created a characteristic model of normal behavior and subsequently compared newly observed command sequences with the learned profile. The aim was to identify behavior deviating from the standard patterns associated with a given user profile.

979-8-3195-3729-4/26/$31.00 ©2026 IEEE

The use of a homogeneous Markov chain for anomaly detection was presented by Tian et al. [7]. In their model, the states of the Markov chain correspond to individual types of system calls. Transition probabilities are estimated from sequences of system calls collected during normal operation, and the probabilities of newly observed sequences are subsequently evaluated. Sequences with low probability are considered potentially anomalous.

The use of Markov models in a SIEM environment is discussed by Tynymbayev [8]. The author employs a Hidden Markov Model within User and Entity Behaviour Analytics (UEBA) using real-world data obtained from a SIEM system. The input dataset contains approximately 1.2 million events collected over a period of 30 days and includes authentication events, process creation, registry changes, and network communication. Prior to modeling, the events are normalized and correlated.

The paper [12] presents the AnoMark tool, which uses Markov chains and NLP methods to detect anomalies in command lines recorded during process creation. The command lines are decomposed into n-grams, from which a Markov model is constructed, and a probability score is subsequently calculated for newly observed commands. The method was also applied to detect malicious behavior and is designed to be integrated into SIEM systems.

The studies mentioned above support a general procedure for applying Markov chains in this field: heterogeneous security logs are first transformed into a unified event representation, arranged in chronological order, and subsequently represented as a sequence of discrete states. In the proposed model, selected categories of observed events are directly used as states. The transition probabilities between these states are then estimated from historical sequences representing normal network operation. The resulting Markov chain represents a model of standard behavior against which newly observed event sequences can be compared in order to identify unusual or low-probability transitions.

III. Discrete-Time Markov Chain

A Markov chain belongs to the class of stochastic processes. A stochastic process can be defined as a collection of random variables

$$\{X(\omega,t)\}_{\omega\in\Omega,\ t\in T}, \tag{1}$$

where Ω denotes the sample space of elementary random events and T denotes the set of time instants. For a fixed t, $X(\omega,t)$ is a random variable, whereas for a fixed ω, $X(\omega,t)$ represents one realization of the process.

In this paper, we consider only discrete time

$$T=\{0,1,2,\ldots\} \tag{2}$$

and a countable set of states

$$S=\{s_1,s_2,\ldots,s_m\}. \tag{3}$$

In this case, we refer to a *discrete-time random chain*. If, in addition, the chain satisfies the Markov property

$$\begin{aligned} &P(X_{n+1}=s_j \mid X_n=s_i, X_{n-1},\ldots,X_0)\\ &\quad=P(X_{n+1}=s_j \mid X_n=s_i), \end{aligned} \tag{4}$$

it is referred to as a discrete-time Markov chain.

The probabilities of transitions between two states within one time step are denoted by

$$p_{ij}=P(X_{n+1}=s_j \mid X_n=s_i), \tag{5}$$

where we assume that the chain is homogeneous, i.e., the transition probabilities do not change over time. The following conditions hold:

$$\forall i,j\in S,\quad p_{ij}\geq 0,\quad \sum_{j=1}^{m} p_{ij}=1. \tag{6}$$

The transition probabilities are represented by the matrix

$$\mathbf{P}=(p_{ij})_{i,j=1}^{m}. \tag{7}$$

The probability distribution of the states of the Markov chain at time n is denoted by

$$\mathbf{p}(n)=\big(p_1(n),\ldots,p_m(n)\big),\qquad p_k(n)=P(X_n=s_k) \tag{8}$$

For a homogeneous Markov chain,

$$\mathbf{p}(n)=\mathbf{p}(n-1)\mathbf{P}=\mathbf{p}(0)\mathbf{P}^n. \tag{9}$$

Under long-term stable behavior of the Markov chain, for example, when it is aperiodic and transitive, its probability distribution as a function of time n converges to the steady-state distribution of the chain:

$$\lim_{n\to\infty}\mathbf{p}(n)=\boldsymbol{\pi},\qquad \sum_i \pi_i=1. \tag{10}$$

A more detailed analysis of Markov chains can be found in [9], [10] and [11].

