Intelligent Sensor of Information and Technical Impact (ITI) on the Network Subsystem of a Man-Made Facility

Yuriy Sosnovskiy [1], Victor Milyukov [1] and Veronika Ilyina [1]

[1] *V.I. Vernadsky Crimean Federal University, Vernadsky av.,4, Simferopol, 295007, Crimea*

Abstract

The aim of the article is to develop an information model of an intelligent sensor of information and technical impact (ITI) on the communication subsystems of microprocessor-based control systems used in technogenic objects that can be classified as the objects of critical information infrastructure.

The article substantiates the importance of identifying abnormal traffic in the communication subsystem of the lower and medium levels of the controlling information systems. Such traffic may indicate both serious errors in the MCS (microprocessor-based control system) itself, which were not detected during the system implementation and testing, and the fact of a computer attack or ITI on these levels of MCS.

The kernel of Modbus request-response packet is considered as a communication subsystem protocol, irrelevant of the communication environment type: RTU or TCP.

The information model in terms of extended Petri nets (EPN) developed in the paper allows us to describe in a formalized way the place of a smart sensor in the MCS structure and conditions of sensor triggering upon detection of a computer attack.

The software implementation of sensor using machine learning method – XGBoost – is performed, the algorithm of data preparation for training and cross-validation of the method is given. The results of testing the method on sets of traffic dumps with signs of a computer attack (CA) on MCS showed satisfactory performance of the method to identify the computer attack (CA). The results are presented in the paper.

Keywords

information and technical impact (ITI), microprocessor-based control systems (MCS), intelligent sensor, abnormal traffic sensor

1. Introduction

The increase in the number of systems operated by digital control systems, microprocessor-based control systems and information control systems (ICS), as well as the growing complexity of such systems leads to additional vulnerabilities in the software, which raises the likelihood of an intruder implementing ITI threats on them [1]. Most of the ICS consist of several sub-system, the main part of them is PLC. Another subsystems are Human Machine Interface (HMI), field-level communication subsystem, Master Terminal Unit (MTU) and Remote Terminal Unit (RTU) [2]. Due to the simple programming, variable control program and existing modules, high reliability and convenient expansion of PLC, many designers of the ICS have favored it [3].

Previously it was considered, that the industrial control system network is isolated from the external network, so that PLC is a safety device. Some virus attacks in recent years, such as most famous Stuxnet, have confirmed the erroneousness of this idea. However, an increasing number of ICS have an network and internet connection today [4].

AISMA-2021: International Workshop on Advanced in Information Security Management and Applications, October 1, 2021, Stavropol, Krasnoyarsk, Russia

EMAIL: sosnovskiy.yv@cfuv.ru (Yuriy Sosnovsky); milyukov.vv@cfuv.ru (Victor Milyukov); nika.ilyina@mail.ru (Veronika Ilyina)

ORCID: 0000-0003-3807-5297 (Yuriy Sosnovskiy); 0000-0002-0429-8540 (Victor Milyukov)); 0000-0003-4165-5620 (Veronika Ilyina)

There are lot of examples of different vulnerabilities with high grade CVSS score. For example, in the month of 2022 alone, there are more than 100 entries in the search engine vulners.com. The list of manufacturers is extensive, these are Siemens (CVE-2014-2251), Schneider (CVE-2012-0931), Omron (CVE-2015-0987), Rockwell (CVE-2012-4690), Mitsubishi (CVE-2015-3938), WAGO and many other. Most records have a CVSS value of more than 5, often more than 7. It has been shown by the search engine SHODAN[5] that thousands of industrial control systems are directly accessible via the Internet. Yet today there is no consideration of the information security in the phase of building, testing PLCs, which causes various vulnerabilities and makes PLC programs vulnerable to tampering attacks[6][7][8].

For the cyber security of inner PLC programs, Daniel et al. converted SCL language to the model NuSMV [9] and proposed a formal verification method of complex properties for PLC programs[10].

