Приложение Б

Тексты публикаций

Performance Evaluation of the Message Queue Protocols to Transfer Binary JSON in a Distributed CNC System

Maxim Ya. Afanasev, Yuri V. Fedosov, Anastasiya A. Krylova, Sergey A. Shorokhov
School of Computer Technologies and Control, Faculty of Control Systems and Robotics,
Instrumentation Technologies Department, ITMO University
49 Kronverksky Pr., St. Petersburg, 197101, Russia
Email: amax@niuitmo.ru, yf01@yandex.ru, {ananasn94, stratumxspb}@gmail.com

***Abstract*—This paper presents a performance study of the message queue protocols that are used for binary encoded JSON data transmission in a distributed network, which is the foundation of the CNC system. The distributed architecture of the control system makes it possible to simplify the integration of automated and robotic equipment into the cyber-physical production system (CPPS) by creating an open program interface for the interaction between the components of a distributed CNC system. For effective interaction, a unified protocol is needed that combines the high performance of TCP sockets with the possibility of many-to-many routing between endpoints. A message queue was selected as the protocol; the main format of the transmitted data is JSON. A comparative analysis of the bandwidth of TCP and UNIX sockets was performed; the average bandwidth of various network connections used in the distributed CNC system was determined. The maximum compression ratio for various JSON packers was calculated. The dependence of the compression ratio and the packing/unpacking time on the size of the transmitted data is obtained. The four message queue protocols ware compared. Based on the analysis, the most productive bundle of the data transmission pattern and the package/unpacking algorithm was determined. A technique for determining the total bandwidth of the message queue and network overhead for various network scenarios has been developed.**

I. Introduction

Considering the development trends of modern production it should be noted that new organization forms of interconnection between technological equipment components are gradually phasing out the older ones. Increasingly, enterprises create unified equipment control centres and introduce novel methods of production process control. The boundaries between physical components (such as machine tools, industrial robotics, etc.) and the virtual model are getting smoother. Consequently, a new term CPPS (Cyber-Physical Production System) has appeared [1].

CPPS is an information technology concept, which involves the integration of computing resources with physical processes. Obviously, a computer network acts as a go-between in such systems. The network is the CPPS component that needs to meet the strictest requirements because workability of the whole system depends on the network reliability and fault-tolerance. Currently, flexible decentralized productions slowly replace centralized networks integrated with different automation tools. In other words, interoperability is phasing out integration.

However, machine tools with CNC systems and industrial robotics are still the main components of automated production. Such devices are quite complex and its integration via SCADA just complicates production control process. A monolithic architecture of the system for machine tools and robotics control leads to the monolithic SCADA. Given the above does not match the main CPPS principle—simplification of the production process, creating a dynamic distributed environment and carrying out appropriate actions autonomously. Thus, we can conclude that a novel distributed CNC equipment system architecture should be developed in order to create a highly effective CPPS.

The base of such architecture is the open interface that allows hardware components stay autonomous but at the same time able to communicate with other components if necessary. Obviously, it is impossible to create such interface without any unification of the data transferring protocol. This protocol should work above all protocols and provide the ability to transparently exchange information between operation system processes as well as between heterogeneous network nodes that can act as a software service running under the operating system and a physical controller [2].

All these requirements are met by message queues. A message queue is an asynchronous communication protocol so the receiver and sender do not communicate directly and use a message queue instead. Generally, message queues have a message size constraint. Message queues improve system fault-tolerance as messages stay in the queue and can be handled even if the transmitter node fails.

Message queues guarantee message delivery, at least until one node is active and can handle it, and do not violate message order. In academia, a queue becomes both a handy instrument to implement complex asynchronous message-driven systems and it is a wide-ranging topic in computer science research. In other words, the authors consider a queue of two sides.

978-1-5386-0837-1/17/$31.00 ©2017 IEEE

The first one copes with an inner queue optimization. There are many papers on improved queuing schedulers that were found by a queue algorithm modification [3], [4]. For example, K. S. Kahlon modified a weighted fair queuing algorithm for the WiMAX network that contains both a real and a non-real-time traffic [5]. Also, queues are used in energy-awake scheduling, so Matthew Andrews and Lisa Zhang offered a rate-adaptive version of the Weighted Fair Queuing scheduling algorithm and proved its energy consumption [6]. In order to gain a high performance, researchers started developing novel hybrid genetic queuing algorithms for scheduling [5], [7].

The other one solves some application domain tasks by queue employment. For example, J. Acosta-Cano uses the message queuing system MSMQ and web service-oriented solutions for coupling a shop floor software with a set of production equipment [8]. In order to create an enterprise messaging system, Gursev Singh Kalra used ActiveMQ along with the Java Messaging Service [9]. Researchers from Kielce University of Technology implemented a wireless network for mobile robot applications using MQTT protocol [10]. Also, ZeroMQ framework usage is described in several papers [11]–[13]. Message queues have become pretty ubiquitous and used in many spheres including manufacturing [14], [15].

This article gives the comparative analysis and performance evaluation of different message queue protocols for binary JSON data transmission. Currently, this protocol is the industry standard for both data storing and big data transmission. Also, it is widely employed for the creation of distributed networks of the Internet of Things (IoT) as well as web services. The format is textual, which simplifies the work with it. At the same time, there is a sufficient number of software tools that make it possible to implement binary serialization of JSON in order to reduce the message size and increase the message queue, which in turn increases the responsiveness of the distributed CNC system. The rest of the article will be organized as follows. In Section II, specific motivations for using a message queue in the distributed CNC system will be discussed. The basic network scenarios will be presented in Section III. A description of the research methodology will be presented in Section IV. The results are displayed in Section V along with the analysis. The main findings will be summarized and main conclusions will be drawn in Section VI.

II. Motivation

The distributed CNC system that is part of the CPPS is a challenge for networks and computers. It is assumed that the protocol based on the message queue will become the main protocol of interaction within a large production environment. Accordingly, there are numerous scenarios for using this protocol as well as limitations imposed on software and hardware.

Obviously, the distributed CNC system will be implemented on the basis of the existing corporate computer network, consisting of a wide variety of computers, programmable logic controllers, low-level controllers, servers, etc. As a medium for data transmission, a high-speed wire connection (based on copper or optical cables) or wireless can be used. This is especially important in production because technological equipment can be located in sufficiently large workshops where it is economically inexpedient to use only a wired connection.

It should not be forgotten that the corporate network of an enterprise could be geographically distributed. Today, the industry is increasingly using cloud technologies for such things as performing complex calculations related to the preparation of control programs for CNC equipment or the creation of virtual models of equipment.

All this leads to the necessity of economical use of network resources to achieve the maximum performance of a distributed CNC system. The purpose of this work is to estimate the network overhead associated with the packaging and unpacking of binary JSON data as well as transfer it through the message queue. A comparison will be made between the bandwidth of the message queue and "raw" TCP sockets both for transmission between the operating system processes and for transmission over various network connections. As a result, the fastest variant of data transmission in the distributed CNC system under consideration will be determined [16]–[20].

III. Network Scenarios

The possible scenarios of network interaction between components of a distributed CNC system were considered.

A. Operation System Processes Interaction

It is proposed that the distributed CNC system components can be installed on a personal computer (PC) or a general-purpose server. The possibilities of inter-process communication (IPC) for POSIX compliant operating systems were considered. This scenario is the most common in the implementation of high-level CNC logic. For example, the G-code control program-parsing module interacts with trajectory optimization and moves the planning module in this way. It is obvious that both of these modules do not solve tasks in real time, and as a rule they can be installed on ordinary PCs.

B. Interaction Hardware and Software Modules in a Heterogeneous Computer Network

This is the most common interaction scenario in a distributed CNC network. It is assumed the possibility of transparent communication between low-level controllers, computers and general-purpose servers. Each of the participants of this interaction can act both as a client and as a server. It can be implemented as a two-point or a broadcast connection. For example, a low-level programmed logical controller that collects data from sensors and broadcasts them to the network can act as a data source.

In this case, all nodes interested in obtaining these data will act as the receivers: motion controllers, servers for collecting statistical data on the production process, interface modules of the operator, etc. Another example is a video camera that broadcasts a video signal from the processing area. This signal can be displayed on the physical monitor or on the user

interfaces widget, transmitted to machine vision system for analysis, etc. On the other hand, specialized control signals are transmitted directly between two predetermined nodes.

C. Interaction with the User (Operator)

The users interface is one of the most important parts of every CNC system. The use of the decentralized approach for the design of such systems provides enormous advantages over the classical monolithic systems. In the classic systems, the users interface is inseparable from the CNC controller. In early CNC systems, user interaction was mostly done by using physical control components such as buttons, switches, character indicators, and so on. But in modern systems the interface becomes virtual, that is, it is displayed on the CNC system screen. The number of physical control components is reduced, and only components that provide operational safety (for example, an emergency stop button) or ease of management (for example, a pendant hand-wheel) remain.

It is obvious that the use of a virtual interface that is tightly bound to the controller is inconvenient and inexpedient. Such interfaces have low flexibility and it is hard to adapt for users needs as well as almost impossible to organize remote management. Therefore, to create the operators interface for a distributed CNC system, the use of web technologies is assumed.

In this case, the main control terminal will be a web browser installed either on a personal computer or on a mobile device. At the same time some of the computing tasks will be transferred directly to the browser, which requires the ability to receive data from the distributed network directly through it. This creates a separate network scenario because the client web applications use their own stack of technologies that impose certain restrictions on the operation of the distributed CNC system.

IV. Methodology

The measurements described in this article were carried out by transferring data between two processes connected via TCP/IP. This allowed us to create a stable test environment, which makes it possible to exclude overhead costs associated with the preparation of data. First, a comparison of the methods of inter-process communication was carried out. The performance of the TCP/IP protocol was analysed and compared to UNIX sockets and shared memory. Then load testing of the channels for various network scenarios was performed.

Furthermore, the performance analysis of various binary JSON data packaging/unpacking tools was performed. After that, the bandwidth analysis for several message queue libraries was performed on the basis of which the most productive variant was selected. At the end, the final calculation of the maximum bandwidth for all the network scenarios under consideration was performed.

A. Test Set-ups

Four test set-ups were used, representing different architectures and performance components of a distributed CNC system:

1) *General PC* is a low-performance computer used in a distributed CNC system to solve a wide range of tasks. Specifications: Intel Core i5 CPU M 520 @ 2.40 GHz, 4 GiB RAM. Operating system: Ubuntu 16.10, GNU/Linux 4.8.0-41 x86-64.
2) *Laptop* is a portable computer that can be used by the operator of a distributed CNC system as a terminal, or for remote control. Specifications: MacBook Pro, Intel Core i5 CPU @ 2.60 GHz, 8 GiB RAM. Operating system: OS X 10.10.5, Darwin 14.5.0.
3) *Server* is a high-performance computer used for the most resource-intensive tasks of the distributed CNC system. Specifications: 2 x Intel Xeon E5620 CPU @ 2.40 GHz, 32 GiB RAM.
4) *Programmable logic controller (PLC)* is a embedded system-on-chip used to implement low-level algorithms in real-time mode. Specifications: Amlogic S905 Quad Core Cortex-A53 @ 1.5 GHz 64bit ARMv8 CPU with Mali-450 GPU, 2 GiB RAM. Operating system: Ubuntu 16.04 LTS.
5) *Virtual private server (VPS)* is a cloud-based virtual machine used to optimize the processing power of a distributed CNC system by transferring part of the software components to the dedicated servers. Specifications: Intel Xeon CPU E5645 @ 2.40 GHz, 512 MiB RAM. Operating system: Ubuntu 16.04.2 LTS, GNU/Linux 2.6.32-042stab120.18 x86-64.

