Detection of malicious requests aimed at disrupting the availability of cyber-physical systems

Anastasia Iskhakova*

Financial University under the Government of the Russian Federation, Moscow, 125993, Russia

Abstract. The work is devoted to solving the problem of algorithmization of the security management processes of cyber-physical systems by detecting malicious requests aimed at disrupting the availability of management interfaces. Particular attention is paid to attacks aimed at denial of service of cyber-physical systems by sending HTTP-flood to web management interfaces. This paper proposes algorithmic provision for comprehensive adaptive analysis of incoming requests. The proposed algorithm for the detection of malicious requests analyses the activity of the investigated components of the cyber-physical system's web service at various network levels. The work applies a visual analysis and data processing method based on the representation as a single normalized set. The raw data of the analysed queries is grouped in a special way to detect a particular anomaly as a suspected threat. Examples of data changes and security responses are given. The experimental results confirm that the proposed algorithmic software achieves first- and second-order error reduction compared to the commonly used regression models in modern application-layer firewalls. The results obtained can be applied to the further development of the theory of information security, in particular the information security of cyber-physical systems and systems of processing of especially protected confidential information.

1 Introduction

The urgency to improve the defense against attacks on the availability of cyber-physical systems is based on the high level of importance of the possibility of degradation of the services as a result of cyber-physical attacks [1, 2]. Such attacks are often intended to paralyze or damage critical infrastructure such as power systems, transportation networks, industrial processes, and other facilities that depend on computer systems to function. One example of such attacks is a DDoS (Distributed Denial of Service) attack, in which attackers use botnets (networks of infected computers) to simultaneously direct a huge number of requests to a target system. This can lead to server overload and temporary inaccessibility for legitimate users. Another example of an availability attack is physically targeting infrastructure. For

* Corresponding author: shumskaya.ao@gmail.com

example, attackers may attempt to hack into the control system of a vehicle, such as a car or drone, in order to disable or control it in a malicious manner [3].

Such attacks can lead to serious consequences, including system disruption, threat to human life and health, and significant economic losses. Protecting cyber-physical systems from such attacks is therefore a critical task for organizations and nations [4, 5]. This includes developing and implementing security measures, training personnel, keeping software up-to-date, and monitoring systems for possible vulnerabilities and anomalies.

In recent years, we have observed information confrontations that often utilize DDoS attacks against various information objects. This type of attack disrupts not only the local functioning of the system, but also disrupts a number of critical processes, particularly affecting cyber-physical systems. This class of attacks focuses on significantly impacting the physical space by exploiting vulnerabilities in the computing circuit and communication infrastructure that provide interfaces to systems for monitoring various sensors and controlling individual actuators. For example, in a successful attack on an authentication system, an attacker could take control and manage the computational or communication components of gas, water, and heating pipelines, causing damage to property or the environment and putting people's lives at risk. As a result, security is widely regarded as one of the most important challenges in the design of robust cyber-physical systems [6].

Previously, author has proposed algorithmic software to detect the sources of malicious requests in cyber-physical systems [7]. In this study, a new approach for detecting abnormal activity in traffic is discussed and demonstrated: based on deep packet analysis of traffic to detect flood attacks and payloads for NGFW systems, using classical machine learning algorithms.

2 Current state of DDOS protection in cyber-physical systems

Detecting the sources of malicious requests in cyber-physical systems is an important step to defend against availability attacks. There are several methods that can be used for this purpose:

- Network traffic monitoring: This method involves observing network traffic passing through the system to detect anomalies or unusual activity. Monitoring systems that analyze data packets going to and from the system and look for signs of attack or unusual behavior can be used [8, 9].
- Analyze event logs: Event logs contain information about actions and activity on the system. Analyzing these logs can help identify unusual or suspicious events that may indicate the presence of malicious requests. You can use tools to collect and analyze event logs to automatically detect suspicious activity [10-12].
- The use of intelligent intrusion detection systems (IDSs): IDSs are systems that monitor a network or computer system for attempts at unauthorized access or use. IDSs can be configured to detect malicious requests based on known attack signatures or anomalies in network traffic [13, 14].
- Utilizing machine learning: Machine learning techniques can be applied to detect malicious requests in cyber-physical systems. Machine learning models can be trained on known normal activity data and used to detect deviations from that norm. For example, classification algorithms can be used to determine whether a request is malicious or legitimate [15-18].

