

A Survey of Data Quality in Industrial Networks and Enhancing Their Reliability Complex Indicator for the Needs of the Industry

Ekaterina Hristova

Faculty of Technical Sciences
Konstantin Preslavski University of
Shumen
Shumen, Bulgaria
ekaterinakonstantinova@abv.bg

Daniel Denev

Faculty of Technical Sciences
Konstantin Preslavski University of
Shumen
Shumen, Bulgaria
d.denev@shu.bg

Tsvetoslav Tsankov

Faculty of Technical Sciences
Konstantin Preslavski University of
Shumen
Shumen, Bulgaria
c.cankov@shu.bg

Abstract. With the evolution of the industrial sector, the establishment and administration of production units have emerged as primary concerns for each industrial enterprise. For the control process, emphasis is placed on programmable devices and industrial networks. Through them, manufacturing enterprises face two main questions, "What reliability could they provide?" and "Is it possible to improve the current devices they use." In this article, an experimental study related to the information level of an industrial network is carried out. It aims to increase the quality and speed of her workflow. Monitoring and data extraction of the communication of programmable devices is carried out as well. A method for improving data transmission in the same network is proposed. A statistical method is used to analyze the general reliability. Also the present article aims to develop a computational model of a complex indicator of the reliability of a system of programmable devices. The choice is based on the fact that many enterprises rely on the rapid processing of information and greater reliability of programmable devices.

Keywords: *Monitoring, Programmable logic controllers, Reliability, Risky technical system.*

I. INTRODUCTION

With the development of automation and industrial networks, many enterprises have come to rely on the fast and reliable transmission of data, as well as the general and field reliability of each machine. Based on this, a massive search for optimization solutions begins. A fundamental step towards achieving this goal is the statistical analysis of a given industrial system. Statistics is a mathematical method used to collect, systematize and interpret quantitative data, with an emphasis on examining the characteristics of the entire population by drawing inferences from a representative sample. The process of

statistical analysis involves collecting and analyzing in detail the vast amount of information needed to detect trends and gain valuable insights. In a professional context, statisticians process raw data and examine relationships between variables in order to uncover patterns and trends that provide important information to relevant interested parties. Statisticians working in a variety of sectors play a key role in generating new scientific discoveries, formulating business strategies and fostering innovation [1].

With the progress of time and technological processes, many new possibilities for improving network communication are discovered. Each new methodology can be viewed as a better version of its previous one. One of the latest advances in the advancement of network communication is the implementation of intelligent devices such as switches and routers. While years ago, most devices performed simple tasks, today, thanks to new technologies, they perform much more complex functions. These devices retain their functionality, providing opportunities for use by both scientists and ordinary users. Scientists and engineers use them to research and improve work processes. The present study drew attention to the early work of other researchers. Ivanka Georgieva, author of Automation Systems with Programmable Logic Controllers, explains the best and most efficient way to work with programmable devices. She shares her experiences on preparing simulation software so that the researcher or the average user can get the most out of any automation system. Another researcher, Ivan Popchev, examines risk in network communication and its reliability. Evgeni Gindev from the G.S. Rakovski Military Academy in Sofia also focuses on general and field reliability of risky technical systems. The following studies follow the research of Nikolay

Print ISSN 1691-5402

Online ISSN 2256-070X

<https://doi.org/10.17770/etr2024vol2.8036>

© 2024 Ekaterina Hristova, Daniel Denev, Tsvetoslav Tsankov. Published by Rezekne Academy of Technologies.
This is an open access article under the [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

Petrov also, who examines the reliability indicators of various systems.

II. MATERIALS AND METHODS

Each specific statistical study has for its object a certain extensive phenomenon that manifests itself in certain time and space frames. The process involves three interrelated steps: statistical observation, statistical clustering, and statistical analysis.

- The first stage, statistical observation, focuses on gathering initial empirical information as part of the overall study;
- In the second stage with the use of statistical clustering, the primary data must be summarized and systematized so that they are prepared for the application of statistical analysis methods;
- The third and final stage includes the statistical analysis itself, which must meet the goals and expected results that preceded the conduct of the overall statistical study.