B. Libraries and Frameworks

The performance evaluation required analysing the development tools (software libraries) for packing/unpacking the binary JSON data and the message queue implementation, and selecting the ones most appropriate to solve the given problem. The following requirements for libraries and frameworks for packaging/unpacking binary JSON data were formulated:

- Open source license
- Support for programming languages C, Python, Go and JavaScript (the main programming languages used to implement the distributed CNC system)
- A small amount of the software library
- High-quality documentation

As a result of the performed analysis, the following libraries were selected for testing: MessagePack, BSON (Binary JSON), UBJSON (Universal Binary JSON), CBOR (Concise Binary Object Representation).

The requirements for message queue libraries are similar to the requirements for the packaging/unpacking libraries; additionally an open API is required to add new transport protocols. As a result of the analysis, the following libraries were selected for testing: ZeroMQ, NanoMsg, RabbitMQ, and ActiveMQ. The first two can work without a dedicated message broker. For others, the use of a special message broker is mandatory.

V. Results

A. Comparing the Bandwidth of TCP and UNIX Sockets

To compare the bandwidth of TCP and UNIX sockets, the following methods were used:

1) socat utility, transfer from memory to disk, shared memory is used.

```
$ sudo dd if=/dev/urandom of=/dev/shm/data.dump
    bs=1M count=512
$ sudo socat -u -b32768 UNIX-LISTEN:/tmp/unix.sock
    ./data.dump &
$ sudo socat -u -b32768 "SYSTEM:dd
    if=/dev/shm/data.dump bs=1M count=512"
    UNIX:/tmp/unix.sock
$ sudo rm -rf /dev/shm/data.dump
```

2) `socat` utility, transfer from memory to memory, shared memory is used.

```
$ sudo dd if=/dev/urandom of=/dev/shm/data.dump
    bs=1M count=512
$ sudo socat -u -b32768 UNIX-LISTEN:/tmp/unix.sock
    /dev/shm/data.dump.out &
$ sudo socat -u -b32768 "SYSTEM:dd
    if=/dev/shm/data.dump bs=1M count=512"
    UNIX:/tmp/unix.sock
```

3) `socat` utility, transfer from memory to the null device, shared memory is used.

```
$ sudo dd if=/dev/urandom of=/dev/shm/data.dump
    bs=1M count=512
$ sudo socat -u -b32768 UNIX-LISTEN:/tmp/unix.sock
    /dev/null &
$ sudo socat -u -b32768 "SYSTEM:dd
    if=/dev/shm/data.dump bs=1M count=1024"
    UNIX:/tmp/unix.sock
```

4) `socat` utility, transfer from `/dev/zero` (special file that provides as many null characters as are read from it) to the null device.

```
$ sudo socat -u -b32768 UNIX-LISTEN:/tmp/unix.sock
    /dev/null &
$ sudo socat -u -b32768 "SYSTEM:dd if=/dev/zero
    bs=1M count=512" UNIX:/tmp/unix.sock
```

5) `nc` utility, data transfer through a common file.

```
(SEND) $ sudo pv /dev/zero | nc -U /tmp/socket
(RECV) $ sudo nc -lU /tmp/socket > /dev/null
```

6) `iperf` utility, data transfer through a TCP socket, TCP window size is 2.5 MB.

```
(SEND) $ iperf -s; (RECV) $ iperf -c localhost
```

The results of the comparative analysis are presented in Table I. It can be seen that the best way for inter-process communication in POSIX-compatible operating systems is to use TCP sockets.

B. Determining the Average Bandwidth of the Network Connections

To determine the average bandwidth of the network connections, the following methods were used:

1) HTTP server, 10 GB of data is transferred.

```
(SEND) $ dd if=/dev/urandom of=test bs=1024K
    count=10000
(SEND) $ sudo python -m SimpleHTTPServer 80
(RECV) $ wget <server-ip>/test
```

TABLE I
INTER-PROCESS COMMUNICATION BANDWIDTH COMPARISON

Test Set-up	Bandwidth, Gbit/s					
	Meth. 1	Meth. 2	Meth. 3	Meth. 4	Meth. 5	Meth. 6
General PC	0.69	5.76	7.46	12.03	6.33	22.3
Laptop	1.79	4.47	5.98	10.18	3.65	23.2
Server	2.41	2.76	2.75	3.23	3.13	31.6
PLC	1.98	1.96	2.38	2.86	2.04	4.61

2) `iperf` utility, TCP window size is 84.3 KB, 10 GB of data is transferred.

```
(SEND) $ ipref -s; (RECV) $ ipref -c <send-ip>
```

3) `netperf` utility, 10 GB of data is transferred.

```
$ netperf -H <any-ip-in-network> -4 -T TCP_STREAM
```

4) `nuttcp` utility, 10 GB of data is transferred.

```
(SEND) $ nuttcp -S
(RECV) $ nuttcp -vvv -il <send-ip>
```

5) `nttcp` utility, 10 GB of data is transferred.

```
(SEND) $ nttcp -i; (RECV) $ nttcp -t -T <send-ip>
```

The results of the comparative analysis are presented in Table II. Obviously, the use of the HTTP protocol leads to the emergence of additional overhead, so this method is excluded from the calculation of the average bandwidth.

TABLE II
BANDWIDTH OF THE NETWORK CONNECTIONS

Conn.	Bandwidth, Mbit/s					
	Meth. 1	Meth. 2	Meth. 3	Meth. 4	Meth. 5	Avg.
Wi-Fi*	**5.25**	5.95	8.00	6.86	5.47	6.57
Cable*	**559.00**	941.00	907.77	941.27	940.72	932.69
Cloud**	**12.44**	13.70	13.67	13.60	14.42	13.85

*connection between *General PC* and *Server* in the local network

**connection between *Server* and *VPS* via the Internet

C. Performance Evaluation of the JSON Packers

Fig. 1 shows the dependence of the compression ratio on the JSON file size. The JSON files consist of an ordered list of elements. Each element consists of a field name and a value. Field names are strings. Values include numbers, strings, Booleans, or null objects. All tests are written in Python. The figure shows that for all libraries the dependence of the compression ratio on the JSON file size is practically not observed.

Then the dependence of the packing/unpacking rate on the size of the JSON file was obtained. The results are shown in Fig. 2 and 3. As one can see, for all the libraries under consideration this dependence is linear. The CBOR library demonstrated the best time for both packaging and unpacking. The next step was to test the performance of the CBOR libraries, which were written in different programming languages. There were 1000 packaging/unpacking cycles; the size of the original JSON file was 15 MB. The test results are shown in Table III.

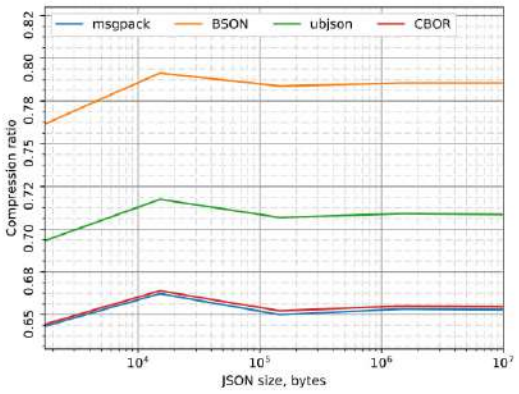

Fig. 1. Compression ratio as a function of the JSON file size.

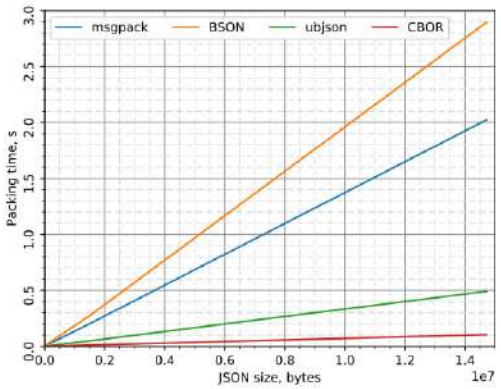

Fig. 2. Packing time as a function of the JSON file size.

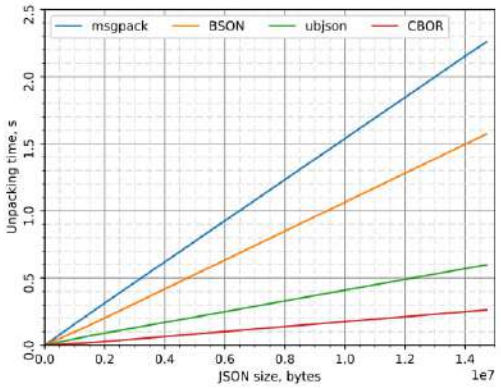

Fig. 3. Unpacking time as a function of the JSON file size.

D. Performance Evaluation of the message queue protocols

Fig. 4 shows the dependence of the bandwidth on the message size. For each method, one million messages were transmitted, and inter-process communication is used. Based on the results of the analysis, we can conclude that the most rational message size ranges from 1 KB to 100 KB. The NanoMsg and ZeroMQ libraries demonstrated the best performance. NanoMsg was selected as the main message transfer protocol

TABLE III
BANDWIDTH OF THE CBOR LIBRARIES

Library	Bandwidth, Mbit/s	
	Packing	Unpacking
Python 3.5.2 CBOR	112.73	19.13
Python 3.5.2 CBOR2	7.59	4.30
C libcbor (GCC 6.2.0 20161005)	244.13	273.07
JavaScript cbor (Google Chrome 56.0.2924.87)	76.19	100.21

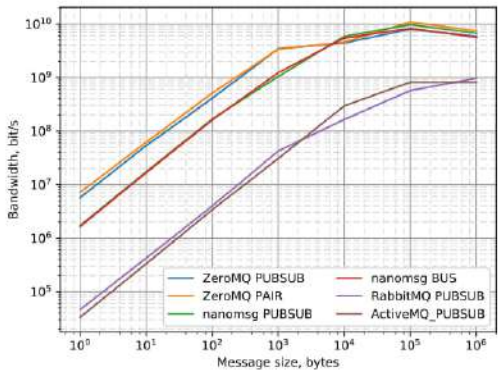

Fig. 4. Message queue protocols bandwidth as a function of the message size.

in the distributed CNC system under consideration. The main reasons for this solution are better compatibility with POSIX and a more flexible API.

Table IV shows the results of measuring the bandwidth of the NanoMsg protocol for various network connections. All tests were written in Python where the message size is 1 KB and one million messages were sent in each test.

TABLE IV
BANDWIDTH OF THE NANOMSG PROTOCOL

Library	Bandwidth, Mbit/s	
	Nanomsg PUBSUB	Nanomsg BUS
Wi-Fi	5.93	7.34
Cable	880.74	880.38
Cloud	15.18	14.98

E. Overall Evaluation of the Performance

To give an overall evaluation of the performance of the message queue protocol, we obtained the formula for the average bandwidth (1).

$$\mathcal{B}_a = \frac{\mathcal{M}}{t} = \frac{\mathcal{M}}{t_p + t_t + t_u} = \frac{\mathcal{M}}{\mathcal{M}/\mathcal{S}_p + \gamma\mathcal{M}/\mathcal{B}_t + \mathcal{M}/\mathcal{S}_u} = \\ = \frac{\mathcal{M}}{\frac{\mathcal{M}(\mathcal{B}_t\mathcal{S}_u)+\gamma\mathcal{M}(\mathcal{S}_p\mathcal{S}_u)+\mathcal{M}(\mathcal{S}_p\mathcal{B}_t)}{\mathcal{S}_p\mathcal{B}_t\mathcal{S}_u}} = \frac{\mathcal{S}_p\mathcal{B}_t\mathcal{S}_u}{\mathcal{B}_t\mathcal{S}_u + \gamma\mathcal{S}_p\mathcal{S}_u + \mathcal{S}_p\mathcal{B}_t}, \quad (1)$$

where $\mathcal{M}$ is a message size, t is total point-to-point time, t_p is packing time, t_t is transfer time, t_u is unpacking time, $\mathcal{S}_p$ is a JSON data packing speed, $\mathcal{S}_u$ is a JSON data unpacking speed, $\mathcal{B}_t$ is a message queue bandwidth, γ is a JSON compression ratio.