It is important to note that these methods can only be effective if properly configured and continuously updated. It is also important to use a combination of different methods to increase the effectiveness of malicious request detection and minimize false positives.

Most of the publications on the topic of this study are based on datasets and results from the study on cyber-security at the Canadian university [8]. For example, many papers, such

as [9-11], review classical machine learning algorithms and their software implementations. However, when attempting to replicate the experiments presented in these articles with real traffic, the model classified the traffic as normal. When analysing the dataset, it was found that the throughput in the traffic collection experiments was significantly reduced compared to the real-world throughput of today's networks. This led to the creation of a suitable dataset and the development of training models to test the hypothesis more closely to reality. This change entailed some adjustments to traffic collection and labelling parameters, as well as changes to model training parameters. All of these works are academic in nature and aim to study and compare algorithms, but it is incorrect to compare algorithms on an incorrect dataset.

Thus, machine learning can be applied to such problems and it is capable to perform well on test data. However, most of the reviewed studies have one drawback – they were tested only on the test data from the dataset itself; the test data were obtained by cross-validation, which cannot guarantee the performance of the model on real network traffic in a conventional network [19].

3 Methods and datasets

Some of the best-known and publicly available data sets are the following: DARPA1998, KDD Cup 1999, Kyoto 2006, NSL-KDD 2009, ISCX 2012, CTU-13, UNSW-NB15, CIDDS-001, UGR-16, CICIDS 2017, CICIDS 2018 and others. These sets are used by the vast majority of researchers to validate the detection algorithms under investigation. Given the requirements to the data relevance in the set and the availability of a quality traffic sniffer, one of the popular datasets – CIC-IDS – was chosen. It is worth noting that CIC-IDS is a dataset that includes data collected between 2017 and 2019, during which time three sets with different types of attacks were generated. The data itself is presented as 80 features collected from traffic based on the CICFlowMeter sniffer. As this sniffer is freely distributed software, this allows using it on real traffic for testing.

Infiltration – in this case, the attack collects information about the network that has become vulnerable due to a virus file having penetrated the system. The attack scenario is as follows: the virus infiltrates the system through a file and the attacker uses the vulnerability created by the file to execute a port scanning attack on the network.

Bot –t Ares botnet uses a special attack tool written in Python programming language to coordinate and execute its malicious actions. This tool – Ares Bot – is software that is installed on infected computers, turning them into bots that obey commands and are controlled by attackers. Ares Bot communicates with the remote command and control server (C&C server) of the Ares botnet which controls the entire botnet infrastructure. The commands sent from the C&C server include instructions for bots, for example, to launch DDoS attacks on target systems, to distribute malware, to collect sensitive data, or to perform other malicious operations.

BENIGN – this category represents normal and legitimate network traffic with no signs of malicious activity.

DDOS attack-HOIC – a DDoS attack using a HOIC (High Orbit Ion Cannon) tool. HOIC is a tool designed to attack a target system in a coordinated manner by sending large numbers of requests or packets to overwhelm it.

DoS attacks-Slowloris – a DoS (Denial of Service) attack using the Slowloris method. Slowloris performs the attack by occupying available connections on the target server by sending incomplete HTTP requests and delaying their completion, which results in blocking new connections.

DoS attacks-GoldenEye – a DoS attack using the GoldenEye tool. GoldenEye is designed to deplete a target system's resources by sending multiple invalid requests, making it unavailable to legitimate users.

DoS attacks-Hulk – a DoS attack using the Hulk tool. Hulk is also aimed at draining the target system's resources, but uses the technique of sending a large number of requests with invalid headers, thus draining the system of its resources.