Much scientific research is concerned with comparing the distribution of two or more variables. The peculiarity of these comparisons is that the conclusions drawn must refer to entire populations, and the data available to the researcher cover only a sample of it. Therefore, assumptions - hypotheses are initially formulated, and subsequently a check is made to see if the data from the sample confirms or rejects them.

- Zero (H0) - claims that there is no statistically significant difference in the compared statistical indicators. Although some variation may be observed in the samples, it is random and cannot be generalized to the general populations;
- Alternative (H1) - claims that the observed difference in the compared statistical indicators in the samples is statistically reliable and can be generalized for the general populations.

The degree of certainty with which the alternative hypothesis is accepted as true is called the guarantee probability (P). The risk of making a mistake by accepting the alternative hypothesis as true is called the significance level (α). After forming the hypotheses, an appropriate statistic (criterion) is selected, which is calculated according to the parameters detailed in the hypotheses. A final decision is made by comparing the tabular (theoretical) value of the criterion with the empirical (calculated from sample data).

To conduct a test procedure and for the purposes of the research in three stages, a test setup is organized with a control and test group of industrial devices and a switch playing the role of a key test element. The test group setup is shown in Fig. 3. Fig. 1 shows the approach for organizing the three stages of the test set-up. Four types of industrial controllers (Toshiba V200, Omron NX, Allen Bradley SLC-500, Siemens S7-1500) were selected for the purposes of the study. Each current test setup includes five industrial controllers of the given type applied to the respective stage. In addition, a Cisco Nexus 9300 smart switch is added for the test setup.

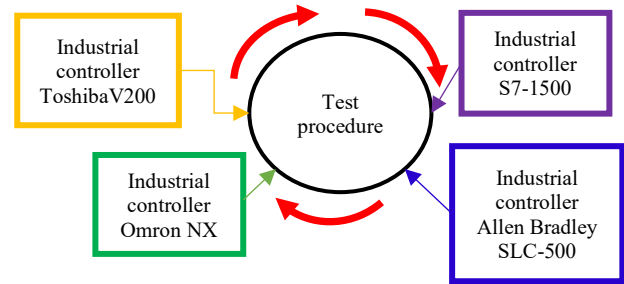


Fig. 1. Organizing the test setup.

Using activity diagrams in Fig. 2 an algorithm for conducting a test valid for each stage is presented.

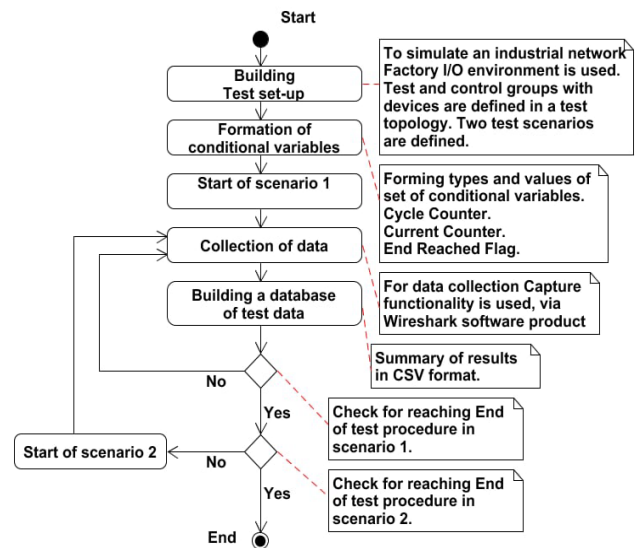


Fig. 2. Algorithm for conducting a test from a given stage.

In both types of test setups, a workstation with the following purpose is used:

- Management of test applications;
- Storage of intermediate data;
- Storage of test results.

For the purposes of the study, the following applications were selected:

- Factory I/O – integrated environment for network simulations;
- Wireshark – an application for analyzing network protocols and network data;
- Rstudio – an integrated environment for the synthesis of scripts intended for analyzing data.

To form an empirical set of results, the following set of criteria is established:

1. Port number;
2. Number of packages;
3. Packet size in bytes;
4. Total number of packages;
5. Number of sent packages;
6. Number of packets received.