ICS protection is becoming a particularly urgent task. At the top level of the network, standard methods and tools for protecting information systems are used. Methods to ensure the security of the upper-level communication subsystem of the ICS are well-known and continue to be actively developed. However, the field level and the level of programmable logic controllers for control and monitoring systems require using special measures to control information processes and be protected against computer attacks and other impacts.

A deeper protection approach is also being implemented. For example, ViPNet SIES MC is responsible for managing the components of the ViPNet Security for Industrial and Embedded Solutions (SIES) solution. It allows deploying the solution in a trusted way, putting its components into operation and updating both the components themselves and their key information. The ViPNet SIES solution is an embedded security tool for elements of ACS, M2M and industrial Internet of Things (IIoT) systems (Whitepaper "Application of cryptographic protection of information in intelligent electricity metering systems" Official documentation InfoTeCS, can be found at https://infotecs.ru/product/vipnet-sies-mc.html).

In-depth analysis performed on the Siemens PLC environment, communication protocol S7CommPlus. This protocol enables communication between the engineering software from the vendor and PLCs like the S7–1211C [11]

The key element of MCS security system is a sensor (indicator hardware or software module), allowing to detect the fact of a computer incident, i.e. a successful implementation of ITI by an intruder.

Thus, the task of developing an intelligent sensor of abnormal traffic in the communication subsystem of MCS, using methods to ensure easy configuration of the sensor and little dependence on the types and formats of data transmitted in different systems, is relevant and of practical interest.

Problem statement

The following assumptions were made in this research:

- the creation, training and testing of the smart sensor requires appropriate traffic dumps, for generation of which a special simulation model is used, and the developed technique for test sequences (dumps) generation is implemented;
- the high level of consistency of the MCS communication subsystem simulation model is ensured by the fact that the software elements of the model can interact via a computer network with the actual components of the MCS, used in the model, without being fundamentally changed;
- Modbus TCP was selected as the network communication protocol, but for the intelligent sensor the TCP part of the data packet is discarded and the content relating directly to the Modbus protocol is used. It is assumed, that analysis of TCP-packages for their relevance to standard criteria (correctness of addresses, integrity, absence of spam, etc.) is made on a higher level of information system by tools, which become standard.

Thus, the aim of the work is to develop information support, as well as software and hardware support, for the intelligent sensor of ITI on MCS communication subsystems of man-made facilities, on the basis of machine learning techniques.

Taking into account the above-mentioned assumptions, to achieve the goal the following tasks are set:

- to develop the information support for the intelligent sensor of ITI on communication subsystems of MCS of man-made facilities;

- to select a method for identifying abnormal MCS traffic, choosing from a variety of machine learning techniques, and to write a software program for the sensor model implementation;
- to "train" the sensor with the help of special traffic dumps and test the sensor's output characteristics on the test dumps with rare "abnormal" events.

2. Information support for intelligent sensor of ITI on communication subsystems of man-made facilities MCSs

The basis for the development of the methodology of MCS protection assessment under ITI conditions is the model of MCS functioning under ITI conditions, which allows a comprehensive analysis of the interrelated processes of MCS functioning, implementation of ITI and elimination of their consequences.

The scheme of the MCS model functioning under ITI conditions and in terms of the extended Petri net is presented in [12][13]. The model of MCS functioning under the conditions of ITI and in terms of extended Petri nets (EPN) contains three operating circuits, as with the regular functioning, the circuit of simulation of ITI at MCS and the circuit of elimination of the consequences of ITI on MCS.

It is assumed that in the ITI implementation an intruder may exploit undeclared capabilities [14] in both the MCS hardware and software as well as in the programmable data network routers. Moreover, an intruder can be not only external, but also internal, who is familiar with specifics and time constraints of the technological process (triggering conditions of automatic and automated actuators), and is able to implement unknown impact, realizing "zero" day vulnerability [15].