SSH-Bruteforce – an attack on the SSH protocol using brute force techniques. An attacker tries to gain unauthorized access to a system by brute-forcing different combinations of usernames and passwords.

DDOS attack-LOIC-UDP – DDoS attack using LOIC (Low Orbit Ion Cannon) tool and UDP protocol. LOIC performs the attack by sending large numbers of packets over the UDP protocol, which can overwhelm the target network or system.

4 Data precessing

A review of the previous studies revealed that, although most of the classical machine learning methods perform well on the F-measure metric, the best results remain with decision tree-based models [20]. Basic criteria for selection:

- The model must learn fast enough;
- The model must be easily interpretable;
- The model must not require a large amount of data preparation.

These criteria are appropriate for a decision tree. Also, to improve the detection probability of the model, it was decided to use an ensemble of trees rather than just trees.

The final model looks like this:

- Decision tree (in ensemble);
- Random forest (in ensemble);
- Logistic regression – processing the tree predictions and final output.

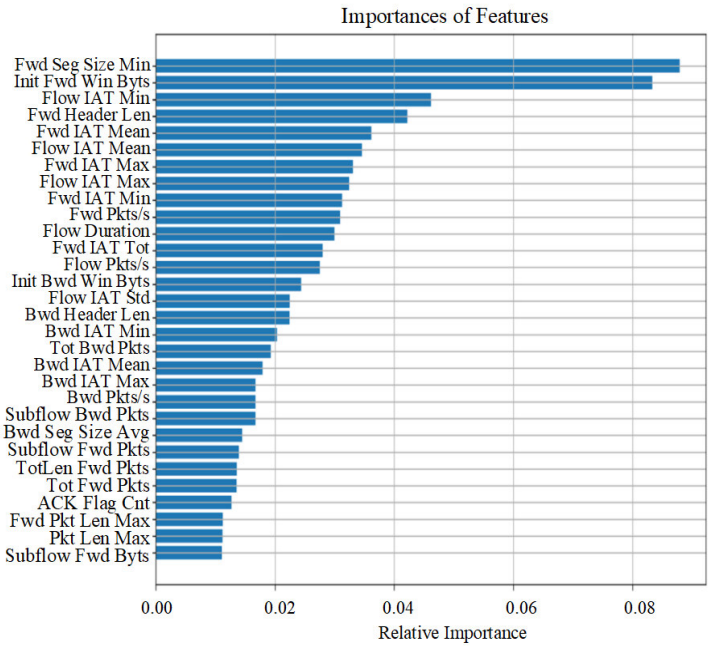

Fig. 1. Importance of Features.

The model pipeline consists of the following steps:

- Loading the dataset;
- Processing the dataset (removing empty values, duplicates, reducing the dimensionality);
- Cross-validation;
- Training the ensemble on a portion of the data;
- Testing the model on the rest of the data;
- Testing the model with feature selection through trees.

As a result of this feature selection method, we obtained the following distribution of the effect of the features on the prediction result for the RandomForest model (Figure 1)

A pre-perfomance characterization assessment in the form of the correlation matrix was constructed (Figure 2) to ensure that the model is not trained on dependent features.

Based on the results of the matrix, the features that correlate more than 0.9 (linear relationship) were discarded.