According to these criteria, sets of tables 1, 2, 3 with the empirical data were formed.

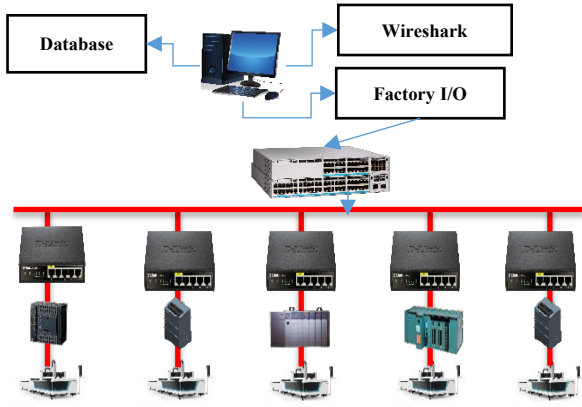


Fig. 3. Industrial network diagram with test set of devices.

TABLE 1 PROFINET NUMERICAL DATA FOR CONTROLLER OMRON NX AFTER IMPROVEMENT

Addresses	Packets	Bytes	Total Packets	Percent Filtered
Omron NX	40	8848	40	100.00%
Omron NX	30	455	30	100.00%
Omron NX	39	744	43	92.00%
Omron NX	55	311	55	100.00%
Omron NX	10	350	10	100.00%
Omron NX	15	886	15	100.00%
Omron NX	5	578	5	100.00%

TABLE 2 PROFINET NUMERICAL DATA FOR CONTROLLER ALLEN BRADLEY SLC 500 AFTER IMPROVEMENT

Addresses	Packets	Bytes	Total Packets	Percent Filtered
AB SLC 500	80	5961	80	100.0%
AB SLC 500	500	18966	500	100.0%
AB SLC 500	92	400	101	91.0%
AB SLC 500	20	1566	20	100.0%
AB SLC 500	40	2610	40	100.0%
AB SLC 500	500	1515	560	88.0%
AB SLC 500	29	451	32	90.0%

III. RESULTS AND DISCUSSION

On the basis of the theoretical formulation, the statistical study was carried out to test hypotheses and prove the results of the improvement of the industrial network. Graphical representation of information and data using visual elements such as charts, graphs, maps, and visualization tools provide an accessible way to see and understand trends, deviations, and patterns in data. It provides an excellent way to present data to non-technical audiences without confusion. For the purpose of the study, the data from the aforementioned controllers are presented in scatter plots. Such observations are called paired or repeated, i.e. to each x_i corresponds y_i . [2], [9], [10]

TABLE 3 PROFINET NUMERICAL DATA FOR CONTROLLER S7-1500 AFTER IMPROVEMENT

Addresses	Packets	Bytes	Total Packets	Percent Filtered
S7-1500	100	8410	100	100.0%
S7-1500	445	22886	500	89.0%
S7-1500	186	714	200	93.0%
S7-1500	45	2285	50	91.0%
S7-1500	25	8365	30	86.0%
S7-1500	285	1410	300	95.0%
S7-1500	36	366	40	92.0%

The figures show a distinct difference in the measurements before and after the additional optimization level was applied. But even so, no conclusion can be drawn about the exact magnitude of the effect of the Cisco Business 220 Smart Switch on the controllers in the network.

- Omron NX – (Fig. 4, Fig. 5)

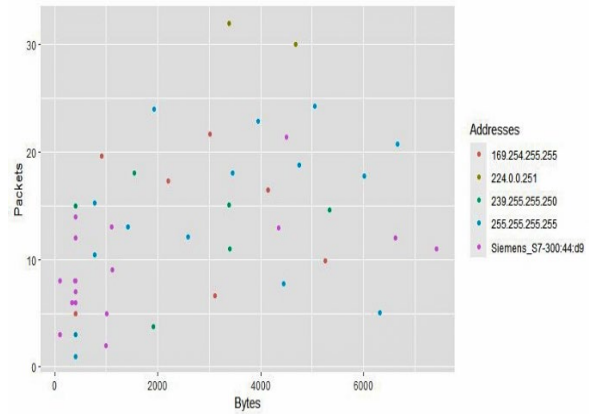


Fig. 4. Dot plot of an Omron NX controller.