IV. Description of Normal IP Network Behavior

The gradual construction of the states of a Markov chain for describing the normal behavior of an IP network will be illustrated using a simple educational example. Let us consider a departmental network with four users: a secretary, a teacher, a network administrator, and the head of the department. Their workstations are denoted by W_S, W_T, W_A and W_B.

For the individual servers, we will use the following unified notation:

E – e-mail server,
I – university information system,
F – departmental file and document server,
L – e-learning server,
A – administration server.

Typical sequences of the secretary's activities are presented in Table I.

TABLE I
TYPICAL STATE SEQUENCES FOR THE SECRETARY

Situation	Normal sequence
1.processing a received document	$W_S \to E \to F \to W_S$
2.working with a student	$W_S \to I \to F \to W_S$
3.receiving an e-mail attachment	$W_S \to E \to F \to E$
4.checking data and replying	$W_S \to I \to E \to W_S$
5.administrative work	$W_S \to F \to I \to W_S$

Let us assume that we have a historical record of 100 sequences observed over a certain period, distributed according to their order in the table into a frequency vector, e.g. $(20, 30, 10, 5, 35)$. Based on the given sequences, we can construct the transition graph shown in Fig.1.

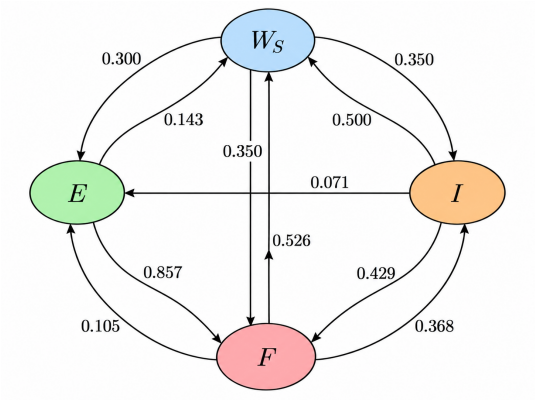


Fig. 1. State transition diagram for the secretary

If we denote the state space for the secretary, in the given order, by $S^{(S)} = \{W_S, E, I, F\}$, we can write the corresponding transition probability matrix as:

$$P^{(S)} = \begin{pmatrix} 0 & 0.300 & 0.350 & 0.350 \\ 0.143 & 0 & 0 & 0.857 \\ 0.500 & 0.071 & 0 & 0.429 \\ 0.526 & 0.105 & 0.368 & 0 \end{pmatrix}.$$

The transition graph does not include self-loops, i.e., the possibility that the chain remains in its current state. We work with a discrete-time model; therefore, the choice of the time-slot length during which system changes are observed depends on the analyst's decision or on the system requirements. For example, if the analyst selects a time slot of 1 second and the secretary is expected to remain at her workstation for 4 seconds under normal conditions, a self-loop with probability 0.8 will occur in state W_S. In our example, where the self-loop probabilities are zero, we may, without loss of generality, assume that the time slot is sufficiently long for the system to transition to a different state.

Typical sequences of the teacher's activities are presented in Table II.

TABLE II
TYPICAL STATE SEQUENCES FOR THE TEACHER

Situation	Normal sequence
1.preparation for teaching	$W_T \to L \to E \to W_T$
2.communication with students	$W_T \to E \to I \to W_T$
3.assignment submission to the LMS	$W_T \to L \to I \to W_T$
4.checking study results	$W_T \to I \to E \to W_T$
5.updating study materials	$W_T \to E \to L \to W_T$

For the state space $S^{(T)} = \{W_T, E, I, L\}$ and the frequency vector $(20, 20, 40, 5, 15)$ we again calculate the transition probability matrix:

$$P^{(T)} = \begin{pmatrix} 0 & 0.350 & 0.050 & 0.600 \\ 0.417 & 0 & 0.333 & 0.250 \\ 0.923 & 0.077 & 0 & 0 \\ 0.200 & 0.267 & 0.533 & 0 \end{pmatrix}.$$

Typical sequences of the administrator's activities are presented in Table III.