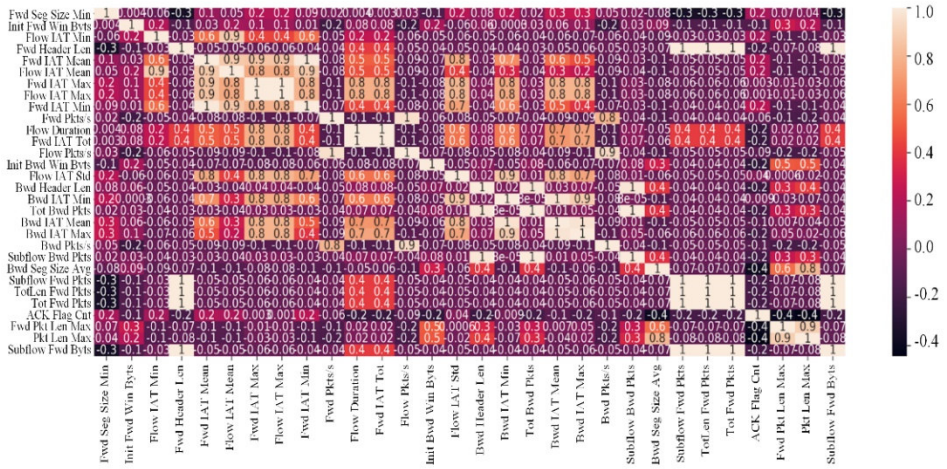

Fig. 2. Correlation Matrix.

In the end, a total of 14 features were selected that contribute the most to the final result: Fwd Seg Size Min; Init Fwd Win Byts; Flow IAT Min; Fwd Header Len; Fwd IAT Mean; Flow IAT Mean; Fwd IAT Max; Fwd Pkts/s; Flow Duration; Init Bwd Win Byts; Flow IAT Std; Bwd Header Len; Bwd IAT Min; Bwd Pkts/s.

The model was trained and tested on these features (on a synthetic dataset, and on real traffic). Learning outcomes in the form of classification effectiveness indicators (Precision, Recall, F1-score) are presented in Table 1.

Table 1. Classification effectiveness indicators.

	Precision	**Recall**	**F1-score**	**Support**
BENIGN	0.90	0.97	0.93	4059
DDOS attack-HOIC	1.00	1.00	1.00	2689
DoS attacks-Slowloris	1.00	1.00	1.00	2685
Bot	1.00	1.00	1.00	1833
DoS attacks-GoldenEye	1.00	1.00	1.00	1731
DoS attacks-Hulk	1.00	1.00	1.00	1805
Infilteration	0.93	0.75	0.83	1772

SSH-Bruteforce	1.00	1.00	1.00	1843
DDOS attack-LOIC-UDP	1.00	1.00	1.00	552
Accuracy			0.97	18969
Macro avg	0.98	0.97	0.97	18969
Weighted avg	0.97	0.97	0.97	18969

The analysis of the synthetic dataset showed that the feature values in the dataset were collected on low-bandwidth traffic, which is completely inconsistent with modern networks. Because of this, a part of the features that are used for prediction may be distorted, resulting in the inability to use the model trained on this dataset to detect anomalies in real traffic. In addition, this model cannot successfully detect denial-of-service attacks replicated at the L7 layer.

5 Suggested approach

At the heart of the proposed algorithmic support is a process that integrates network activity within the interaction of network traffic sniffer and web-oriented management interfaces of a cyber-physical system into a single dataset.

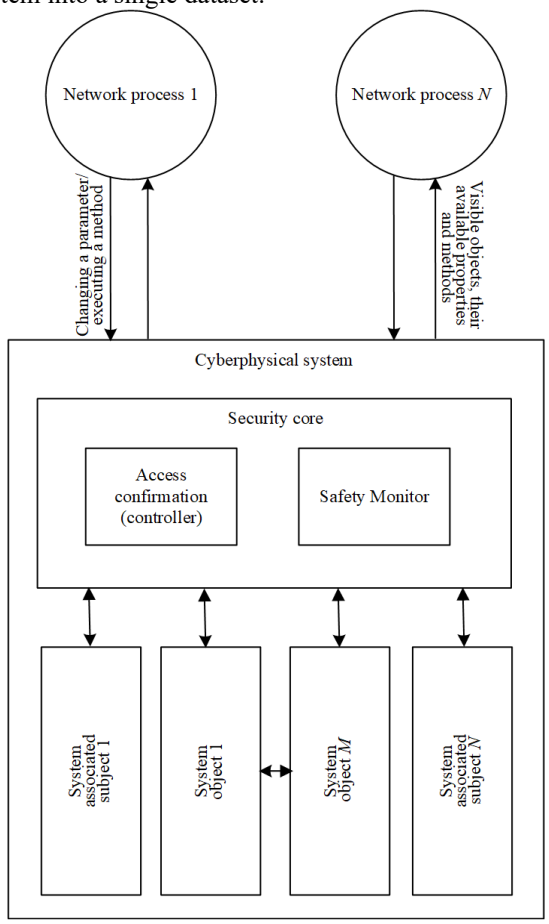

Fig. 3. Call monitor scheme.