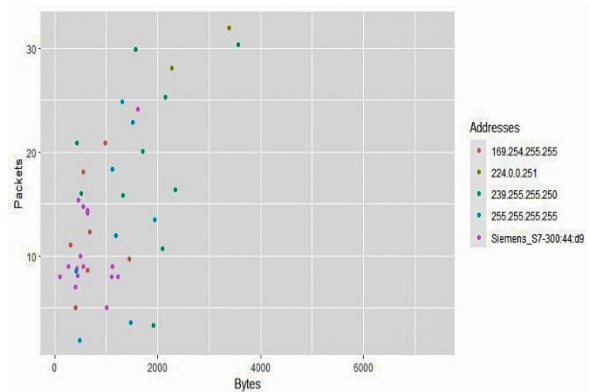


Fig. 5. Dot plot of an Omron NX controller after improvement.

- S7-1500 – (Fig. 6, Fig. 7)

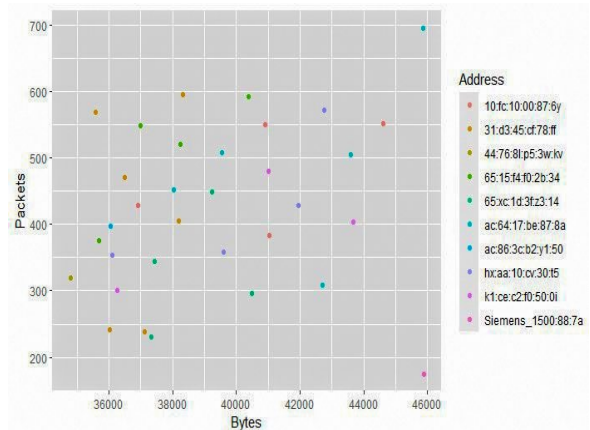


Fig. 6. Dot plot of an S7-1500 controller.

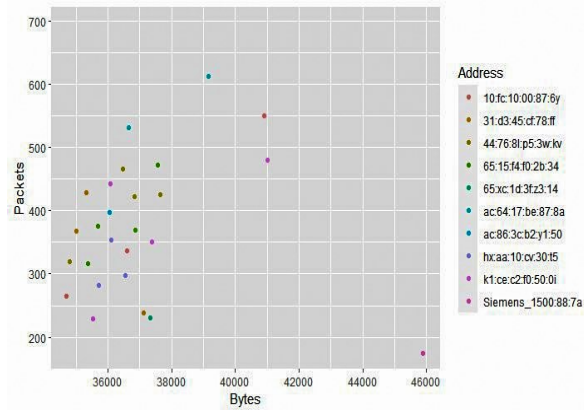


Fig. 7. Dot plot of an S7-1500 controller after improvement.

Despite the inability to estimate the exact magnitude of the switch's impact from the dot plots, the numerical values in the samples indicate a possible improvement in the number of packets and bytes transmitted between controllers in the industrial network.

- $H_0: m_y - m_x = 0$, i.e. no significant difference between data downloaded before adding the switch;
- $H_1: m_y > m_x$, i.e. there is a significant improvement in network reliability, packet transmission, and system performance optimization after connecting additional devices;

Statistics are used:

$$U_x = \sum_{i=1}^{n_x} \sum_{j=1}^{n_y} \delta_{ij} \quad (1)$$

where

$$\delta_{ij} = \begin{cases} 1, & x_i > y_j \\ \frac{1}{2}, & x_i = y_j \\ 0, & x_i < y_j \end{cases} \quad (2)$$

When testing H_0 , the confidence region for testing the hypothesis W takes the form

$$W = \{U_{1-\alpha} \leq U_x\}, P(W) = \alpha \quad (3), (4)$$

In practice, it is most often worked with $\alpha = 0,05$ which show the huge significance after optimization.

