Proposal of system for industrial network security with possibility of error predicition

Návrh systému zabezpečenia priemyselnej siete s možnosťou predikcie poruchových stavov

Igor Halenár, UIAM MTF STU

Abstract: The identification and prediction of faults in industrial systems and their communication networks is one of the challenges by implementation of an Industry 4.0 standards into practice. The article describes the possibility of analysis in industrial communication networks using modern technology for data processing. The proposal of presented algorithm is a solution which links data mining technologies with traditional approach to manage and secure data transfers in industrial networks.

Key words: industry, communication, safety, prediction

Abstrakt: Identifikácia a predikcia porúch v komunikačných sieťach priemyselných systémov predstavuje jednu z výziev pri implementácii štandardov Industry 4.0 do praxe. V článku sa popisuje možnosť analýzy komunikácie v priemyselných sieťach s využitím moderných technológií pre spracovanie dát. Návrh prezentovaného algoritmu predstavuje riešenie spojenia dolovania dát s klasickými technológiami na riadenie a zabezpečenie komunikácie dátových a priemyselných sietí..

Kľúčové slová: priemysel, komunikácia , bezpečnosť, predikcia

1. Introduction

Nowadays, modern industries go through a transformation in accordance with the standards of Industry 4.0. The Industry 4.0 standard represents mainly the German way of approach to implementing new technology in the manufacturing (7), otherwise known as Smart Factory (SF) (8), or Cyber Physical Systems (CPS) (9). New technologies, like Internet of Things (IoT) (10) represents the next generation of products and communication between them. The reliability of modern communication networks depends on many factors. The complexity of communication systems means many ways how to compromise data communication in network. A consequence of these facts is need to use a new ways of faults detection.

The article describes possibility of use data mining techniques in association with classic detection systems such as IDS and firewall. A part of article is a proposal of system that using listed techniques and description of data flows in proposed system. Reliable detection of faults by the transmission of data in communication networks is one of the basic prerequisites of a well functioning transmission infrastructure. As a communication error of the data network we can in principle consider all events that disturb one of the basic functions of communication systems. Between primary functions of communication network we can include technical reliability of data transmission. The quality of transmission and availability of communication is achieved mainly through full compliance with the standards in the design and implementation of communication routes. Yet, this function is in modern network design taken for granted. More important attribute of correct communication transfer is logical data security, which means a protection of the network against attacks and non-authorized access from outer (perhaps inner) side. Basic requirements of logically secure communication network including guaranteed access and authenticity, protection of data unambiguous authentication of elements in network environment (transfers, files, users, interfaces, etc.). In addition to these is very important continuous availability of all services in communication network, mostly in communication networks for critical applications of processes control.

2. Proposal of security and detection system

Protection of communication network is happening on several levels and is performed by various systems such as firewalls, IDS (Intrusion Detection System), IPS (Intrusion Prevention System) devices or protocols. In modern communication systems they are an equipment, which is commonly for integrity protection of data network. The efficiency of protection of communication channels depends on many factors, especially good combination of techniques and methodical systems update. The most widely used machine for protection of data networks is firewall, mainly stateful packet filter. More sophisticated protective systems are formed by more active and passive systems, not only firewall. There are used intrusion detection systems, intrusion preventing systems and alternatively systems using the honeypots technology.

The combination of listed security elements in communication network can be considered like a sufficient level of protection of data networks in normal "civil" computer networks. Clearly, this combination stops a high amount of attacks and security breaking attempts. In principle it is not possible to stop all security threats. There are well known techniques (5), which is impossible to capture by any currently available element of active safety.

The level of successfully detection of security bugs in network can be increased by joining the classical technology (firewall, IDS) with special methods of data processing. The combination of IDS and neural networks is able to detect non-standard errors and data transfers. Another way can be implementation of technology from data mining environment (1), which are using neural networks technology.

The communication in data network is, in this case, represented like some kind of data and is processed with the use of database engine, perhaps a data warehouse. Because the amount of data passing through the networks is too much for classical processing methods, this is the area for the application of methods of knowledge discovery in databases (KDD) and data mining (DM).

Process of knowledge discovery in databases is the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data (6). Possibility of error prediction is based on statistical and other data preprocessing (neural networks, six sigma techniques) used in KDD (11).