3. Identification of abnormal MCS traffic based on machine learning methods and software sensor implementation

On the basis of preliminary research using dumps of normal and abnormal traffic, the most effective solution in terms of the combination of ROC AUC, Recall and Precision metrics was selected - an open-source library XGBoost, which provides a high-performance implementation of decision trees on gradient boosting.

Jupyter Notebook was used as the software environment and Python was chosen as the programming language.

The stage of data preparation for further processing is presented as a series of steps.

1. Data conversion. For convenience, the data are converted into tabular form pandas.DataFrame. The obtained data are merged into dataFrame, which contains results of normal operation and error handling. A markup is added to these, indicating the data corresponding to the error operation of MCS.

2. Checking for additional attributes. One of the most important characteristics that will correlate strongly with the target variable is whether the device id, its register or function code is "new", the one that was not present during normal operation of the system. To add these features, the appropriate code has been implemented, excerpts of which are shown in Figure 1.

```
:  data['new_register_address'] = 0
   data['new_count_registers'] = 0

   for i in range(len(data)):
       if data['register_address'][i] not in norm_vals_reg_address:
           data['new_register_address'][i] = 1
       if data['count_registers'][i] not in norm_count_registers:
           data['new_count_registers'][i] = 1
```

Figure 1: Implementation of additional attributes to check for new register values or number of registers

3. In case of insufficient amount of training data, abnormal data corresponding to additional binary features are generated. Due to the fact that it is resource-intensive to generate examples of "bad" data in the emulator, this process can be automated in the case of additional features, for example when only one device or register id value needs changing. This operation also increases the amount of data, which will have a positive impact on the results obtained from the model as there will be more examples to study.

4. Machine learning: building and training the model. XGBoost package was chosen as the model. The resulting dataframe is broken down into features and target variable. These sets will be used for both cross validation and delayed sampling training. The training of the model on the resulting data is shown in Figure 2.

```
xgb.fit(X_train, y_train)

XGBClassifier(base_score=0.5, booster=None, colsample_bylevel=1,
              colsample_bynode=1, colsample_bytree=1, gamma=0, gpu_id=-1,
              importance_type='gain', interaction_constraints=None,
              learning_rate=0.300000012, max_delta_step=0, max_depth=6,
              min_child_weight=1, missing=nan, monotone_constraints=None,
              n_estimators=100, n_jobs=0, num_parallel_tree=1,
              objective='binary:logistic', random_state=0, reg_alpha=0,
              reg_lambda=1, scale_pos_weight=1, subsample=1, tree_method=None,
              validate_parameters=False, verbosity=None)

y_predict = xgb.predict(X_test)
```

Figure 2: Training and prediction by means of XGBoost algorithm

The function fit() builds the composition (training) of the algorithm, the training sample and the labels of the target variable for each object of the training sample are passed inside the function as parameters. The function also displays the hyperparameters of the algorithm, the best combination of which can be chosen based on the current results. The predict() function takes a sample for which we want to make a prediction and returns an array of "predicted values" - the variable y predict. Results of the metrics and the result of building the error matrix are shown in Figure 3.

You can see from it that there occur errors of the first kind which are more undesirable than errors of the second kind because erroneous queries are missed. However, the relative number of missed queries is small.