Frequently, the Python libraries demonstrate a performance that is approximately equal to the performance of C libraries. This is due to the fact that these libraries are just wrappers of the C libraries. In this case, the use of the Python libraries is preferable because of the simpler syntax of the Python programming language. However, the Python libraries under consideration did not show sufficient performance, so it is obvious that a C library was chosen to implement the server part. When implementing the client part, the JS library was used, since this is the only option for implementing programs in the browser.

Table V shows the results of calculations. The calculations show that average bandwidth of the described link channel is in the range of 6 % (the worst case, packaging/unpacking is required at both endpoints) to 145 % (the best case, packing/unpacking is not required at all, both endpoints use binary JSON data, which implies the ability to compress messages due to the fact that the binary JSON format takes up less space compared to the text JSON).

TABLE V
COMMUNICATION BANDWIDTH, CABLE CONNECTION, $\gamma = 0.65$

$\mathcal{S}_p \setminus \mathcal{S}_u$	JS library	C library	No unpacking
JS library	n/a^*	**57.06**	70.94
		(**6.22** %)	(7.60 %)
C library	67.50	117.70	203.18
	(7.24 %)	(12.62 %)	(21.78 %)
No packing	93.31	227.27	**1354.98**
	(10.00 %)	(24.37 %)	(**145.28** %)

*connection between client is not implemented in the system under consideration

VI. CONCLUSION

In this paper, we examine the performance of the transmission of JSON data through the message queue, which is the foundation of a distributed network of a CNC system. It was shown that the best time of packing/unpacking and the compression ratio of the JSON data were demonstrated by the CBOR library.

It was determined that the compression ratio does not depend on the volume of the original JSON file. Based on the analysis of advantages and disadvantages, the best message queue protocol is NanoMsg. Based on the results of the analysis, it was also concluded that the most suitable message size is from 1 KB to 100 KB.

The calculations have been made of the overall performance of the bundle of the NanoMsg protocol and the CBOR packer. Calculations showed that the bandwidth of this data channel is in the range of 6 % to 145 % of the average network bandwidth when using the raw TCP sockets, which can be considered a good result for this type of software.The source code of the tests is available at `github.com/futoke/message-queue-protocols`.

REFERENCES

[1] B. Vogel-Heuser, C. Diedrich, D. Pantförder, and P. Göhner, "Coupling heterogeneous production systems by a multi-agent based cyber-physical production system," in *2014 12th IEEE International Conference on Industrial Informatics (INDIN)*, 2014, pp. 713–719.

[2] M. Y. Afanasev and A. A. Gribovskiy, "Kontseptsiya adaptivnoy platformy tekhnologicheskogo oborudovaniya [An adaptive platform of technological equipment concept, in Russian]," *Isvestiya vusov. Priborostroenye*, vol. 58, no. 4, pp. 268–272, 2015.

[3] P. Valente, "Reducing the execution time of fair-queueing packet schedulers," *Computer Communications*, vol. 47, pp. 16–33, 2014.

[4] L. Rizzo and P. Valente, "On service guarantees of fair-queueing schedulers in real systems," *Computer Communications*, vol. 67, pp. 34–44, 2015.

[5] Akashdeep, K. Kahlon, and M. Kaushal, "Analysis of a queue length aware and latency guaranteed fuzzy-based adaptive resource allocator for WiMAX networks," *Optik—International Journal for Light and Electron Optics*, vol. 127, no. 1, pp. 357–367, 2016.

[6] M. Andrews and L. Zhang, "Rate-adaptive weighted fair queueing for energy-aware scheduling," *Information Processing Letters*, vol. 114, no. 5, pp. 247–251, 2014.

[7] S. Rashidi and S. Sharifian, "A hybrid heuristic queue based algorithm for task assignment in mobile cloud," *Future Generation Computer Systems*, vol. 68, pp. 331–345, 2017.

[8] J. Acosta-Cano and F. Sastrónáguena, "Loose coupling based reference scheme for shop floor-control system/production-equipment integration," *Journal of Applied Research and Technology*, vol. 11, no. 3, pp. 447–469, 2013.

[9] G. S. Kalra, "Threat analysis of an enterprise messaging system," *Network Security*, vol. 2014, no. 12, pp. 7–13, 2014.

[10] R. Kazala, A. Taneva, M. Petrov, and S. Penkov, "Wireless network for mobile robot applications," *IFAC-PapersOnLine*, vol. 48, no. 24, pp. 231–236, 2015.

[11] K. Kirsanov, "Software architecture of control system for heterogeneous group of mobile robots," *Procedia Engineering*, vol. 100, pp. 278–282, 2015.

[12] B. Goertzel, D. Hanson, and G. Yu, "A software architecture for generally intelligent humanoid robotics," *Procedia Computer Science*, vol. 41, pp. 158–163, 2014.

[13] V. Andreev, K. Kirsanov, P. Pletenev, Y. Poduraev, V. Pryanichnikov, and E. Prysev, "Technology supervisory control for mechatronic devices via the internet," *Procedia Engineering*, vol. 100, pp. 33–40, 2015.

[14] D. Stock, M. Stöhr, U. Rauschecker, and T. Bauernhansl, "Cloud-based platform to facilitate access to manufacturing it," *Procedia CIRP*, vol. 25, pp. 320–328, 2014.

[15] C. Morariu, O. Morariu, T. Borangiu, and S. Raileanu, "Manufacturing service bus integration model for implementing highly flexible and scalable manufacturing systems," *IFAC Proceedings Volumes*, vol. 45, no. 6, pp. 1850–1855, 2012.

[16] X. Huang, "Enhancing STEP-NC compliant CNC controller for distributed and reconfigurable environment in production line," in *2010 International Conference on Computer, Mechatronics, Control and Electronic Engineering*, vol. 2, 2010, pp. 106–109.

[17] V. Lesi, Z. Jakovljevic, and M. Pajic, "Towards plug-n-play numerical control for reconfigurable manufacturing systems," in *2016 IEEE 21st International Conference on Emerging Technologies and Factory Automation (ETFA)*, 2016, pp. 1–8.

[18] H. A. ElMaraghy, "Flexible and reconfigurable manufacturing systems paradigms," *International Journal of Flexible Manufacturing Systems*, vol. 17, no. 4, pp. 261–276, 2005.

[19] K. Latif and Y. Yusof, "New method for the development of sustainable STEP-compliant open CNC system," *Procedia CIRP*, vol. 40, pp. 230–235, 2016.

[20] H. Hong, D. Yu, X. Zhang, and L. Chen, "Research on the data hungry problem in CNC system based on the architecture of real-time multitask," in *2011 3rd International Conference on Computer Research and Development*, vol. 2, 2011, pp. 103–108.

Modular Industrial Equipment in Cyber-Physical Production System: Architecture and Integration

Maxim Ya. Afanasev, Yuri V. Fedosov, Anastasiya A. Krylova, Sergey A. Shorokhov
ITMO University
St. Petersburg, Russia
amax@niuitmo.ru, yf01@yandex.ru, {ananasn94, stratumxspb}@gmail.com

***Abstract*—The design of numerical control systems for industrial machinery is a difficult task, especially when you create universal modular equipment with computer numerical control (CNC). This article presents a modular approach to the design of such systems. The modular control system under consideration is based on a multi-agent network, in which each entity (module) acts as an integral and indivisible part of the object as well as the enlarged structure. This approach allows one to combine the advantages of classical hierarchical control systems with the flexibility and reliability of decentralized multi-agent networks and also to carry out seamless integration of equipment built on the basis of this architecture into a cyber-physical production system (CPPS). The proposed architecture is implemented in the control system of a universal industrial platform. As an example, the apparatus for selective laser curing of a photopolymer to the surfaces of arbitrary shapes is represented. The general structure of the installation determined by the basic hardware and software modules and the network communication protocol are described.**

I. Introduction

Since microprocessors became powerful enough and as their memory capacity increased in the 1980's, the possibility of development from a Numerical Control (NC) to a Computer Numerical Control (CNC) system was obtained. Due to rapid scientific and technological advancement, a great amount of manufacturing equipment was designed and foundational units of CNC system were determined.

Each CNC system consists of three main units: Numerical Control Kernel (NCK), Human-Machine Interface (HMI), and Program Logical Control (PLC). Moreover, the Numerical Control Kernel usually includes an interpreter, an interpolator, an acceleration/deceleration controller, and a position controller, that are literally the main functions. Given the above, various architectures of CNC systems have appeared. However, despite the diversity, all of them were of a closed type. Therefore, changes and additions of new functions or hardware modules were not allowed.

The closed, "monolithic" type of industrial systems architecture is the most common, because it allows minimizing internal errors, eliminating internal unauthorized interference, and also standardizing internal components and control commands. Such equipment is aimed at large enterprises with mass production of products. However, for small companies with a wide range of products, the acquisition and operation of machines with closed architecture is unprofitable. The process of changing the production lines will take a long time, some equipment may be idle or too loaded.

Nevertheless, industrial equipment became a complex set of hardware and software in which providing resource and work coordination, tasks and queries management are necessary things. A CNC system should be flexible, scalable, and interchangeable. Another important fact is an open system architecture that allows you to change the configuration of industrial equipment easily.

Consequently, at the present time the manufacturing community has an increasing desire to gain as much flexibility, agility, and intelligence as possible. These requirements are usually considered at the high level of the *Manufacturing Execution System* [1], [2], where the control of multiple production process elements such as inputs, stuff, machines and support services are enabled. However, the CNC system architecture plays a significant role too. To overcome the closed type architecture problem, this paper introduces a open modular approach to implement an open and flexible architecture.

The proposed approach involves a control system as the consolidation of modules that have their own unique identity but are part of a larger whole. All modules are joined by the multi-agent network, where each module provides an open interface for communication. That is, the system is not just modular but also compatible with the protocol layer. Therefore, the difference between software and hardware modules disappears and makes interconnection easier.

II. Related work

In academia, the multi-agent approach is employed quite actively. Despite the age of the approach, which is rooted from the late 1960's, it is still under development. Researchers have been continually offering various formalization and hybrid architectures.

For example, the combination of web technologies and multi-agent control is described by Ebisa [3]. The system is based on web-services and applied to smart power grids. In addition, another extension of the multi-agent paradigm to smart grids is proposed by Ferreira [4].

Additionally, cloud-based systems are pretty ubiquitous nowadays. However, for manufacturing systems, cloud

computing and integration are quite novel. The cloud-based approach poses essential challenges such as security, reliability, availability, etc. Such systems are mostly referred to service-oriented types, but Issa suggested fuzzy design to represent cloud-based intelligent manufacturing [5]. In pursuit of increasing system intelligence, other researchers employ bees algorithms and artificial neural networks to make a agent evolve [6].

Currently, researchers are engaged in various industrial production fields employing agent technologies from production cooperation within collaborative manufacturing to mobile industrial robotics control [7], [8], automated stores management, and stock accountings. Undoubtedly, intelligent manufacturing has many tasks to solve [9], [10]. Supply tasks [11], [12], scheduling tasks [13], [14], job allocation [15], operational control [16], etc. can be solved using the modular multi-agent methodology. Virtually all systems supporting steps of the product lifecycle can be implemented that way.

However, a heterarchical architecture of control is the major multi-agent application field. To overcome CNC challenges a multi-agent CNC controller was developed in Katholieke Universiteit Leuven using an object-oriented paradigm [17]. Also, Lesi V. represents a modularized and decentralized CNC system [18]. The control was released for the conception of modular equipment and met challenges for a plug-n-play automation system. Moreover, a quickly customized and developed CNC system architecture is described by Li S.Z. [19]. As seen above, a huge amount of academic research on intelligent manufacturing has been done including CNC.