TABLE III
TYPICAL STATE SEQUENCES FOR THE ADMINISTRATOR

Situation	Normal sequence
1.user account management	$W_A \to A \to F \to W_A$
2.file service monitoring	$W_A \to F \to A \to W_A$
3.e-learning system management	$W_A \to L \to A \to W_A$
4.incident resolution	$W_A \to A \to L \to W_A$
5.configuration backup	$W_A \to F \to A \to F$

For the administrator, we arrange the states as $S^{(A)} = W_T, A, F, L$ and specify the numbers of state sequences as $(15, 15, 10, 55, 5)$:

$$P^{(A)} = \begin{pmatrix} 0 & 0.700 & 0.200 & 0.100 \\ 0.250 & 0 & 0.200 & 0.550 \\ 0.429 & 0.571 & 0 & 0 \\ 0.846 & 0.154 & 0 & 0 \end{pmatrix}.$$

Typical sequences of the head of the department's activities are presented in Table IV.

TABLE IV
TYPICAL STATE SEQUENCES FOR THE HEAD OF THE DEPARTMENT

Situation	Normal sequence
1.approval of administrative data	$W_B \to I \to E \to W_B$
2.checking data and the schedule	$W_B \to I \to A \to W_B$
3.communication with staff	$W_B \to E \to I \to W_B$
4.verification of user permissions	$W_B \to A \to I \to W_B$
5.resolution of an administrative issue	$W_B \to E \to A \to W_B$

For the head of the department, we have the states $S^{(B)} = \{W_T, E, I, A\}$ and the frequency vector $(20, 20, 40, 5, 15)$:

$$P^{(B)} = \begin{pmatrix} 0 & 0.650 & 0.300 & 0.050 \\ 0.235 & 0 & 0.471 & 0.294 \\ 0.600 & 0.267 & 0 & 0.133 \\ 0.875 & 0 & 0.125 & 0 \end{pmatrix}.$$

For all users, we create a common state space with the states $S = \{E, I, F, L, A\}$ and determine the transition probabilities only from transitions between the individual servers, omitting the workstations:

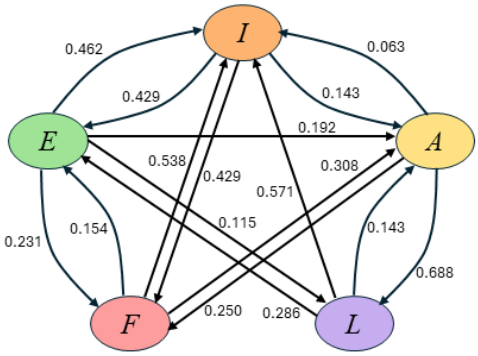


Fig. 2. State Transition Diagram for the Common Environment

The transition probability matrix has the form:

$$\mathbf{P} = \begin{pmatrix} 0 & 0.462 & 0.231 & 0.115 & 0.192 \\ 0.429 & 0 & 0.429 & 0 & 0.143 \\ 0.154 & 0.538 & 0 & 0 & 0.308 \\ 0.286 & 0.571 & 0 & 0 & 0.143 \\ 0 & 0.063 & 0.250 & 0.688 & 0 \end{pmatrix}.$$

Using the transition matrix, we can easily determine the probabilities of state sequences of a given length. For example, the least probable 5-state sequence is

$$A \to I \to A \to I \to A \to I$$

with probability

$$0.063 \cdot 0.143 \cdot 0.063 \cdot 0.143 \cdot 0.063 = 0.0000051132$$

Such very low values indicate that the given sequence is unlikely to occur. The occurrence of such a situation, involving repeated switching between the information system and the administration server, may represent a security issue.

Each user has a defined initial probability distribution of connections to the servers, $\mathbf{p}(0)$, for example, for the secretary:

$$\mathbf{p}(0) = (0.30, 0.30, 0.35, 0, 0)$$

This is essentially a modified first row of the user's transition matrix. Using the system transition matrix, we can predict the probabilistic behavior of the user

$$\mathbf{p}(n) = \mathbf{p}(n-1)\mathbf{P} = \mathbf{p}(n-2)\mathbf{P}^2 = ... = \mathbf{p}(0)\mathbf{P}^n \quad (11)$$

The presented example represents a simplified model that clearly illustrates the principle of constructing a Markov chain. In a real IP network, the basic procedure remains the same: events are ordered chronologically, assigned to states, and the transition matrix is estimated from the observed transition frequencies. A real network contains a large number of users,