- Enhanced data density measurement methods

```
library(readxl)
library(fitdistrplus)
Profinet_before_raw <-
read_excel("Statistics/Profinet_OmronNX
_before.xlsx",
range = "C2:C29")
Profinet_after_raw <- read_excel("Statistics/
Profinet_OmronNX_after.xlsx",
range = "C2:C29")
as.numeric(unlist(Profinet_before_raw))
as.numeric(unlist(Profinet_after_raw))
Diff = Profinet_after - Profinet_before
plotdist(Diff, histo = TRUE, demp = TRUE)
```

```
library(readxl)
library(fitdistrplus)
Profinet_before_raw<- read_excel("Statistics/
Profinet_S7-1500_before.xlsx",
range = "C2:C29")
Profinet_after_raw <- read_excel("Statistics/
Profinet_S7-1500_after.xlsx",
range = "C2:C29")
as.numeric(unlist(Profinet_before_raw))
as.numeric(unlist(Profinet_after_raw))
Diff = Profinet_after - Profinet_before
plotdist(Diff, histo = TRUE, demp = TRUE)
```

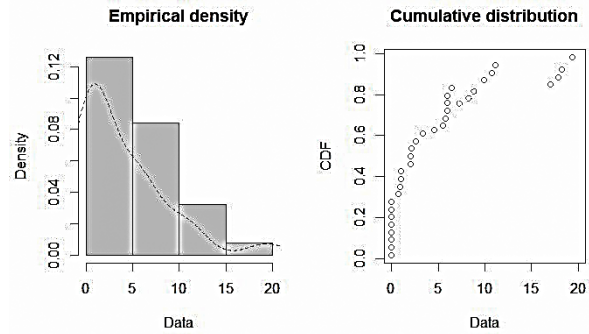


Fig. 8. Empirical density and cumulative frequency function plot of an Omron NX controller.

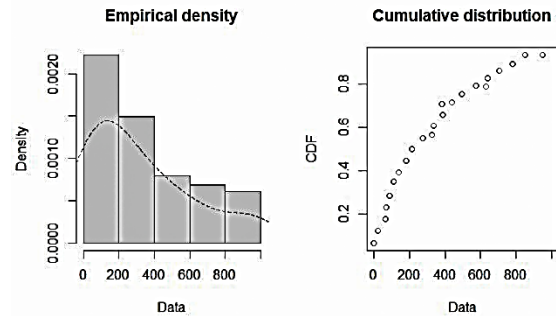


Fig. 9. Empirical density and cumulative frequency function plot of an S7-1500 controller.

- Wilcoxon test on the two-sample data for an Allen Bradley SLC-500 controller

```
median(Profinet_before_Allen Bradley SLC-500)
median(Profinet_after_Allen Bradley SLC-500)

wilcox.test(Profinet_before, Profinet_after,
paired = TRUE, exact=FALSE)
```

Wilcoxon signed rank test with continuity correction

alternative hypothesis: true location shift is not equal to 0

p-value (0,01415) < α (0,05)

Due to the existence of a significant difference between p and α , the hypothesis H_0 is rejected and H_1 is accepted as true.

- Improvement Effect Size Investigation on a Toshiba V200 Controller

```
wilcox.test(Profinet_before_Toshiba-V200,
Profinet_after_Toshiba-V200, paired = TRUE,
exact=FALSE)
Zstat_3<-qnorm(test_3$p.value/2)

#Effect size using
abs(Zstat_3)/sqrt(10)
```

```
##[1] 0.7858644
d = 0,79 => large
```

Effect size: very small - 0.01, small - 0.20, medium - 0.50, large - 0.80, very large - 1.20, huge - 2.0.

Cisco Nexus 9300 intelligent switch shows a significant impact on the operation of network controllers by optimizing their operation, improving the reliability of transmitted information and electronic products, and improving the speed of packet transmission between devices.

Fig. 10 observes the magnitude of the effect on all involved programmable controllers in the industrial network.