The rest of the paper is structured as follows. Section III shows a multi-agent approach to the design of the control system, its mathematical model with regard to domain knowledge features and an example of integration into a CPPS. In Section IV the proposed system is described along with modules such as the driver, vision, laser head, graphical user interface modules, etc. Also, its protocol specification is given in detail. Section V contains the conclusion and future plans concerning system development and improvements.

III. Control system design

A. Multi-agent communication

As a basis for the developed modular control system for CNC equipment, a multi-agent network has been selected. In its essence, any such network is a heterarchical flat structure, that consists of independent entities, called agents. Resources and/or specific tasks are also usually represented by agents. Such an approach allows for the creation of simple and fault-tolerant systems because agents do not require any predefined information about other agents. Consequently, small agent activity deviations can be easily fixed. Global data is forbidden due to agent independence. Therefore, the total performance is highly influenced by agent communication rules.

Obviously, this imposes certain limitations on the performance of the control system, which is unacceptable in solving real-time problems, such as the control of modular industrial equipment. Consequently, various approaches to the implementation of multi-agent networks in the industry were considered [20]. Based on the results of the analysis, the agent-holonic approach was recognized as the most appropriate solution.

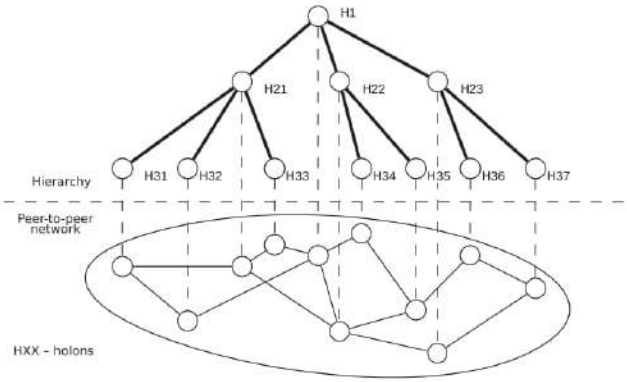

Fig. 1. The general structure of a holonic multi-level system

This approach was first proposed in the 1990's by IMS consortium (Intelligent Manufacturing Systems) [21]. This organization initiated a number of projects aimed the creation of "future manufacturing." One of these projects was Holonic Manufacturing System.

The "holon" term was first used by Arthur Koestler (Kösztler Artúr) to explain biological and social systems evolution [22]. Such systems organize stable transitional forms, that are independent during the whole developing process. However, it is quite difficult to determine where there are wholes and parts in such system organization (Fig. 1).

Hard and soft holons are able to create and modify hierarchical relationships (holarchy) as well as take part in several holarchies at the same time. Even so, holons are absolutely detached as they can receive information from higher order holons and be insensitive to environment perturbations.

Each holon in this distributed network has the ability to be self-monitoring, do a diagnosis, and complete a self-healing function. If an emergency situation occurs, holons complement each other. Every holon possesses the communication interface for information exchange with other holons and holarchy via a bidirectional protocol. Therefore, the current state holon information is available not only for neighbourhood holons but the whole holonic system as well. In addition, each holon has a number of utility functions for different emergency cases. Wherein, a faulty holon can be excluded from the holarchy without any negative consequences.

A distributed holonic system can be based on a holarchy, and consists of autonomous holons (hardware and software modules) with a decentralized control. This architecture of the control system allows for expanding the functionality of the industrial equipment, but can increase fault tolerance, stability, and reliability.

B. Mathematical model

A mathematical model of the holonic control system is determined by a unified abstract representation of the multi-agent environment (MAE), allowing it to further proceed to the holonic structures. Unification is achieved through a uniform

definition, which allows on the basis of data about the agent (holon) the ability to draw conclusions about the state of its environment and vice versa [23].

This approach, in turn, makes it possible to build an isomorphic MAE on the basis of a given MAE, where a group of agents can be combined to create a single holonic agent (holonic synthesis) and a single agent can be divided into the group of the holonic agents (holonic decomposition) [24].

The holonic multi-agent environment of the represented control system can be represented as a tuple (1):

$$\mathfrak{h}\mathfrak{E} \stackrel{\text{def}}{=} \langle \mathfrak{A}, \mathfrak{E}, \mathfrak{P}, \mathfrak{F}\mathfrak{e} \rangle, \tag{1}$$

where $\mathfrak{A} = \{\mathcal{A}_{\mathcal{G}1}, \ldots, \mathcal{A}_{\mathcal{G}n}\}$ is a set of the agents. Each agent is a tuple (2):

$$\mathcal{A}_{\mathcal{G}i} = \langle \mathcal{S}_i, \mathcal{P}_i, \mathcal{A}_i, \mathcal{F}_i \rangle, \tag{2}$$

where $\mathcal{S}_i$ is a set of the agents' states, $\mathcal{P}_i$ is a set of the perceptions of the agents, $\mathcal{A}_i$ is a set of the actions of the agents, $\mathcal{F}_i$ is the agent function (3):

$$\mathcal{F}_i \colon \mathcal{S}_i \times \mathcal{P}_i \to \mathcal{S}_i \times \mathcal{A}_i. \tag{3}$$

$\mathfrak{E}$ is a set of the environment's states, $\mathfrak{E} = \{\varepsilon_1, \ldots, \varepsilon_n\}$. $\mathfrak{P} \colon \mathfrak{E} \to (\mathcal{P}_1 \times \ldots \times \mathcal{P}_n)$ is a perception function, $\mathfrak{F}\mathfrak{e} \colon \mathfrak{E} \times (\mathcal{A}_1 \times \ldots \times \mathcal{A}_n) \to \mathfrak{E}$ is a environment function.

It is assumed that there is a discrete time scale, where the time step is given by the transition from one time point to another time point. Each agent $\mathcal{A}_{\mathcal{G}i}$ for all states of the environment $\varepsilon \in \mathfrak{E}$ and all states of agents (4):

$$(s_1, \ldots, s_n) \in \mathcal{S}_1 \times \ldots \times \mathcal{S}_n \tag{4}$$

at each step of computation gets its local object of perception $\mathfrak{P}^i(\varepsilon)$ through the perception function. Agent calculates its new action (5) based on its current state s_i, and its own perception of this state.

$$\mathcal{A}_i = \mathfrak{F}_i^1(s_i, \mathfrak{P}^i(\varepsilon)), i \in \overline{1, n} \tag{5}$$

In the other words, a state of the environment is changing under the influence of agents. Equation (6) defines a successive state of the environment, and (7) defines a successive state of the all agents i in the environment.

$$\varepsilon' = (\varepsilon, \mathcal{A}_1, \ldots, \mathcal{A}_n) \tag{6}$$

$$s_i' = \mathfrak{F}_i^2(s_i, \mathfrak{P}^i(\varepsilon)), i \in \overline{1, n} \tag{7}$$

The transition state function (8):

$$\widehat{\mathfrak{F}\mathfrak{e}} \colon \mathfrak{E} \times \mathcal{S}_1 \times \ldots \times \mathcal{S}_n \to \mathfrak{E} \times \mathcal{S}_1 \times \ldots \times \mathcal{S}_n \tag{8}$$

defined as follows (9):

$$\widehat{\mathfrak{F}\mathfrak{e}}(\varepsilon, s_1, \ldots, s_n) = (\varepsilon', s_1', \ldots, s_n'), \tag{9}$$

combines the agents' states and environment's states. Consequently, the perception function, the agent function, and the environment function are part of the transition state function $\widehat{\mathfrak{F}\mathfrak{e}}$.

Based on the previous explanations, it is clear that any holon of the considered control system should take its own object of perception, calculate action and next state from previous state without the reference to the other holons' states with any environment state.

Two multi-agent environments (10) and (11):

$$\mathfrak{h}\mathfrak{E} = \langle \mathfrak{A}, \mathfrak{E}, \mathfrak{P}, \mathfrak{F}\mathfrak{e} \rangle \tag{10}$$

$$\mathfrak{h}\mathfrak{E}' = \langle \mathfrak{A}', \mathfrak{E}', \mathfrak{P}', \mathfrak{F}\mathfrak{e}' \rangle \tag{11}$$

are isomorphic if there exists a bijective function (12):

$$\mathfrak{B} \colon \mathfrak{E} \times \mathcal{S}_1 \times \ldots \times \mathcal{S}_n \to \mathfrak{E}' \times \mathcal{S}_1' \times \ldots \times \mathcal{S}_n' \tag{12}$$

such that, for all states in the environment (13):

$$(\varepsilon, s_1, \ldots, s_n) \in \mathfrak{E} \times \mathcal{S}_1 \times \ldots \times \mathcal{S}_n, \tag{13}$$

we have the equality (14):

$$\widehat{\mathfrak{F}\mathfrak{e}}'(\mathfrak{B}(\varepsilon, s_1, \ldots, s_n)) = \mathfrak{B}(\widehat{\mathfrak{F}\mathfrak{e}}(\varepsilon, s_1, \ldots, s_n)) \tag{14}$$

In connection with this, attention is drawn to the fact that every MAE has two special configurations. The first is the environment without the agents, in which all state changes are defined by an environment function $\mathfrak{F}\mathfrak{e}$. The second is the MAE with only one permanent state and only one agent, where all states of this agent defined by agent function $\mathfrak{B}$.

This formalization enables us to reduce the MAE, which includes a several agents, to the environment with the only one agent. Concurrently, all the other agents are represented as the subject of this environment.

If (15) is multi-agent environment,

$$\mathfrak{h}\mathfrak{E}^\dagger = \langle \{\mathcal{A}_{\mathcal{G}1}, \ldots, \mathcal{A}_{\mathcal{G}n}\}, \mathfrak{E}, \mathfrak{P}, \mathfrak{F}\mathfrak{e} \rangle \tag{15}$$

then for each $i \leq n$, there exists an isomorphic multi-agent environment. Without loss of generality, let $i = 1$. Let us create an environment (16):

$$\mathfrak{h}\mathfrak{E}^\dagger = \langle \{\mathcal{A}_{\mathcal{G}i}\}, \mathfrak{E}, \mathfrak{P}, \mathfrak{F}\mathfrak{e} \rangle \tag{16}$$

such that $\mathfrak{E}' = \mathfrak{E} \times \mathcal{S}_2 \times \ldots \times \mathcal{S}_n$, $\mathfrak{P}' = \mathfrak{P}^1$, for all $(\varepsilon, s_2, \ldots, s_n) \in \mathfrak{E}'$ and $\mathcal{A}_1 \in \mathcal{A}_1$ environment function (17):

$$\mathfrak{F}\mathfrak{e}' \colon \mathfrak{E}' \times \mathcal{A}_1 \to \mathfrak{E}' \tag{17}$$

defined as follows (18):

$$\mathfrak{F}\mathfrak{e}'((\varepsilon, s_2, \ldots, s_n), \mathcal{A}_1) = (\varepsilon', s_2', \ldots, s_n'), \tag{18}$$

where (19):

$$\varepsilon' = \mathfrak{F}\mathfrak{e}(\varepsilon, \mathcal{A}_1, \mathfrak{F}_2^1(s_2, \mathfrak{F}^2(\varepsilon)), \ldots, \mathfrak{F}_n^1(s_n, \mathfrak{F}^n(\varepsilon))) \tag{19}$$

is a successive state ε in the $\mathfrak{h}\mathfrak{E}^\dagger$ for $\mathcal{A}_1$ and the other actions of agents.

Equation (7) defines a successive state s_i of the all agents i in the $\mathfrak{h}\mathfrak{E}^\dagger$ for $2 \leq i \leq n$, which implies the existence of an isomorphism between two MAE, and we can begin to identify the concepts of "holonic synthesis" and "holonic decomposition."

Hence, on the basis of the above equations, we can conclude that it is possible to combine several agents into one holon and vice versa. Thus, we can dynamically change the structure of the designed computer numerical control system.