TABLE V
EVOLUTION OF STATE PROBABILITIES ACROSS INDIVIDUAL STEPS

k	E	I	F	L	A
0	0.3000	0.3500	0.3500	0	0
1	0.2040	0.3269	0.2194	0.0345	0.2154
2	0.1839	0.2456	0.2412	0.1717	0.1584
3	0.1916	0.3228	0.1875	0.1302	0.1693
4	0.2046	0.2744	0.2251	0.1385	0.1593
5	0.1920	0.3047	0.2048	0.1331	0.1676
6	0.2003	0.2854	0.2170	0.1374	0.1625

devices, services, and event types, while the state space must be designed so that it remains interpretable. Sequences and states are generated automatically from logs, network flows, and authentication records. As in the illustrative example, the resulting Markov model describes the probabilistic structure of normal network traffic, against which new sequences can be evaluated and unusual behavior can be identified.

V. METHODOLOGY FOR CONSTRUCTING A MARKOV MODEL

The construction of a Markov model of an IP network for the purpose of anomaly detection can be divided into several consecutive steps.

A. Data Collection and Normalization

The input may consist of historical data obtained from network flows, authentication records, Windows Event Logs, Sysmon, or SIEM systems. Each event can be represented, for example, in the form

$$e_k = (t_k, u_k, h_k^{src}, h_k^{dst}, q_k, a_k), \quad (12)$$

where t_k is the event time, u_k is the user account, h_k^{src} is the source device, h_k^{dst} is the destination device, q_k is the event type, and a_k is the set of additional attributes. Let $\mathcal{E}$ denote the set of all normalized events:

$$\mathcal{E} = \{e_1, e_2, \ldots, e_m\} \quad (13)$$

Before constructing the model, the data must be chronologically ordered, user and device identifiers must be unified, and heterogeneous records must be transformed into a common format.

B. Construction of the State Space and State Sequences

Let S be the state space of the Markov model. Individual states S_i do not represent specific log records, but semantically and operationally interpretable classes of events.

$$S = \{s_1, s_2, \ldots, s_m\} \quad (14)$$

The states may represent, for example, DNS communication, web communication, access to a file server, e-mail communication, authentication, etc.

Each normalized event e_k must be assigned to exactly one state from the state space S. This assignment is performed by the mapping function

$$g(.) : \mathcal{E} \to S, \qquad g(e_k) = s_k \tag{15}$$

By successively mapping the normalized events, we obtain the corresponding state sequence, for example:

$$(e_1, e_2, e_3, e_4, e_5, \ldots) \longrightarrow (s_5, s_1, s_2, s_3, s_4, s_1, \ldots) \tag{16}$$

The sequence constructed in this way represents the input for the subsequent estimation of the transition probabilities of the Markov chain:

$$P(X_{k+1} = s_j \mid X_k = s_i) = p_{ij},$$

It is appropriate to construct sequences for a specific user account, device, or pair of communicating devices. For long records, network traffic can be divided into time windows or sessions. The aim is to ensure that consecutive states represent events for which the temporal ordering is meaningful to analyze.

C. Selection of the Time Window

The length of the time window Δt affects the characteristics of the model. A time window that is too short may divide naturally related activity into several independent sequences, whereas a time window that is too long may combine events that are not semantically related. In a practical implementation, it is therefore appropriate to compare several values of Δt and select one for which the model is stable while remaining sufficiently sensitive to changes in the ordering of events.

It is also important to decide whether repeated identical events are retained as multiple transitions $S_i \to S_i$, merged before mapping, or whether a time window Δt is selected such that the probability of remaining in the same state satisfies $p_{i,i}(\Delta t) = 0$. This decision directly affects the diagonal elements of the transition matrix (see the previous example).

D. Estimation of Transition Probabilities

From historical data representing normal operation, we count $N_{i,k}$, the number of transitions from state s_i to all other reachable states $s_k \in S$. The maximum likelihood estimate of the transition probabilities p_{ij} has the form

$$\hat{p}_{ij} = \frac{N_{ij}}{N_i} = \frac{N_{ij}}{\sum\limits_{\forall k} N_{ik}}. \tag{17}$$

In practical applications, some transition counts may be zero, $N_{ij} = 0$. In such a case, we can use so-called Laplace smoothing

$$\hat{p}_{ij} = \frac{N_{ij} + \alpha}{\sum\limits_{\forall k} N_{ik} + \alpha |S|}, \qquad \alpha \to 0. \tag{18}$$

where α is the Laplace smoothing constant, and $|S|$ is the number of considered states.