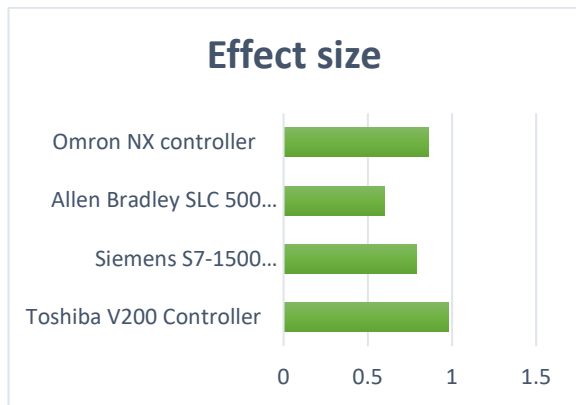


Fig. 10. Magnitude of the effect on all involved PLC.

A complex indicator of the reliability of risky technical systems (in particular, programmable devices and equipment) is the probability of normal functioning $P_{H\Phi}(\tau, t)$, expressing the probability that at any moment τ the product is operational and will continue to work without failure for a certain interval from time t , i.e. in the interval $\tau, \tau + t$. [3], [7], [8]

This general indicator of reliability for programmable devices in the industry can be defined as "the probability that the planned task (production of a product or processing of information) will not fail due to the fault of the electronic component composition". It is determined by:

$$P_{H\Phi}(\tau, t) = K_{\Gamma}(\tau) \cdot P_{BP}(\tau, \tau + t) \quad (5)$$

where: $K_{\Gamma}(\tau)$ - the readiness ratio of risky technical systems at a moment τ ;

$P_{BP}(\tau, \tau + t)$ – probability of failure-free operation of risky technical systems in an interval $(\tau, \tau + t)$.

With an established SR process (stationary random process), when no failure has occurred in the interval $(\tau, \tau + t)$ of $P_{H\Phi}(\tau, t)$ it is called the coefficient of operational readiness. $K_{O\Gamma}(t)$ is defined by the expression:

$$K_{O\Gamma}(t) = K_{\Gamma}(t) \cdot P_{BP}(t) \quad (6)$$

The operational readiness coefficients $K_{O\Gamma}(t)$ and the readiness coefficient $K_{\Gamma}(t)$ are provided in the standards as standardized complex indicators. Which of the two should be chosen depends on the way the respective product functions. [4]

For products with high reliability $K_{\Gamma}(t)$ is suitable, because the efficiency of their use will depend most on the recovery time.

For products where reliability is lower and critical, $K_{O\Gamma}(t)$ will be suitable.

Example: Through operational observations of a group of the same type of recoverable products, it was found that a total of $n = 5$ failures were obtained (table 4), and the total working time between failures is $\sum_{i=1}^n t_{0i} = 450 h$, the total recovery time is $\sum_{i=1}^n t_{Bi} = 22,5 h$ and the total stay in scheduled maintenance and repairs in the observed interval is $\sum_{i=1}^j t_j = 92,5 h$. It is assumed that there is an established normal operation process where the event flows (failures, recoveries) are stationary. To determine the reliability indicators at an operating time $t_p = 30 h$ and a recovery time $t_B = 2 h$.

TABLE 4 NUMERICAL RELIABILITY INDICATORS

Parameter	Values
Number of failures	5
Total working time between failures	450 h
Total recovery time	22,5 h
Total stay in scheduled repairs	92,5h
Working time	30 h
Recovery time	2 h

Solution: For the fail-safe metrics, we get:

$$n \cdot h_{daily}, 5 \cdot 3h_{daily} = 15 h_{daily} \quad (7)$$

$$30_{day} \cdot 15h_{daily} = 450 h \quad (8)$$

$$\bar{T}_0 = \frac{\sum_{i=1}^n t_{0i}}{n} = \frac{\sum_{i=1}^5 t_{0i}}{5} = \frac{450}{5} = 90 h \quad (9)$$

$$\bar{\omega} = \frac{1}{\bar{T}_0} = \frac{1}{90} = 11 \cdot 10^{-3} h^{-1} \quad (10)$$

$$P_{BP}(t) = \exp(-\bar{\omega} \cdot t_p) \quad (11)$$

$$P_{BP}(t) = \exp(-11 \cdot 10^{-3} \cdot 30) = 0,71892 \approx 0,719 \quad (12)$$