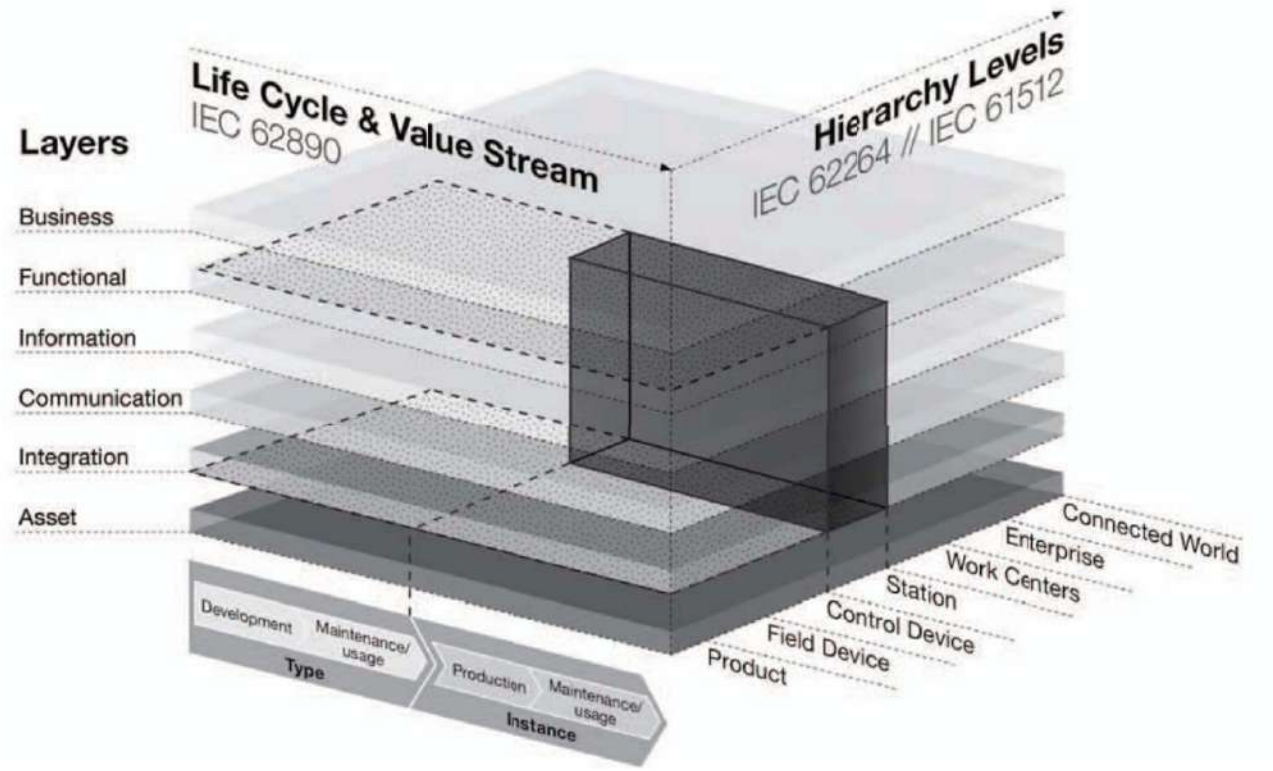

Fig. 2. Mapping the proposed architecture on the RAMI 4.0 reference model

C. CPPS integration

The developed mathematical model in itself allows creating a control system for modular industrial equipment. This system will have the flexibility of multi-agent networks and the stability of hierarchical management structures. However, this is not enough to argue that industrial equipment built on the basis of this control system can become a part of the cyber-physical production system. The fact of the matter is that for today there are some standards of representation of industrial cyber-physical systems. From the diversity of standards associated with the concepts of industrial cyber-physical systems and Industry 4.0, two most well-known ones can be distinguished. This is *Reference Architectural Model Industrie 4.0 (RAMI 4.0)* [25] and *Industrial Internet Reference Architecture (IIRA)* [26]. From these two options, the RAMI 4.0 model was chosen. The main prerequisite of this decision was the greater compliance of this reference architecture with Russian production standards, due to the fact that the IIRA is more oriented to the industry of the United States of America, and RAMI 4.0—European Union.

The RAMI 4.0 model is hierarchical and three-dimensional, and in three axes it displays the subject area in terms of information technology, product life cycle and production organization.

On the axis "Hierarchy Levels", modular industrial equipment is located at the levels of "Station" and "Control Device" because such equipment used in repetitive or discrete production. It is also obvious that such equipment is not included in the "Field Device" area, since it has no connection with them directly. The axis of the "Value Stream" is of the greatest interest, since this axis is directly related to IT, which means that it represents a control system for modular technological equipment. Thus, the multi-agent control system considered earlier is located at the levels "Integration", "Communication", "Information" and "Functional" (Fig.2).

At the level of "Integration", modular equipment interacts with the "real world", that is, with the operator through the HMI. At the level of "Communication", it communicate with the internal modules and other equipment to solve problems in real time. At the level of "Information", it provides compatible data representation and access for all components of the CPPS. At the functional level, the modular equipment control system describes its capabilities as services for other components of the CPPS, which is the basis for describing the business process (the next level).

From all of the above, it can be concluded that the modular equipment can be projected onto the reference model RAMI 4.0. In this case, the control system architecture corresponds to the hierarchy of levels of the given model. Modular industrial equipment can become a component of the CPPS.

IV. Application Example

The proposed control system is used in the universal industrial platform. As an example, let us consider the implementation of one of the types of equipment constructed on the basis of this platform. This is an apparatus for selective

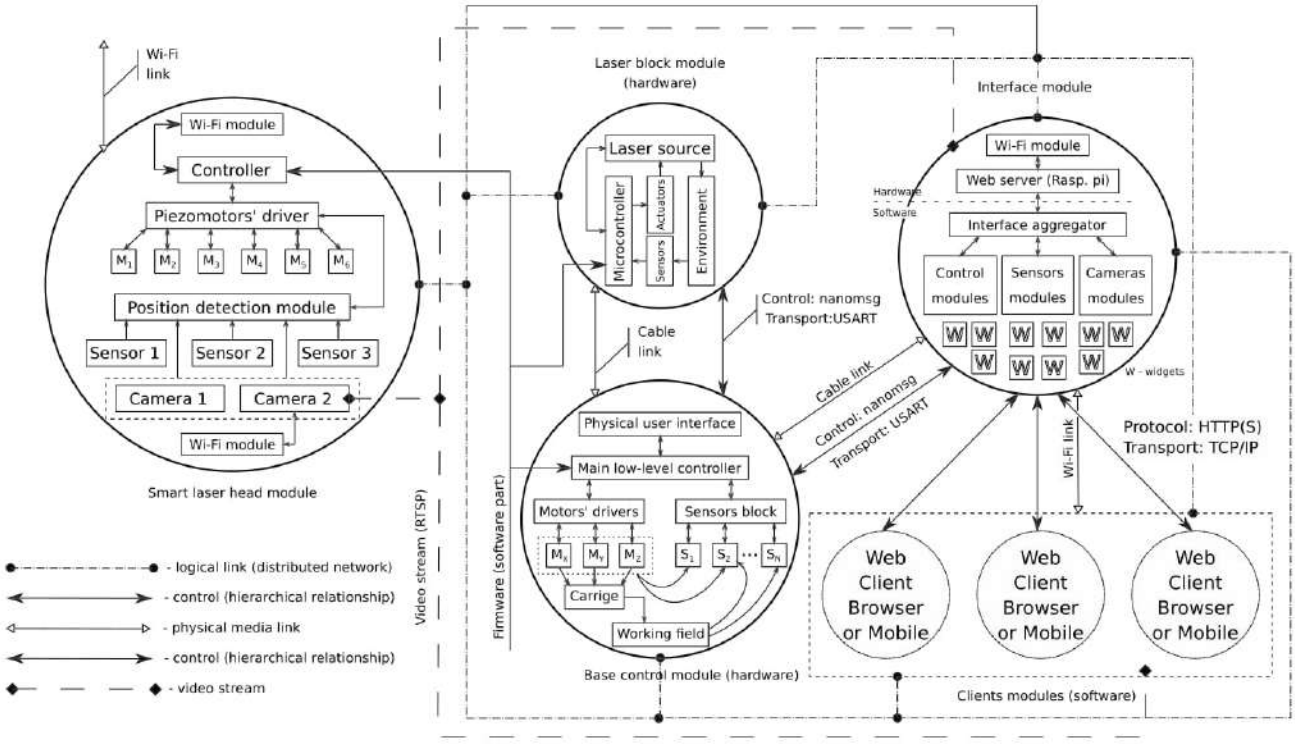

Fig. 3. The architecture of the modular equipment

laser curing of the photopolymer to the surfaces of arbitrary shape [27–31].

This equipment uses laser processing technology the same as the LDF (Laser Direct Structuring) process. The experimental testbed represents a device for arbitrary surfaces processing by laser emition. The equipment contains a fixed coordinate table where processing the object takes place, a laser head is placed over the work table and is movable in a horizontal plane along two coordinates with a laser emitting source (laser module) connected optically with the laser head. Structure of its modules and their interaction with each other are described in more detail in previous paper [32].

The laser head has a gyro-stabilized suspension system based on the modified Stewart platform. Linear piezo motors are used for platform moving. The suspension system allows the laser beam deflection on two self-perpendicular surfaces and focus distance changing by the head objective moving along Z-coordinate.

Every module of this apparatus is a close independent entity, highly specialized and solved only one of a few tasks, determined in the system, such as query processing, data providing or CNC program generating. Entities can interact on the unified protocol level and, as already reported above, software, hardware, and mixed parts are similar from the system's point of view. The proposed system uses an open protocol which allows adding new modules "on the fly." It is a necessary property for making system design easier and for providing an interchangeability of modules. The system examined below uses holonic units for blocks' presentation which contains a set of holons interacting inside their unit.

Platform components are represented as a set of small simple parts joined in a big device with a few functions, which interact with neighbours by data transitions. At this point each part must have a clear understanding of its own task regardless of what role it has: master or slave. The architecture of the modular equipment is depicted in Fig. 3, 3D-model of the apparatus in Fig. 4. Equipment can contain many types of modules that are classified into three main groups by functions.

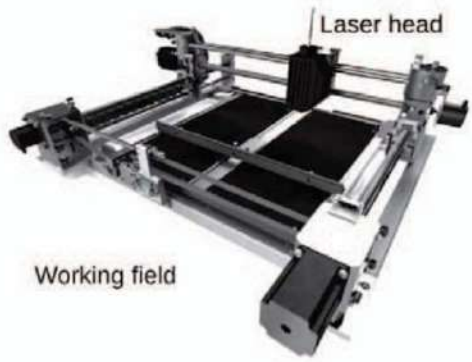

Fig. 4. Experimental apparatus

The interface module provides a link between users and equipment. The user connects to the server by the web-application and sends a request to the device with CAD-file or CNC-program; the CNC-program should be checked for accuracy. Furthermore, the application monitors the operation process and sends notifications to the user about important events. This unit is mainly software.

The base control module has the master role and contains the CNC-module and a set of drivers and sensors for the control of physical devices. CNC-program processing, sending current status signals to the server and device's display, and making decisions about the current situation are its main tasks. This unit is mainly hardware.

The smart laser head module is a slave for the controls unit and associates software and hardware modules. Basic types of holons:

- main camera, which runs availability and the size of blank detecting, reference point labels searching and watching the processing operations;
- the machine vision component with special frame processing algorithms;
- a laser module with surface incline detection camera;
- a set of sensors for self-diagnostics and modules work.