In doing so, it is intended to provide:

1) collecting data and bringing it to a single normalised form;
2) grouping data according to certain features and attributes;
3) detecting incidents based on the detection of the correlation and alerting security personnel;
4) visualization of the processed data as a tool for analysis and investigation of incidents
5) generation of reports on the state of assets of the protected system. The approach used in the algorithm involves developing a call monitor (Figure 3) the main function of which is to perform parsing of all incoming requests as well as to perform the primary control of the fields whether the values are present in stop lists.

To analyse the data, a slice of traffic is supposed to be prepared for retrospective evaluation at some time interval. Figure 4 shows the conceptual scheme of the algorithmic software used.

The basic evaluated features can be the number of requests, the average time between requests, the standard deviation of time between requests, the proportion of errors with code 5XX in application responses to a given user, the proportion of errors with code 4XX in application responses to a given user, and the uniqueness of the resources requested by the user [9]. Based on the detection of sources of malicious requests, additional visual analytics were performed on various evaluation metrics for the web services of the cyber-physical system management interface.

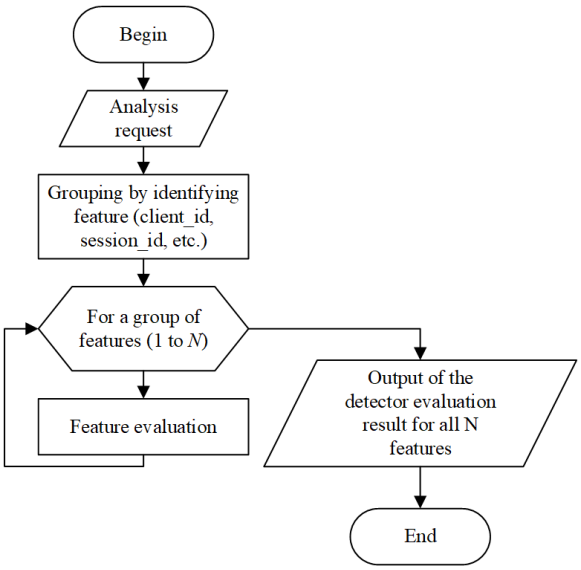

Fig. 4. Conceptual scheme of the proposed approach.

6 Experiment and result

The experiment was conducted on a virtual polygon emulating the operation of cyber-physical systems. Five scenarios (different techniques) of a controlled DDoS attack on predefined endpoints of all control interfaces were implemented. Table 2 shows the results of the experiment to evaluate the effectiveness of defense against attacks on the pool of cyber-physical system control servers in the following modes*:

1) practicing built-in security mechanisms like RateLimit;
2) detecting and blocking DDoS attacks using an information protection tool like an application-level firewall;
3) implementation of the proposed algorithm in addition to Method 2.

Table 2. Evaluating the effectiveness of detecting and blocking sources of malicious requests.

Security mode* No.	Metric	Accessibility violation attack scenario				
		1	2	3	4	5
1	FAR	0.043	0.023	0.095	0.052	0.020
	FRR	0.203	0.157	0.177	0.307	0.150
2	FAR	0.031	0.050	0.102	0.043	0.031
	FRR	0.173	0.237	0.221	0.114	0.331
3	FAR	0.030	0.022	0.097	0.041	0.125
	FRR	0.092	0.112	0.201	0.082	0.104

As an example of the benefits of the proposed approach, the diagram (Figure 5) shows the distribution of the number of subject requests by time, grouped using different colours according to the "URI referral path".