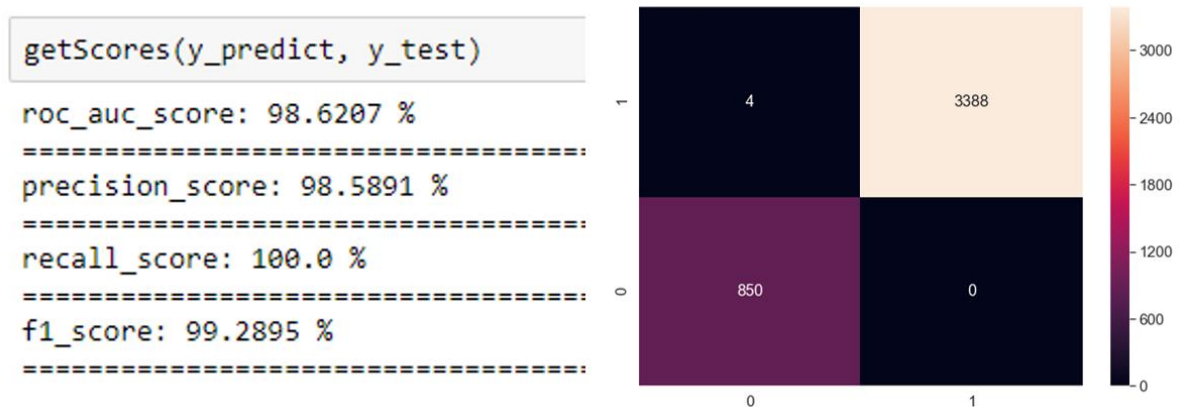

Figure 3: Metric values and Error matrix

To gain a complete picture it is also necessary to consider the results of cross-validation, which are: precision 100,00%, recall 99,94%, roc_auc 99,99%.

Based on the cross-validation results, it can be concluded that the results were satisfactory during the validation phase of the model functioning on training data.

Intelligent sensor testing

Test data is used to validate the developed and trained computer attack sensor model. These are MCS traffic dumps corresponding, in general, to the normal operation of the system, with some sporadic abnormal events. For example, wrong register numbers, register values beyond normal operation ranges, etc.

Files named dumps/dump2_testV1_TRM12_CTW and dumps/dump2_testV2_TRM12_CTW are used for this purpose. Steps 1, 2 of this algorithm are repeated for the test data. After that, metric values are calculated, which are shown in Figure 4 for the test data, and error matrix.

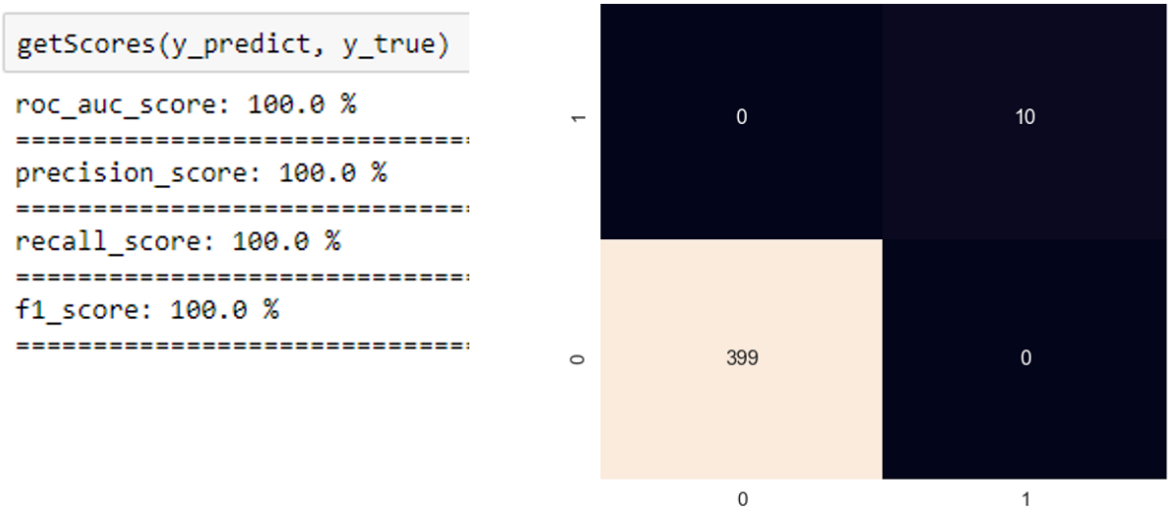

Figure 4: Metric values and Error matrix

The result of applying the method is positive; all 10 objects in the test data structure were found to be erroneous queries. The error matrix in Figure 4 also verifies the correctness of the results. It shows that the results are correct.