For the reparability indicators, it is obtained:

$$\bar{T}_B = \frac{\sum_{i=1}^n t_{Bi}}{n} = \frac{\sum_{i=1}^5 2h}{5} = \frac{22,5}{5} = 4,5 h \quad (13)$$

$$\bar{\mu} = \frac{1}{\bar{T}_B} = \frac{1}{4,5} = 0,22h^{-1} \quad (14)$$

$$P_B(t_B) = 1 - \exp(-\bar{\mu} \cdot t_B) \quad (15)$$

$$P_B(t_B) = 1 - \exp(-0,22 \cdot 2) = 0,35882 \quad (16)$$

For the complex reliability indicators, it is obtained:

$$K_{\Gamma\Pi} = \frac{\sum_{i=1}^n t_{oi}}{\sum_{i=1}^n t_{oi} + \sum_{i=1}^n t_{Bi} + \sum_{j=1}^k t_j} = \frac{450}{450+22,5+92,5} = \frac{450}{565} = 0,79646 \quad (17)$$

$$K_{\Gamma} = \frac{\sum_{i=1}^n t_{oi}}{\sum_{i=1}^n t_{oi} + \sum_{i=1}^n t_{Bi}} = \frac{450}{450+22,5} = \frac{450}{472,5} = 0,95238 \quad (18)$$

$$K_{O\Gamma}(t) = K_{\Gamma} \cdot P_{BP}(t) = 0,952 \cdot 0,719 = 0,68469 \quad (19)$$

In the theory of reliability, models have been developed for estimating the probability function for normal operation $P_{H\Phi}(\tau, t)$ and for the cases when one or more failures may occur and be removed in the process of operation. For the electronic component composition in the systems of programmable devices, these models are applicable when the random time T_B for failure removal does not exceed the definitely permissible time $t_{B,доп}$ for the controller to stay in a faulty state, i.e. $T_B \leq t_{B,доп}$. Such a model applicable to programmable devices is presented in [5]. In it, it is assumed that the total time to restore operability in the case of the n number of failures of the assembly line is much smaller than the considered operation time of the controller in the mode of information processing and task execution t_{BD} , i.e. $\sum_{i=1}^n t_{Bi} \ll t_{BD}$.

When determining the probability of normal operation $P_{H\Phi}(\tau, t)$ it is assumed that for the time t_{BD} no failure should occur. If it is assumed that in an established process of operation of the programmable devices, after the moment τ in the interval $(\tau, \tau + t)$, a failure has occurred, which will be removed by the maintenance personnel before the expiration of the time t_{BD} . For the calculation of the probability of normal functioning the following equation is suggested:

$$P_{H\Phi}(t, t_{BD}) = K_{\Gamma} \cdot P_{BP}(t) \cdot \{1 + [1 - P_{BP}(t)] \cdot P_B(t_{BD})\} \quad (20)$$

Example: With an established process of operation of a given series of programmable devices (Omron NX controller – Fig. 11), it is known that the readiness factor is $K_{\Gamma} = 0,99$; the probability of failure-free operation for a certain time for carrying out an operation t_{BD} is $P_B(t_{BD}) = 0,8$, and the probability of restoring the operability of the controller in the production enterprise t_{BD} is $P_B(t_{BD}) = 0,96$. To determine the probability of

normal operation of $P_{H\Phi}(t, t_{BD})$ of a programmable device of this series, if the appearance and removal during time t_{BD} allows only one failure.



Fig. 11. Programmable controller Omron NX [1].

Solution: After substitution in (20) we get:

$$P_{H\Phi}(t, t_{BD}) = K_{\Gamma} \cdot P_{BP}(t) \cdot \{1 + [1 - P_{BP}(t)] \cdot P_B(t_{BD})\} = 0,99 \cdot 0,8 \cdot [1 + (1 - 0,8) \cdot 0,96] = 0,944 \quad (21)$$