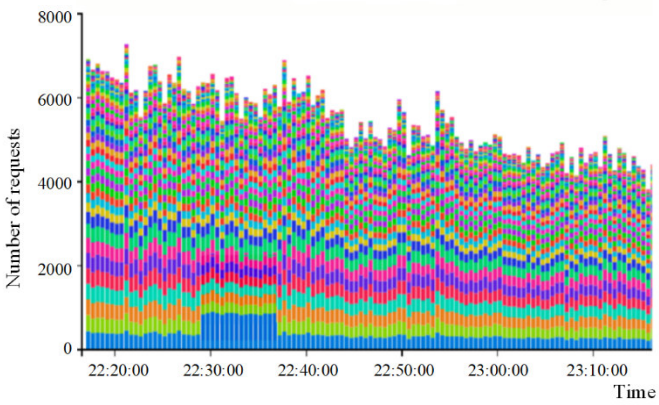

Fig. 5. Example of detecting a DDoS L7 attack on the management interface.

Thus, the peaks of requests to the cyber-physical system interface between 21:20-21:40 were erroneously detected by the embedded software protection as an attack due to the atypical time range of the access subjects. The software implementation of the proposed approach successfully detected this type of attack.

7 Conclusion

The study confirms the need to adapt the models tested on test datasets to real-world traffic and the specifics of the protection object. The proposed approach implies that the computer model of attack detection built at both the network and application levels should be further trained as the dataset expands, and tested on attacks implemented from different locations of the network infrastructure of the cyber-physical system.

Research on automated analysis of malicious requests in web-based services and rapid detection of their sources extends the theoretical base of methods for detecting potentially dangerous information flows. Detailed consideration of the problem allows us to model the protection means on the basis of the classification of incoming requests by means of applying data mining methods. The complex of theoretical and methodological developments obtained as a result of this study will form the basis for the formation of science-based principles to improve the system of countering attacks on web-oriented components of cyber-physical systems.

The article is based on the results of research carried out at the expense of budgetary funds on the state assignment of the Financial University.

References

1. M. Fraiwan, F. Al-Quran, B. Al-Duwairi, *Defense Analysis Against Store and Forward Distributed Reflective Denial of Service Attacks*, in 2018 International Conference on Innovations in Information Technology (IIT), 2018, Al Ain, United Arab Emirates (2018)
2. A.Y. Nur, M.E. Tozal, *Defending Cyber-Physical Systems against DoS Attacks*, in 2016 IEEE International Conference on Smart Computing (SMARTCOMP), 2016, St. Louis, MO, USA (2016)
3. F. Zahid, G. Funchal, V. Melo, M.M.Y. Kuo, P. Leitao, R. Sinha, *DDoS Attacks on Smart Manufacturing Systems: A Cross-Domain Taxonomy and Attack Vectors*, in 2022 IEEE 20th International Conference on Industrial Informatics (INDIN), 2022, Perth, Australia (2022)
4. N. Sun, M. Ding, J. Jiang, W. Xu, X. Mo, T. Yonghang, J. Zhang, IEEE Communications Surveys & Tutorials **25(3)** (2023)
5. M. Tehaam, S. Ahmad, H. Shahid, M.S. Saboor, A. Aziz, K. Munir, *A Review of DDoS Attack Detection and Prevention Mechanisms in Clouds*, in 2022 24th International Multitopic Conference (INMIC), 2022, Islamabad, Pakistan (2022)
6. A. Iskhakov, R. Meshcheryakov, S. Iskhakov, *Problems of Using Compromise Indicators for Proactive Threat Detection in Robotic Systems*, in 2021 14th International Conference Management of large-scale system development (MLSD), 2021, Moscow, Russian Federation (2021)
7. A. Iskhakova, R. Meshcheryakov, *Automatic search of the malicious messages in the internet of things systems on the example of an intelligent detection of the unnatural agents requests*, in 2017 Second Russia and Pacific Conference on Computer Technology and Applications (RPC), 2017, Vladivostok, Russia (2017)
8. M.D.T. Bennet, M.P.S. Bennet, D. Anitha, *Securing Smart City Networks - Intelligent Detection Of DDoS Cyber Attacks*, in 2022 5th International Conference on Contemporary Computing and Informatics (IC3I), 2022, Uttar Pradesh, India (2022)
9. F. Rebecchi, J. Boite, P. -A. Nardin, M. Bouet, V. Conan, *Traffic monitoring and DDoS detection using stateful SDN*, in 2017 IEEE Conference on Network Softwarization (NetSoft), 2017, Bologna, Italy (2017)
10. M.E. Şahin, S. Özdemir, *Detection of Malicious Requests on Web Logs Using Data Mining Techniques*, in 2019 4th International Conference on Computer Science and Engineering (UBMK), 2019, Samsun, Turkey (2019)
11. I. Ghafir, V. Prenosil, *DNS traffic analysis for malicious domains detection*, in 2015 2nd International Conference on Signal Processing and Integrated Networks (SPIN), 2015, Noida, India (2015)
12. M.R. Rahman, R. Mahdavi-Hezaveh, L. Williams, *A Literature Review on Mining Cyberthreat Intelligence from Unstructured Texts*, in 2020 International Conference on Data Mining Workshops (ICDMW), 2020, Sorrento, Italy (2020)
13. C. Li, L. Dai, Z. Xu, Y. Ding, Y. Han, *A Message-Based Malicious Detection Scheme of Public DNS Services*, in 2021 IEEE 23rd Int Conf on High Performance Computing & Communications; 7th Int Conf on Data Science & Systems; 19th Int Conf on Smart City;