For example, if state s_2 has zero transitions to state s_3, $N_{2,3} = 0$, but has a total of 130 transitions to states $s_1, ..., s_5$, $N_2 = 130$, then the estimate of the probability is $\hat{p}_{2,3} = 0/130 = 0$. However, if we choose $\alpha = 1$, knowing that $|S| = 5$, we obtain

$$\hat{p}_{2,3} = 1/135 = 0,007407$$

Naturally, the remaining probabilities $\hat{p}_{ik}$ must be slightly adjusted so that $\sum\limits_{\forall k} \hat{p}_{ik} = 1$ holds, i.e., the sum of the probabilities in each row remains equal to one.

If some transitions between states are impossible from the perspective of the network architecture or are explicitly prohibited, it is important to keep the corresponding transition probabilities equal to zero, $\hat{p}_{l,k} = 0$.

The result of the preparatory modeling phase is a reference matrix, i.e., an estimate of the transition matrix $\widehat{\mathbf{P}} = (\hat{p}_{ij})$, which describes the typical sequential structure of normal operation. High values of $\hat{p}_{ij}$ represent common event transitions, while low values represent rare transitions.

VI. Anomaly Detection Relative to Normal Operation

After constructing the reference Markov model of normal operation, each newly observed sequence can be compared with the probabilistic structure learned from historical data.

Let us consider a newly observed state sequence:

$$S_{new} = (s_0, s_1, \ldots, s_n), \qquad s_k \in S, \tag{19}$$

Its probability is given by

$$P(S_{new}) = p_{s_0}(0) \prod_{k=0}^{n-1} \hat{p}_{s_k s_{k+1}}, \tag{20}$$

It follows from the above relation that even a single highly unlikely transition can significantly reduce the probability of the entire sequence. This fact forms the basis of anomaly detection using a Markov model.

For longer sequences, the product of a large number of small probabilities may become numerically very close to zero. For this reason, it is more convenient to work with the logarithm of the probability.

From (21) we obtain

$$\ln P(S_{new}) = \ln p_{s_0}(0) + \sum_{k=0}^{n-1} \ln \hat{p}_{s_k s_{k+1}}, \tag{21}$$

Since $0 < \hat{p}_{ij} \leq 1 \implies \log \hat{p}_{ij} \leq 0$, frequent transitions with probabilities close to one therefore contribute values close to zero, whereas rare transitions produce strongly negative terms.

In general, the observed sequences may have different lengths; therefore, for their mutual comparison, we define a normalized negative logarithmic score

$$A(S) = -\frac{1}{n}\left[\ln p_{s_0}(0) + \sum_{k=0}^{n-1} \ln \hat{p}_{s_k s_{k+1}}\right]. \qquad (22)$$

For anomaly detection itself, it is necessary to determine a decision threshold θ_A, whose exceedance indicates an anomaly, $A(S) > \theta_A$.

The output of the detection phase therefore does not need to be limited to binary information *normal/anomalous*. For each analyzed sequence, the following attribute vector can be stored:

$$\left(u,\, h^{src},\, h^{dst},\, \mathbf{x},\, A(\mathbf{x}),\, k^*,\, a_{x_{k^*} x_{k^*+1}}\right),$$

where the individual attributes have the following meaning:

u	user identity,
h^{src}	source device,
h^{dst}	destination device,
$\mathbf{x}$	analyzed state sequence,
$A(\mathbf{x})$	overall anomaly score of the sequence,
k^*	position of the most unusual transition,
$a_{x_{k^*} x_{k^*+1}}$	local anomaly score of this transition.

Such an output is suitable for subsequent correlation in a SIEM system (Security Information and Event Management), because the security analyst receives not only information that the sequence has been evaluated as anomalous, but also information about which transition or combination of transitions resulted in a high anomaly score.

VII. Discussion

An advantage of the model is its straightforward mathematical interpretation. A security analyst can retrospectively identify the specific transitions that caused a high anomaly score. The model therefore provides a more transparent alternative to purely black-box methods.