The link channel between units is based on NanoMsg, the socket library for a message exchange by network elements. The NanoMsg acts at the top of a transport layer, provides several types of communication, works at a high speed and is easy to scale. The library is open source project and it makes the system non-proprietary; a unified message standard allows for expanding the system while including new modules in it. The protocol acts between all modules except the client-server link, which uses HTTP-protocol.

V. Conclusion

In this work, advantages of the modular agent-holonic approach to the organization of industrial control system with numerical control was considered. A review of the literature on this topic was conducted. The main advantage of the approach is the equivalence of the components from the point of view of the system, which allows one to adapt the system to changing production conditions, which corresponds the principles of cyber-physical production. The correspondence of the Reference Architectural Model Industrie 4.0 (RAMI 4.0) is also shown. The system is based on the proposed peer distributed network of nodes and each of them performs its function in the system. A mathematical study was carried out to analyse the control system as a whole, and the behaviour of its components. The developed mathematical model describes the dependence of the behaviour of the agents represented the modules of the system on the states of the external environment. Also, this model shows the ability of modules to the composition and decomposition.

The description of the system components was given. According to it, the system has a number of basic blocks that perform a common function: the control unit, the client-server part and the smart unit that performs the support processing function. NanoMsg has been chosen as the basic communication protocol. It allows to make a one-tier system, and it will be open to the introduction of new components. The approach involves almost the complete autonomy of the components, which simplifies scaling and improves fault tolerance.

Acknowledgment

The authors would like to thank the anonymous reviewers for their valuable and constructive comments and suggestions that greatly contributed to improve the quality of the final version of the paper.

References

[1] H. Wang, L. Liu, Y. Fei, and T. Liu, *A Collaborative Manufacturing Execution System Oriented to Discrete Manufacturing Enterprises*. Cham: Springer International Publishing, 2014, pp. 277–285.

[2] S. Iarovyi, W. M. Mohammed, A. Lobov, B. R. Ferrer, and J. L. M. Lastra, "Cyber-physical systems for open-knowledge-driven manufacturing execution systems," *Proceedings of the IEEE*, vol. 104, no. 5, pp. 1142–1154, 2016.

[3] E. Negeri, N. Baken, and M. Popov, "Holonic architecture of the smart grid," *Smart Grid and Renewable Energy*, vol. 4, no. 2, pp. 202–212, 2013.

[4] Â. Ferreira, O. Cardin, and P. Leitão, "Extension of holonic paradigm to smart grids," *IFAC-PapersOnLine*, vol. 48, no. 3, pp. 1099–1104, 2015.

[5] H. Issa, E. Ostrosi, M. Lenczner, and R. Habib, "Fuzzy holons for intelligent multi-scale design in cloud-based design for configurations," *Journal of Intelligent Manufacturing*, pp. 1–29, 2015.

[6] J. A. Ahmed and A. M. A. Brifcani, "A new internal architecture based on feature selection for holonic manufacturing system," *International Journal of Mechanical, Aerospace, Industrial, Mechatronic and Manufacturing Engineering*, vol. 2, no. 8, p. 1431, 2015.

[7] D. Sun, *Synchronization and Control of Multiagent Systems*, ser. Automation and Control Engineering Series. CRC Press, 2011.

[8] J. Shamma, Ed., *Cooperative Control of Distributed Multi-Agent Systems*. John Wiley & Sons Limited, 2007.

[9] H. Zhang, R. Guo, and C. Geng, "An architecture model of cloud manufacturing based on multi-agent technology," in *The 26th Chinese Control and Decision Conference (2014 CCDC)*, May 2014, pp. 3725–3729.

[10] P. Zhao, P. Lou, M. Liu, and X. Wu, "Open CNC system design based on multi-agent technology," in *2010 International Conference on Computational Intelligence and Software Engineering*, Dec 2010, pp. 1–4.

[11] T. Dewa, M. S. Matope, F. Van Der Merwe, A. and L. Nyanga, "A web-based holonic inventory management system for tool and die workshop operations," 2014, pp. 14–29.

[12] P. Bečvář, P. Charvát, M. P. J. Pospíšil, A. Říha, and J. Vokřínek, "Explantech/extraplant: Production planning and supply-chain management multi-agent solution," *EXP—in search of innovation*, vol. 3, no. 3, pp. 116–125, 2003.

[13] A. Araúzo, J. R. Del-Olmo-Martínez, J. Laviós, J. and J. De-Benito-Martín, J. "Scheduling and control of flexible manufacturing systems: A holonic approach [programación y control de sistemas de fabricación flexibles: Un enfoque holónico]," *RIAI—Revista Iberoamericana de Automatica e Informatica Industrial*, vol. 12, no. 1, pp. 58–68, 2015.

[14] K. Wang and H. Choi, S. "A holonic approach to flexible flow shop scheduling under stochastic processing times," *Computers and Operations Research*, vol. 43, no. 1, pp. 157–168, 2014.

[15] M. Owliya, M. Saadat, R. Anane, and M. Goharian, "A new agents-based model for dynamic job allocation in manufacturing shopfloors," *IEEE Systems Journal*, vol. 6, no. 2, pp. 353–361, 2012.

[16] A. Ivanescu, N. M. Parlea, and A. Rosu, "Different approaches regarding the operational control of production in a flexible manufacturing cell," *Studies in Computational Intelligence*, vol. 402, pp. 241–254, 2012.

[17] I. Tanaya, P. J. Detand, J.-P. Kruth, and P. Valckenaers, "An object-oriented execution model for a machine controller holon," *European Journal of Control*, vol. 4, no. 4, pp. 345–361, 1998.

[18] V. Lesi, Z. Jakovljevic, and M. Pajic, "Towards plug-n-play numerical control for reconfigurable manufacturing systems," vol. 2016, nov 2016.

[19] Z. Li, S. and W. Lin, S. "Research on reconfigurable method of open CNC system software based on linux," *Applied Mechanics and Materials*, vol. 529, pp. 549–553, 2014.

[20] M. Y. Afanasev, "Razrabotka i issledovanie mnogoagentnoj sistemy dlja reshenija zadach tehnologicheskoj podgotovki proizvodstva [Development and research of a multi-agent system for solving the problems of technological preparation of production, in Russian]," Ph.D. dissertation, Saint-Petersburg National Research University of Information Technologies, Mechanics and Optics, 49 Kronverksky Pr., St. Petersburg, 197101, Russia, May 2012.

[21] L. Wang and A. Nee, *Collaborative Design and Planning for Digital Manufacturing*. Springer, 2009.

[22] A. Koestler, *The ghost in the machine*, the danube, 2. impr ed. London: Hutchinson, 1979.

[23] C. Gerber, J. Siekmann, and G. Vierke, "Holonic Multi-Agent systems," DFKI, Kaiserslautern, Germany, Tech. Rep. RR-99-03, 1999.

[24] M. Wooldridge, *An Introduction to Multiagent Systemes*. John Wiley & Sons Limited, 2002.

[25] F. Zezulka, P. Marcon, I. Vesely, and O. Sajdl, "Industry 4.0—an introduction in the phenomenon," *IFAC-PapersOnLine*, vol. 49, no. 25, pp. 8–12, 2016, 14th IFAC Conference on Programmable Devices and Embedded Systems PDES 2016.

[26] M. Weyrich and C. Ebert, "Reference architectures for the internet of things," *IEEE Software*, vol. 33, no. 1, pp. 112–116, Jan 2016.

[27] M. Y. Afanasev and A. A. Gribovskiy, "Kontseptsiya adaptivnoy platformy tekhnologicheskogo oborudovaniya [An adaptive platform of technological equipment concept, in Russian]," *Isvestiya vusov. Priborostroenye*, vol. 58, no. 4, pp. 268–272, 2015.

[28] M. Y. Afanasev, Y. V. Fedosov, and A. A. Nemkova, "Designing features of power optical units for technological equipment," *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, vol. 16, no. 2, pp. 244–250, 2016.

[29] M. Y. Afanasev and Y. V. Fedosov, "Optomechanical system," Patent RU 2 583 163, 05 10, 2016.

[30] ——, "Device for laser processing of the surface of random shape," Patent RU 161 667, 04 27, 2016.

[31] ——, "Design of an adaptive system for stabilization of a laser beam for CNC machine," in *Proceedings of the 19th Conference of Open Innovations Association FRUCT*, S. Balandin and T. Tyutina, Eds. FRUCT Oy, Helsinki, Finland, 2016, pp. 31–36.

[32] M. Y. Afanasev, Y. V. Fedosov, A. A. Krylova, and S. A. Shorokhov, "An application of microservices architecture pattern to create a modular computer numerical control system," in *Proceedings of the 20th Conference of Open Innovations Association FRUCT*, S. Balandin and T. Tyutina, Eds. FRUCT Oy, Helsinki, Finland, 2017, pp. 10–19.

An Application of Microservices Architecture Pattern to Create a Modular Computer Numerical Control System

Maxim Ya. Afanasev, Yuri V. Fedosov, Anastasiya A. Krylova, Sergey A. Shorokhov
ITMO University
St. Petersburg, Russian Federation
amax@niuitmo.ru, yf01@yandex.ru, {ananasn94, stratumxspb}@gmail.com

***Abstract*—Currently, the most common approach to Computer Numerical Control (CNC) system design is a monolithic architecture. However, the introduction of the concept of Cyber-Physical Production Systems (CPPS) requires a paradigm shift in the design of control systems. This paper suggests a new approach to developing modular industrial equipment using a microservices architecture pattern. Microservices architecture features are addressed, as well as advantages and disadvantages. A heterogeneous computer network, where nodes communicate via a message queue, is proposed as a basis for the computer numerical control system. Fault tolerance is provided by modules full autonomy and reliable messaging. Furthermore, NoSQL database, guaranteeing high data accessibility and increased data access speed, is applied. An apparatus for selective photopolymer laser curing of free-form surfaces is considered as an example. Common setup structure, as well as main hardware and software modules, are described. Moreover, a distributed network latency simulation was carried out to prove the viability of the proposed microservices architecture.**

I. Introduction

By analysing the evolution of modern industry it can be concluded that by far the most promising development destination is creating flexible distributed automated production lines. The novel mass production concept was established as a consequence of The Fourth Industrial Revolution (or Industry 4.0) as well as gradual CPPS spreading. "Hard" conveyor approaches are consequentially transformed into small batch custom production. Moreover, the development of small innovation enterprises and startups is still going on. On balance, the approach to modern intelligent equipment design is changing.

The first CNC machine tools appeared in the 1950's but their development still continues. However, the development has been mostly focused on mass production. So, CNC systems have always been complex, high-performance, and extremely expensive. Furthermore, the deployment process of such systems is long-term and modern system lifetime is dozens of years. Understandably, these circumstances make the implementation of modern communication technologies more complex in that each correction requires either the restructuring of the whole production system or the creation of additional control layers in order to link outdated equipment with a modern CPPS. Such an approach is the most appropriate during a transition period and ultimately phasing out the older system.

Therefore, it is necessary to rethink the CNC equipment design paradigm itself and the ability to plug new equipment into the informational communication environment via open protocol becomes quite important. Different attempts to implement such approaches have been made. For example, Grigoriev and Martinov in their research propose the approach of designing an agile CNC kernel based on an independent library platform. The open architecture of this system includes different abstract layers related to different human-machine interfaces (HMI) along with description possibilities of system components in different programming languages. The connection of system components is established via Fieldbus [1].

Bin describe an open platform for creating CNC systems and this system consists of a set of multipurpose components that can be employed repeatedly and a set of communication modules to connect them [2]. A similar approach is presented in the paper Ma [3]. Morales-Velazques et al. propose a platform with the open architecture based on a multi-agent system of software and hardware components, named MAD-CON. The hardware units of their proposed system integrate control and monitoring functions providing an FPGA-based open architecture for reconfigurable applications. The software components utilize the XML structure for system description files, gathering features like a flowchart descriptive language and a graphic user interface [4].