7th Int Conf on Dependability in Sensor, Cloud & Big Data Systems & Application (HPCC/DSS/SmartCity/DependSys), 2021, Haikou, Hainan, China (2021)

14. R. Stoleriu, A. Puncioiu, I. Bica, *Cyber Attacks Detection Using Open Source ELK Stack*, in 2021 13th International Conference on Electronics, Computers and Artificial Intelligence (ECAI), 2021, Pitesti, Romania (2021)
15. C. Ma, A. Wu, W. Ma, K. Chen, Y. Liu, X. Liang, *Malicious URL Recognition Based on Multi-feature Fusion and Machine Learning*, in 2022 41st Chinese Control Conference (CCC), 2022, Hefei, China (2022)
16. C.S. Tejaswi, Y. Chaitanya, A. Jesudoss, P. Shyry, *Malicious Attacks Detection Using Machine Learning*, in 2022 4th International Conference on Inventive Research in Computing Applications (ICIRCA), 2022, Coimbatore, India (2022)
17. J. Wu, Z. Yang, L. Guo, Y. Li, W. Liu, "Convolutional Neural Network with Character Embeddings for Malicious Web Request Detection," 2019 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking (ISPA/BDCloud/SocialCom/SustainCom), Xiamen, China, 2019, pp. 622-627. https://www.doi.org/10.1109/ISPA-BDCloud-SustainCom-SocialCom48970.2019.00094.
18. A. Raza, S. Memon, M. A. Nizamani, M. Hussain Shah, *Machine Learning-Based Security Solutions for Critical Cyber-Physical Systems*, in 2022 10th International Symposium on Digital Forensics and Security (ISDFS), 2022, Istanbul, Turkey (2022)
19. R. Prabhla, S. Sankaran, *An Experimental Platform for Security of Cyber Physical Systems*, in 2019 IEEE International Symposium on Smart Electronic Systems (iSES) (Formerly iNiS), 2019, Rourkela, India (2019)
20. N. Gulia, K. Solanki, S. Dalal, *Comparative Analysis to Identify the Effective Machine Learning Method for Prediction of DDOS Attack*, in 2022 10th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), 2022, Noida, India (2022)