If no failure is allowed to occur in the same period, the probability of normal operation $P_{H\Phi}(t, t_{BD})$ (in this case, it is the coefficient of operational readiness $K_{O\Gamma}$) will be:

$$P_{H\Phi}(t, t_{BD}) = K_{O\Gamma} = 0,99 \cdot 0,8 = 0,792 \quad (22)$$

This academic publication presents an evolutionary study aimed at improving industrial networks and creating a complex indicator of their reliability. It examines smart switches and new software solutions, making the next-generation improvement accordingly. The research covers a variety of new techniques that have not yet received sufficient attention in the academic literature. The improvement of industrial networks using smart switches as a potential research area has been less explored. Previous research and improvement approaches are often based on techniques from previous generations. Furthermore, the present study confirms and complements previous qualitative results achieved by other scholars in the field. Also noted are gaps in the development of reliable communication, as well as old methods and approaches that are not optimal. As time progresses, the improvement of network communication becomes more and more important, especially due to the growing need for more reliable data transmission. This kind of network communication enhancement not only supports more reliable data transmission in industrial networks, but is also applicable in classical communication networks. This scientific paper demonstrates that the model presented in it has significant differences compared to all other models.

IV. CONCLUSIONS

From the proposed article, it can be concluded that by combining knowledge, skills and technical means from several scientific and applied fields, the goal is achieved, namely - improvement of the information level in industrial communication. It provides better reliability as well as information protection with less data loss. Smart switches show a significant impact on programmable devices and provide a wide range of opportunities to improve industrial communication. Based on the computational model, it can be summarized and

concluded that an efficient method is presented for industrial enterprises to calculate their reliability index, according to the number of their programmable devices and their operating time. After shaping the final solutions, the manufacturing companies get a general idea of the devices they own and their field reliability. Future development of a software plug-in for the Wireshark program is planned, through which a direct connection to the simulation program will be made. Through it, in a simulation process, industrial networks will be analyzed in real time, which in turn will facilitate other future developments and improvements. A future development is also planned in which the complex metric algorithm will be shortened and offer a more convenient calculation.

V. ACKNOWLEDGEMENTS

This current scientific paper and the research conducted would not have been possible without the wonderful support of Assoc. Prof. Eng. Valentin Atanasov. His commitment, expertise and precision in work were a source of inspiration and encouraged the progress of our business in the right direction. We also express our gratitude to the dean of our faculty, Prof. Sabin Ivanov, who secured the funding of the research through the acquisition of a smart switch.

REFERENCES

- [1] E. Gindev, An Introduction to the Theory and Practice of Reliability, Part 1. Basics of applied reliability, AI, Prof. Marin Drinov", Sofia, 2000, ISBN 954-430-608-0.
- [2] I. Popchev, I. Radeva, The new paradigm and the risk in the relationship "human – digital" environment, BAS Magazine, No. 5, Sofia, 2019, pp. 72-77.
- [3] I. Popchev, Six topics and literature on risk management, NBU, Bulgaria, Sofia, 2012, pp. 50-89.
- [4] I. Georgieva, Integrated Automation Systems with Programmable Logic Controllers, Bulgaria, Blagoevgrad, 2016, ISBN 954-8118-06-8.
- [5] N. Petrov, Reliability as a Technical-Economic Problem in Cybernetization of Society, Dissertation for the award of the National Academy of Sciences, Doctor of Economic Sciences, VSU, Chernorizets Hrabar, 2015, ISBN 978-954-92960-8-2,400.
- [6] N. Petrov, Reliability and risk as a global cyber problem, Studio RB ,”G. Rakovski”, 2019, ISBN 978-954-90476-0-8, pp. 45-55.
- [7] N. Petrov, V. Dimitrov and V. Dimitrova, Reliability of Technology Systems in Industrial Manufacturing, Monograph, "AkiNik Publications", New Delhi, India, 2019, ISBN: 978-93-87072-59-6.
- [8] N. Petrov, D. Baskanbayeva, S. Satybaldi, On exponentially large-scale events and the risk of their occurrence, Magazine "Science, Education, Intellect", No. 15, 2021, ISSN 2603-476X, pp. 61-77.
- [9] St. Stanev, St. Zhelezov, Computer and network security, Shumen, 2005, ISBN 954-577-306-5.
- [10] I. Tsonev, S. Stanev, Computer networks and communications, Shumen, 2008, ISBN 978-954-577-420-1.