Works of Verba [5] and Prazeres [6] deal with an interesting concept, Fog of Things. This concept is a modern interpretation of the Internet of Things (IoT) that, in fact, is a foundation of many CPPS. Fog of Things allows for the creation of a more homogeneous informational communication environment thereby improving and simplifying component communication protocol.

II. Motivation and Case Study

Modern control systems are complex and solid, where each element is a black box with a hard hierarchical architecture. Everything in such systems is focused on quality assurance, reliability, and trouble-proof behaviour. The inertia of such systems forces SCADA developers to use a monolithic architecture because the equipment with distributed control can not be easily added to a distributed production environment.

Consequently, the emphasis is placed on the concept of integration. In other words, developers focused on joining heterogeneous components into one production system instead of using an interoperability concept. The interoperability concept

entails creating an open interface that allows components to stay in an autonomous state with communication ability.

Accordingly, we need to move on to the development of equipment with a modular architecture. Such architecture is based on two main postulates: unification and hybridization. Unification is an open software and hardware structure that allows for the assembly of new equipment and software as a "the smart children's designer." Dividing one product into several interchangeable building blocks with described input and output unification can then be achieved.

In summary, it can be noted that any manufacturing equipment consists of the following parts:

- Processing head actuator.
- Coordinate table that moves processing head actuator.
- CNC unit

It is obvious that the coordinate table is the most general-purpose unit where different processing head actuators can be mounted. Thus, the equipment type can be easily changed. For example, a milling machine tool can turn into a laser cutting machine switching the milling head to a laser head. Moreover, a 3D printer can be obtained by mounting a filament extruder.

Additionally, it is possible to create a setup that combines features of different types of equipment by applying the hybridization principle. For instance, an additive-subtractive machine can be assembled by joining some features of the 3D-printer and a three-coordinate milling machine tool. Also, a combination of milling cutter for roughing and laser for surface finishing and polishing is feasible.

However, it should not be forgotten that replacement of the processing head actuator includes changes in the control algorithm. Obviously, the CNC unit has to know about each possible processing head actuator in advance and a solid system with hierarchical control will be obtained again.

Therefore, it is necessary to determine the foundational part of the architecture and to define specifications as for communication protocol as well as for creating new software and hardware modules that can be dynamically plugged into the system. Subsequently, a foundational part will include a motor control algorithm of coordinate tables, and the rest of it will be external components.

In order to implement such architecture, a distributed network with software and hardware modules have to be developed. All modules are connected to the network, along with a connection that can be both wired and wireless. Each module gets an IP-address via DHCP and then sends a broadcast package to find a supervisor. The supervisor replies using a predetermined protocol and record the modules main physical parameters in a supervisor registry. For instance, a list of commands it can receive and data it sends to the coordinate table controller.

Such an approach provides the opportunity to change the paradigm of a production control environment completely. Unfortunately, it is not all that easy. The CNC system is not only a control algorithm, but a database, user interface, and physical environment communication component (also known as "embedded systems" and "microcontrollers"). To implement such system, a system approach is required and the authors suggest the use of a microservices architecture pattern.

III. State-of-the-art in service-oriented architecture

Service-oriented architecture (SOA) is a rapidly developing concept, which has increasingly been used in modern distribution systems. The basis of this architecture is a set of replaceable, low coupling components that possess unified interfaces for standard protocol communication. SOA was first mentioned in 2009 and used as a framework for complex web applications, aggregating data from multiple sources and representing it in a single view. Google maps, Amazon, and eBay are the most well-known examples. Consequently, SOA became widely used in cloud-based services such as SaaS, PaaS, IaaS, etc [7].

Microservice architecture (MSA) is a modern SOA interpretation that is applied to distribution systems. The service is a process executed by an operation system and communicates with another process via a network interface to reach a common goal. Microservices can be built using any programming language and framework but should interact with each other via a single protocol. Moreover, each microservice works with a limited set of tasks and performs a minimum number of operations.Currently, there is no clear definition of MSA. However, we attempt to pick out the key features of the approach that are most important for the development of the distribution CNC systems [8].

MSA provides the opportunity to easily replace the modules included in the system. For example, HMI for industrial equipment can be considered where a conventional HMI is a set of the physical elements, such as indicators, buttons, switches, etc. This interface is really convenient for mass production when equipment is used as a part of the conveyor line. Although, for a fabrication laboratory (FabLab) such a way of control is overhead. Using the MSA makes it easy to replace the physical controller with a virtual one that will enable the user to operate the equipment through the web-interface from any device connected to the network.

All modules are organized around functions. MSA allows you to split the functionality of each block. For example, the milling head consists of a physical part—motor, collet, controller, sensors, and a logical part—control unit. Physically, both parts are on the same block because the control unit has its inner memory (containing low-level and high-level logic) and can be connected to the consolidated network [9].

In terms of MSA, it is two different services. One of them is in charge of low-level motor control commands, data acquisition, etc. The other is a high-level HMI by which the user can control the settings of a production process or develop programs in ISO-7bit programming language. Each of these services is autonomous and has its own algorithm.

On the other hand, each of them is in charge of an exact function and know nothing about other microservices implementation. Certainly, duplication is possible as the same function can be implemented in different microservices and in different ways. However, it does not violates the low coupling

principle and is mostly considered as code writing discipline. MSA provides the opportunity to connect modules written in different languages seamlessly, but it doesnt forbid the unification of programming tools.

Each microservice is elastic and easy to modify, and along with that, it is a complete program product. This means that developers need to stick to cohesion and coupling principles during CNC system development. Conversely, it helps to focus on development and debugging of a particular block, which allows the modification of the entire system behaviour.

Increasingly, publications describing microservices architecture application for reconfigurable production systems are appearing. For instance, da Silva [10] represent architecture for reconfigurable production systems based on MSA and an agent-holonic approach. The suggested system supports knowledge exchange in a heterogeneous production system and provides the opportunity to control equipment placed on geographically remote production areas. The paper demonstrates that the architecture allows such a system to be fault-tolerant and reliable even in difficult production conditions.

Furthermore, Vresk [11] and Butzin [12] describe MSA that they use to increase interoperability and scalability of IoT. It is commonly known that this conception is associated with the notion of Industry 4.0. Therefore, a represented approach can be used to CNC systems development. Based on these examples, we can conclude that MSA is a modern, rapidly developing trend of CPPS, which can be successfully employed for creating a CNC system.

IV. Communication protocol

As previously stated, simple and open MSA has to apply lightweight communication protocol. Because distributed network, that is a basis of the approach, is heterogeneous, it is necessary to use "smart" message receivers and "silly" communication channels. Such a decision gives the opportunity to make system modules more independent and focused on the exact task. Also, it is important to remember that network performance can be obtained using a lightweight message bus and reducing the amount of communications between system components.

It appears that the most obvious way of data exchange in TCP/IP networks is "raw" sockets. Bytes data flow can be transmitted over this interface, which is pretty convenient for many applications. However, all modules in a developed system deal with structured data so packages will be more appropriate. Each package is a message from one module to another. A receiver module, that receives a message, has to confirm it (if content is comprehensible and can be handled) or reject it.

Though such messages are quite similar to those that are used in XML-RPC, SOAP, BPEL and WSDL, an important difference is showing up. All these protocols are used in enterprise applications and are inapplicable for embedded systems. To create a lightweight open and extensible control system necessary to minimize the abstraction layer and combine low-level sockets with routing ability. Thus, the most appropriate communication way between CNC system modules is message queues [13].

A message queue is an asynchronous communication protocol, so the receiver and sender dont communicate directly and use a message queue instead. Generally, message queues have a message size constraint so that data is transmitted via one message or more. Following message queue features allow us to conclude that this way to transmit data is the most efficient for designing a CNC system for several reasons [14]:

- Message queues allow system components to stay most independent from each other and avoid possible deadlocks when one of the participants has to wait for resources to release that are necessary for data transmission.
- Message queues give the opportunity to save system resources because there is no need for network buffers where untransmitted and unhandled data is usually stored. In fact, the message queue is a multipurpose network buffer on its own, which does not depend on any node or process.
- Message queues possess of feasible volume and throughput scalability.
- Message queues smooth network load peaks. It is considerably important for the designed system due to a possibility of different traffic types mixing together. In particular, traffic can be divided into high priority (sensors data, control commands) and low priority (G-code loading, a video stream from processing area) traffic according to suggestions in *Section VI* CNC system. The vehicle control system uses two independent CAN buses (high-speed for sensors and control commands and low-speed for comfort functions) to solve this problem, which doesnt simplify its architecture.
- Message queues improve system fault-tolerance as messages stay in the queue and can be handled even if the transmitter node fails. A g-code program can be considered as an example. An operator creates a g-code program using interface block and then transmits it to the execution block as text command flow. Obviously, the program will be divided into several messages placed in the queue. If the interface block freezes, a watchdog commits that fact and initiates operation system reboot. However, during the reboot execution the block keeps handling messages from the queue and data transmitted process is not interrupted.
- Message queues guarantee message delivery, at least until one node is active and can handle it.
- Message queues do not violate messages order. As a rule, messages can be received in the order they were sent.

Currently, there is a sufficient amount of various implementations of described approach [15–18]. All of them can be divided into commercial (proprietary) and open source (free license). Among commercial implementations the following frameworks and libraries should be noticed: IBM WebSphere MQ, Oracle Advanced Queuing, Amazon Simple Queue Service (SQS), StormMQ, IronMQ. Along with these, message queues are implemented in popular commercial real-time operating systems such as (RTOS) QNX and VxWorks.

The most interesting open source libraries are Apache ActiveMQ, Apache Kafka, Apache Qpid, Beanstalkd, HTTPSQS, JBoss Messaging, JORAM, RabbitMQ, Tarantool, Celery, HornetQ, StormMQ, NATS Messaging, ZeroMQ, nanomsg, Apollo, Darner, Gearman, RestMQ.

Only open source tools were analysed as the developed system is an open one. The following requirements based on analysis were established:

- The library has to be written in C/C++ because the designed system contents components are based on microcontrollers where programs can be written only in these languages.
- The library has to provide access to the message queue without a broker—a special node for addressing, routing, and queuing. This component reduces system fault-tolerance because a broker is a single point of failure, which is unacceptable for CNC systems. Certainly, redundancy can be provided by having additional reserve brokers. This, however, completely negates the main advantage of the system with the broker—reducing the complexity.
- There have to be an open API to add new transport protocols. Most of the message queues are based on TCP and websockets. In developed system connection between interface block and execution block established via UART. Therefore, a message queue has to be adapted to that transport protocol.

Given the above requirements, the most appropriate message queue library is *nanomsg*. This library is written in pure-C without any dependencies. Nanomsg is quite compact and allows developers to use it in embedded systems as well as in POSIX-compatible systems.

At the moment, the library implements following communication patterns:

- PAIR—Simple one-to-one communication.
- BUS—Simple many-to-many communication.
- REQREP—Allows building clusters of stateless services to process user requests.
- PUBSUB—Distributes messages to large sets of interested subscribers.
- PIPELINE—Aggregates messages from multiple sources and load balances them among many destinations.
- SURVEY—Allows querying the state of multiple applications in a single go.

The developed system uses PUBSUB for broadcast messages and BUS for decentralized exchange between system blocks (Fig. 1). Data in these queues, which are based on nanomsg, is transmitted as BLOB. Therefore, *Messagepack* (a binary format that is compatible with JSON and packs data up to 15–20 % more efficiently) is applied to transmit database entries [19], [20].