Methods and tools for protection of communication networks against attack and errors can be divided into two primary groups. The first group is the prevention. That means intrusion detection systems and there are two types of intrusion detection. The first is misuse-based detection (also known as knowledge based method). This method is based on database that contains a lot of signatures about known network attacks. IDS, for example SNORT (2), compare sniffed data with database content. The output is alert about detected abnormality. A problem with this method is that we cannot detect unknown attacks.

Second method is anomaly based detection. This method is based on premise that all anomalous activities are malicious and all the attacks are subset of anomaly activities. By building a model of the normal behavior of the system, then it looks for anomalous activities that do not conform to the established model.

The second group of tools (firewall) is focused on the detection of ongoing attacks or communication errors. The best solution is to join the firewall system with an intrusion detection system, where is possible to modify firewall rules using alerts from intrusion detection system.

This is the area for using data mining and neural network technologies. There exists a lot of works describing use of neural network like an enhancement of an ID system (4). Using data mining methods to enhance intrusion detection engines is very innovative approach. The drafted algorithm and probably deigned system consists of several objects. Objects are individual processes and communication between them provides own communication protocol.

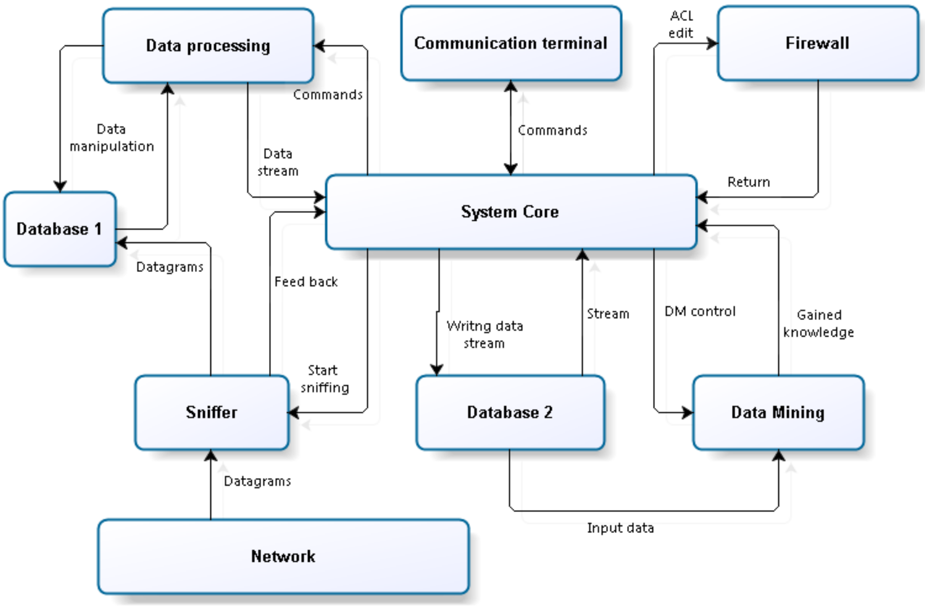

Fig. 1.: Architecture of proposed system

2.1. The principle of operation

The first entity of our proposed system is network data collection system. It can be some widely available suitable sniffer (for example wireshark, tcpdump, winpcap, msn,…). Next parts are modules for data preprocessing and database engine. The base of supposed system is module for data mining processing working on data gathered from network and preprocessed and stored in the database engine.

The recording of communication must be done in the central node in communication network, what means central switch or proxy. It is necessary to count on the fact that the entire record of communications data network represents a large volume of data and the processing-intensive computing power and capacity of disc space size.

Definitely helpful can be knowledge that the communication in classic ethernet network is based on TCP/IP communication model, and over 90% of communication is done thru TCP protocol (3).

Considering the preliminary analysis of possibilities and our previous work, we can propose that specification of data packets would be made at least by the following parameters:

- ID field in IP protocol
- System source port
- Destination port
- Source address
- Destination address
- ICMP type
- FLAGS setting
- TCP window size
- Length of transferred data

The packet is the basic information unit in the proposed system. It enters the system through the interface of communication monitoring (sniffer), which captures the packets and ends them for further processing.