The main limitation is the first-order Markov property. In a real network, the significance of an event may depend on a longer history than the immediately preceding state. This limitation can be mitigated by a more appropriate construction of the state space, by using higher-order Markov chains, or by employing hidden Markov models.

VIII. Conclusion

A standalone model of normal IP network behavior based on a discrete-time Markov chain was proposed. Heterogeneous network and system events are transformed into a random chain over a finite set of states. A reference transition matrix is estimated from historical data, and new sequences are evaluated using a normalized logarithmic anomaly score. The model provides a mathematically simple and interpretable way to identify deviations from normal network operation without the need to explicitly model a specific type of attack.

We presented material that can serve as the basis for a lecture, together with a motivating example illustrating the fundamental principles of IP network modeling using Markov models for the purpose of detecting lateral movement.

Acknowledgment

Co-financed by the European Union through the Slovakia Programme under project No. NFP401101C360: Research and development of advanced artificial intelligence solutions for the detection of cyber threats and defense against sophisticated attacks

References

[1] M. Klimo, O. Škvarek, J. Smieško, S. Foltán and O. Šuch, "Vowel Recognition Supported by Ordered Weighted Average," in *Emergent Trends in Robotics and Intelligent Systems*, Advances in Intelligent Systems and Computing, vol. 316, Springer, 2015, pp. 247–253. doi:10.1007/978-3-319-10783-727.

[2] J. Smieško, P. Segeč and M. Kontšek, "Machine Recognition of DDoS Attacks Using Statistical Parameters," *Mathematics*, vol. 12, no. 1, art. 142, 2024. doi:10.3390/math12010142.

[3] J. Smieško, "IP Network Management of Sources for IP Traffics," *Scientific Papers of the University of Pardubice, Series D, Faculty of Economics and Administration*, vol. 21, no. 32, pp. 109–117, 2014.

[4] J. Korbaš and P. Novotný, "On the Dual Stiefel-Whitney Classes of Some Grassmann Manifolds," *Acta Mathematica Hungarica*, vol. 123, no. 4, pp. 319–330, 2009.

[5] S. Molčan, M. Smiešková, H. Bachratý, K. Bachratá, and P. Novotný, "Classification of red blood cells using time-distributed convolutional neural networks from simulated videos," *Applied Sciences*, vol. 13, no. 13, Art. no. 7967, 2023.

[6] W.-H. Ju and Y. Vardi, "A Hybrid High-Order Markov Chain Model for Computer Intrusion Detection," *Journal of Computational and Graphical Statistics*, vol. 10, no. 2, pp. 277–295, 2001. doi:10.1198/10618600152628068.

[7] X. Tian, M. Duan, C. Sun, and W. Li, "Intrusion Detection Based on System Calls and Homogeneous Markov Chains," *Journal of Systems Engineering and Electronics*, vol. 19, no. 3, pp. 598–605, 2008. doi:10.1016/S1004-4132(08)60126-7.

[8] B. Tynymbayev, "Applying Hidden Markov Models for User and Entity Behavioural Analytics Model," *Systems and Soft Computing*, vol. 8, p. 200467, 2026. doi:10.1016/j.sasc.2026.200467.

[9] K. Janková, S. Kilianová, P. Brunovský and P. Bokes, *Markovove reťazce a ich aplikácie*. Bratislava: EPOS, 2015.

[10] J. Kalas, *Markovove reťazce – skriptum*. Bratislava: Polygrafické stredisko UK, 1993.

[11] J. Smieško, *Operačná analýza II. Základy teórie hromadnej obsluhy*. Žilina: MC Energy s.r.o., 1999.

[12] A. Junius, "AnoMark – Anomaly Detection in Command Lines with Markov Chains," *FIRST Conference 2023*, ANSSI – French National Cybersecurity Agency, 2023.

[13] D. P. Bertsekas, R. G. Gallager, *Data Networks*, 2nd ed. Englewood Cliffs, NJ: Prentice Hall, 1992. ISBN 0-13-200916-1.

[14] Microsoft Sysinternals, *Sysmon – System Monitor*.

[15] Microsoft, *4624(S): An account was successfully logged on*.