4. Conclusion

The proposed information and software intelligent sensor of ITI on the communication subsystem of MCS allows you to solve the important task of identifying abnormal events both at the field level of sensors and at programmable logic controller (PLC) level.

The scientific novelty of the proposed methodology lies in the development of the model of training sampling technique, the model of intelligent sensor and its software implementation based on machine learning method XGBoost.

The testing of software implementation of smart sensor model has shown high level of reliability in terms of anomalies identification in MCS traffic. However, there may be needed further research into the development of MCS specifications and requirements, so that they could provide effective training samples formation for different levels of MCS complexity.

5. References

[1] Smart Environments, Smart Systems, Smart Industries: Series of Reports (Green Book Series) within the framework of the Russian Federation Industrial and Technological Foresight Project / Author's Team; Centre for Strategic Research North-West Foundation. - St. Petersburg, 2012. - Vol. 4. - 62 c.

[2] Sandaruwan, G.P.H., Ranaweera, P.S. and Oleshchuk, V.A. (2013) PLC Security and Critical Infrastructure Protection. 2013 IEEE 8th International Conference on Industrial and Information

Systems, Peradeniya, 17-20 December 2013, 81-85. https://doi.org/10.1109/ICIInfS.2013.6731959
[3] Wang, Y., Liu, J.Y., Yang, C., Zhou, L., Li, S.F. and Xu, Z.Y. (2018) Access Control Attacks on PLC Vulnerabilities. Journal of Computer and Communications, 6, 311-325. https://doi.org/10.4236/jcc.2018.611028
[4] Wei Dong Application Analysis of PLC Technology in Electrical Automatic Control 2020 J. Phys.: Conf. Ser. 1533 022012
[5] Tianyou Chang et al Constructing PLC Binary Program Model for Detection Purposes 2018 J. Phys.: Conf. Ser. 1087 022022
[6] Mclaughlin S, Mcdaniel P. SABOT:specification-based payload generation for programmable logic controllers[C]// ACM Conference on Computer and Communications Security. ACM, 2012:439-449.
[7] J Klick, S Lau, D Marzin, J Malchow, V Roth. Internet-facing PLCs - A New Back Orifice[C].blackhat, 2015.
[8] R Spenneberg, M Brüggemann, H Schwartke.PLC-Blaster :A Worm Living Solely in the PLC[C].blackhat, 2016.
[9] Darvas D, Adiego B F, Viñuela E B. Transforming PLC programs into formal models for verification purposes [J]. 2013.
[10] Darvas D, Adiego B F, Vörös A, et al. Formal verification of complex properties on PLC programs[M]// Formal Techniques for Distributed Objects, Components, and Systems. 2016:284-299.
[11] Henry Hui, Kieran McLaughlin, Sakir Sezer. Vulnerability analysis of S7 PLCs: Manipulating the security mechanism. International Journal of Critical Infrastructure Protection. Volume 35, December 2021, 100470. https://doi.org/10.1016/j.ijcip.2021.100470
[12] Sosnovsky Y.V., Klimov S.M. Methodology of microprocessor control systems security assessment under information-technical impacts // Reliability No. 4 2018. Pp. 36-44.
[13] Sosnovsky Y.V., Klimov S.M., Milyukov V.V. The method of multiversion analysis of the security of the LSG from the effects of network attacks Proceedings of the 4th Central Research Institute of the Ministry of Defense of Russia, issue No. 150, volume 1, Part 1, Article No. 4, pp.21-26 Korolev 2019.
[14] Haolan Wu et al Research on Programmable Logic Controller Security 2019 IOP Conf. Ser.: Mater. Sci. Eng. 569 042031
[15] Klimov S.M., Astrakhov A.V., Sychev M.P. Methodological basis for counteraction to computer attacks. Electronic educational edition. - Moscow: Bauman Moscow State Technical University, 110 p., 2013.