After capturing the record of data communication into a database (DB1), it is necessary to extract the aforementioned values from the sum of data, eventually for better manipulating with data in KDD model, we have to make specific substitution. Regarding that the activity should be automated as much as possible, it is suitable to develop a software conversion bridge that would be able to supply suitable input data directly into the neural network used by KDD. It can be well done with some text manipulating language, such as GNU AWK for example.

Data modification should be executed in stages. Step one is to extract required values from the entire data package. Then we can afford to ignore the service communication records (ARP, RARP communication) as well as the irrelevant data. The following step lies in an appropriate representation of some elements.

Particularly, they are the following parameters (example):

Table 1.: Example of protocol parameters transformation – FLAGS transformation

FLAGS	Numerical substitution
NO_FLAGS	1
RST	2
FIN	3
PSH	4
URG	5
SYN	6

The same way of preprocessing have to be done to other data and protocols (TCP services, UDP services, ARP, ICMP, IGMP,…)

The substitution of protocol parameters is needed for automatic stream processing and to populate the database for data mining process. It is essential to carry out the whole data transformation process as simple and fast as possible.

3. Conclusion

The proposed solution shows possibility of using databases, KDD technologies and neural networks together with classical methods of computer security, such as firewall. The area of research is very complex and this article and methods represented in this article are just for gentle introduction into issues.

4. References

(1) Sharma, B., Gupta, H.: A Design and Implementation of Intrusion Detection System by Using Data Mining. In: Fourth International Conference on Communication Systems and Network Technologies, pp. 700-704. Bhopal, India. (2014)

(2) Open source intrusion prevention system. http://www.snort.org

(3) Naveen, N.C., Natarajan, S., Srinivasan.R.: Application of Change Point Outlier Detec-tion Methods in Real Time Intrusion Detection. In: Advanced Computer Science Appli-cations and Technologies (ACSAT), pp. 110-115. Kuala Lumpur (2012)

(4) Yu-Xin Ding, Min Xiao: Research and implementation on snort-based hybrid intrusion detection system. In: Machine Learning and Cybernetics, pp. 1414 – 1418. Baoding, China (2009)

(5) Gober, S.Z., Javed, B.; Saqib, N.A.: Covert channel detection: A survey based analysis. In: High Capacity Optical Networks and Enabling Technologies (HONET),9th Interna-tional Conference, pp. 57-65. Istanbul. (2012)

(6) Halenar, R.: Real Time ETL Improvement In: International Journal of Computer Theory and Engineering. - ISSN 1793-821X, n. 3 (2002), p. 405-409 [online].

(7) Lihui Wang, Martin Törngren , Mauro Onori, „Current status and advancement of cyber-physical systems in manufacturing." In Journal of manufacturing Systems 37, pp. 517-527, ISSN: 0278-6125, 2015

(8) Stephan Weyer, Mathias Schmitt, Moritz Ohmer, Dominic Gorecky: „Towards Industry 4.0 -Standardization as the crucial challenge for highly modular, multi - vendor productionsystems" in IFACPapersOnLine, Volume 48, Issue 3, pp 579-584, 2015

(9) Jay Lee, Hung-An Kao, Shanhu Yang: „Service Innovation and Smart analytics for Industry 4.0 and Big Data Environment" in Product services Systems and Value Creation. Proceedings of the 6th CIRP Conference on Industrial Product-Service Systems, pp. 3-8, 2014

(10) Yongrui Qin, Quan Z. Sheng, Nickolas J.G. Falkner, Schahram Dustdar, Hua Wang, Athanasios V. Vasilakos, „When things matter: A survey on data-centric internet of

things" in Journal of Network and Computer Applications vol. 64, pp. 137–153, ISSN: 1084-8045, 2016

(11) Andrej Trnka: Aplikácia dolovania dát do metodológie SIX SIGMA : (s využitím SPSS). Księży Młyn Dom Wydawniczy Michał Koliński. 117 p. ISBN: 978-83-7729-175-7. 2012

Author address:

Igor Halenár, Ing., PhD.
Slovak University of Technology
Faculty of Materials Science and Technology in Trnava
J. Bottu 25 , 917 01 Trnava
igor.halenar@stuba.sk