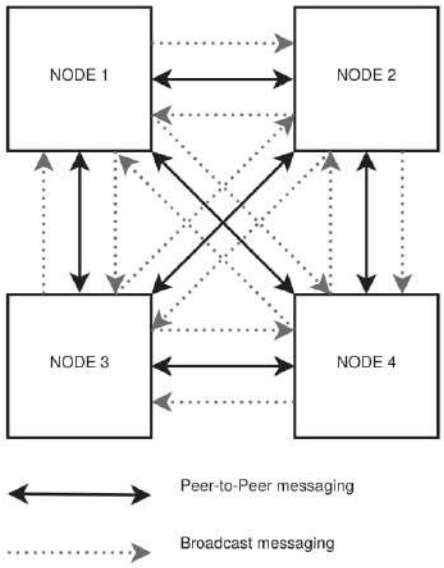

Fig. 1. PUBSUB and BUS communication patterns

V. Distributed Database

For the effective realization of storing data in a decentralized network, which is the core CNC-equipment control system, it is necessary to solve the scaling problem. Obviously, the described open architecture does not allow using traditional relational databases. Despite the fact that the current relational databases support replication and scaling, similar solutions still do not provide the proper data accessibility and consistency.

That is why using the so-called NoSQL databases is most appropriate. The main features of such databases are:

- Availability of data, which means that each call to database is guaranteed to return the requested data or reports that it is impossible to obtain the data.
- The system can change its state even without new input data because state changing can be necessary to reach the consistency of data. NoSQL databases do not have the database schema that contains descriptions of the content, structure, and data integrity constraints, so at some moment the data integrity may be compromised and then automatically restored.
- All data remains consistent, a system does not check all the data, i.e. it does not follow the integrity of each transaction, which improves performance and availability.

Currently, there are three basic NoSQL paradigm realizations. The first, *"key-value" data storages*, which are in fact ordinary hash-tables. They are usually used as an alternative to the distributed file systems and can be used for creating various file caches as well as problems with "Big Data". They often keep data in temporary storage distributed in random access memory that gives a serious performance boost. The following systems have the same solution: Dynamo, Riak, Berkeley DB, FoundationDB, HyperDex, InfinityDB, LMDB, MemcacheDB, and Redis [21–23].

The second, *document-oriented databases*, which represent hierarchically structured data storages. In fact, these are databases based on the JSON standard which makes them especially suitable for use in web applications. The main representatives of this category are CouchDB, MongoDB and eXist.

The third, *graph databases* are used for presenting data with a large number of connections. It is the most suitable for social network organization. Examples of this type of database are Neo4j, OrientDB, AllegroGraph and InfiniteGraph.

Also, it should be noted that other NoSQL databases design approaches exist but they are a logical development of the three basic branches or a combination of them. Obviously, micro service architecture of CNC-equipment control system realization needs a database with advantages combined from document-oriented databases and "key-value" storages. That is why the following requirements were formulated:

- Serverless realization. Many of the considered NoSQL databases have a dedicated server, which organizes interaction and replication. The presence of a server complicates the inclusion of new modules of the management system, as it requires certain actions related to server administration.
- Availability of atomicity during data transfer and transaction support that is necessary to create a system that works in "soft real-time" conditions.
- The possibility to keep all data in one file without having to create temporary files. The designed system suggests the possibility of connecting modules with different capacities and internal storage capacity so the database should use as little resources as possible.
- Cross-platforming and unification on the API level. Systems realized on C/C++ languages are preferred.
- Open-source software project.

According to the criterion, a comparative analysis was made, which resulted in the onset of merits and demerits and the UnQLite database was selected. This system is a NoSQL serverless database. In contrast to the well-known NoSQL engines like MongoDB, Redis, and CouchDB, UnQLite requires no installation or configuration and does not start a separate process to access the data. All data is stored in a single file with standard JSON serialized data. Otherwise, UnQLite is a storage for key-value pairs with cursor support and keeping data in HDD as well as RAM possibility.

In addition to this, UnQLite can work as document storage and support the transactions. The UnQLite engine is fully programmed in ANSI C language which allows using this database even on embedded systems used in a project. In the designed project, UnQLite will be used for storing structured data (modules, related to the implementation of the user interface of the controller and complex logic), as well as rapid storage of data obtained from the sensors (storage type "key-value"). In the last mode, UnQLite supports transactions, and the data stored in this way are best suited for the rapid transmission through a message queue

VI. Application Example

Let us consider the implementation of the proposed CNC controller microservice architecture as an example of an apparatus designed for selective photopolymer curing. This equipment uses laser processing technology the same as LDF (Laser Direct Structuring) process. The experimental installation represents a device for arbitrary surfaces processing by laser emition. The equipment contains a fixed coordinate table (Fig. 2) where processing the object takes place, a laser head (Fig. 3) is placed over the work table and is movable in a horizontal plane along two coordinates with a laser emitting source (laser module) connected optically with the laser head.

The laser head has a gyro-stabilized suspension system based on the modified Stewart platform. Linear piezo motors are used for platform moving. The suspension system allows the laser beam deflection on two self-perpendicular surfaces and focus distance changing by the head objective moving along Z-coordinate. This allows performing the following tasks. The first task is the laser ray is focused at the processing point. The second task is complex surfaces

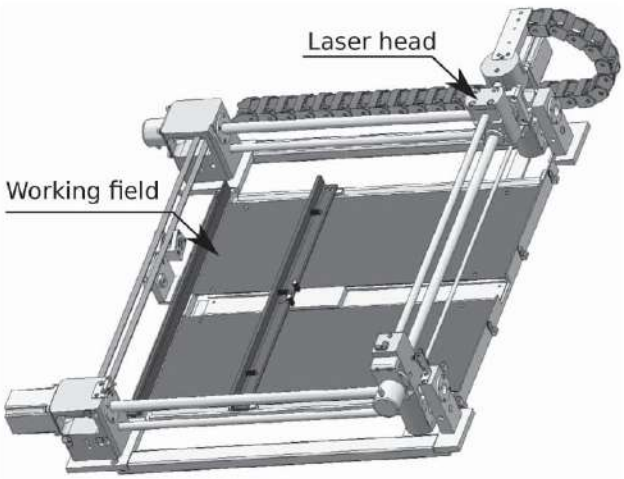

Fig. 2. Coordinate table

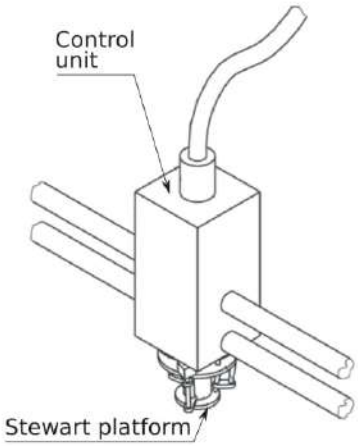

Fig. 3. Laser head

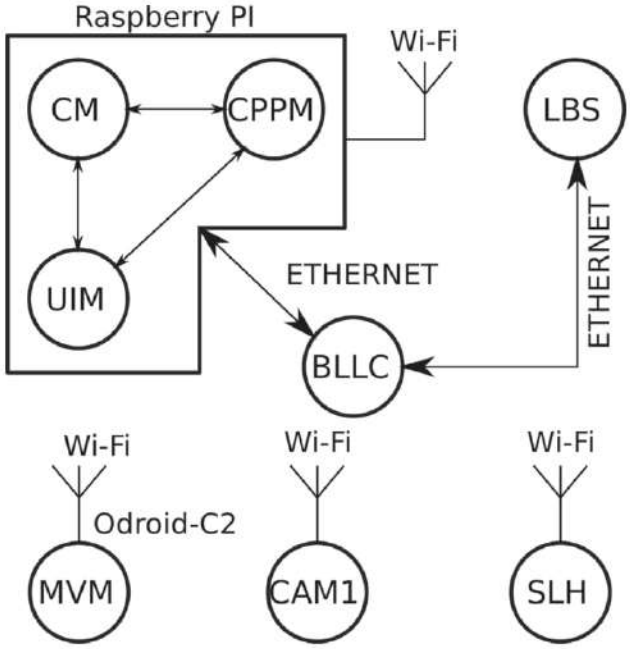

Fig. 4. Equipment control system

processing availability by optical axis deflecting. The slope allows the laser beam to always remain perpendicular to the processed surface and respectively contact patch is not distorted. The third task is active vibrations compensation. High-precision single-axis solid state accelerometers and a camera with a telescopic objective, for real-time spot form checking, are used to record vibrations. In addition, this laser head is equipped with the second camera for automatic setting of a reference point and watching the processing area.

A. Equipment control system

The equipment control system contains (Fig. 4):

Communication Module (CM). Physically it looks like the microcomputer Raspberry Pi with the operating system, Debian Linux, installed on it. This modules functions are tasked with the recording and dispatching of the other modules (fault tolerance of system at all providing).

Base Low-Level Controller (BLLC). Physically it is a control module based on the Cortex-M4 microcontroller with the real-time operating system, FreeRTOS. The main BLLC appointment is to generate control signals for coordinating table actuators and to process sensors signals. The controller is fully self-contained and able to work even without connection to the main control network. This module obtains programs in ISO 7-bit (G-code) programming language as an input data. G-code is loaded by nanomsg protocol, and Ethernet is used as physical level protocol. Also by this protocol, BLLC sends to the net the data from sensors, which can be taken and interpreted by every node of the control network. The rest of the unit is completely independent and after G-code loading and processing, it works in standalone mode. The the work algorithm of BLLC processes all contingencies by itself even without connection with the control network. In addition, it registers information about all the events that BLLC transmitted to the network and is recorded in the event log.

In addition to the control program, BLLC can receive single control signals, for example, "move the carriage to point (X; Y)", "move along X-axis with the velocity V", and "start the self-calibration process." These commands are needed to interact with the operator of the apparatus, and for automatic zero point (reference point) search procedure initiated by the machine vision module (MVM). Additionally, the low-level controller has a debug port, USART, which allows the operator to interact with the controller by command line interface. This function is necessary in the control system debug stage and in case of contingency appearances because of faults in the network. BLLC also transmits to the network the control signal for the laser beam source (LBS). At the point of view of G-code, a laser beam is just a tool like a cutter or a chisel. The controller can send the command to turn the laser on or off or to set the emitting capacity (tool rotation speed analog). There is bilateral data exchange between BLLC and LBS. BLLC sends one of the control signals, LBS gets it and sends to BLLC readiness confirmation and changes the laser state afterwards. In case of failure or another contingency, BLLC takes an appropriate answer from LBS. At the same time, this signal is obtained by other control network components, and all systems turn to the recovery mode. The same happens if the connection with the LBS is lost. The LBS has internal storage to accommodate a database that keeps an event log, CNC programs and user interface components.

Laser Beam Source (LBS). It has the same architecture as the BLLC part, i.e. it is based on a Cortex-M4 microcontroller with the operating system, FreeRTOS, on it and it is also connected to the network. As already mentioned, LBS functions are turning the laser on and off (it is carried out by changing the position of the movable mirror), changing its power (by PWM) and receiving and processing the data from the sensors. LBS has an internal logic and can work off contingencies in standalone mode and transmit sensor data to the network. In the case of failure, LBS transmits an emergency stop and recovery mode signal. Despite the fact that the LBS has the slave role towards BLLC, it constantly interrogates it. In the case of an absence of an answer for some time, it also distributes an emergency stop signal. It turns out that these two network nodes produce a constant mutual monitoring that increases the resiliency of the whole control system. LBS has internal storage to accommodate a database that keeps an event log, CNC programs and user interface components.

Smart Laser Head (SLH). SLH is a software and hardware module based on a Cortex-M4 microcontroller with the operating system, FreeRTOS. The main functions of SLH are: mechanical vibrations compensation, the inclination of the optical axis adjustments for complex surfaces processing and laser beam self-focusing. As the others, the module is autonomous and independently handles all emergencies with the transmission of the corresponding signals in the network. It also has internal storage to accommodate a database that keeps an event log, CNC programs and user interface components.

Machine Vision Module (MVM). It is a software and hardware module based on the Odroid-C2 microcontroller with operating system, Debian Linux. MVM functions are: search for a reference point on the workpiece and transfer the video

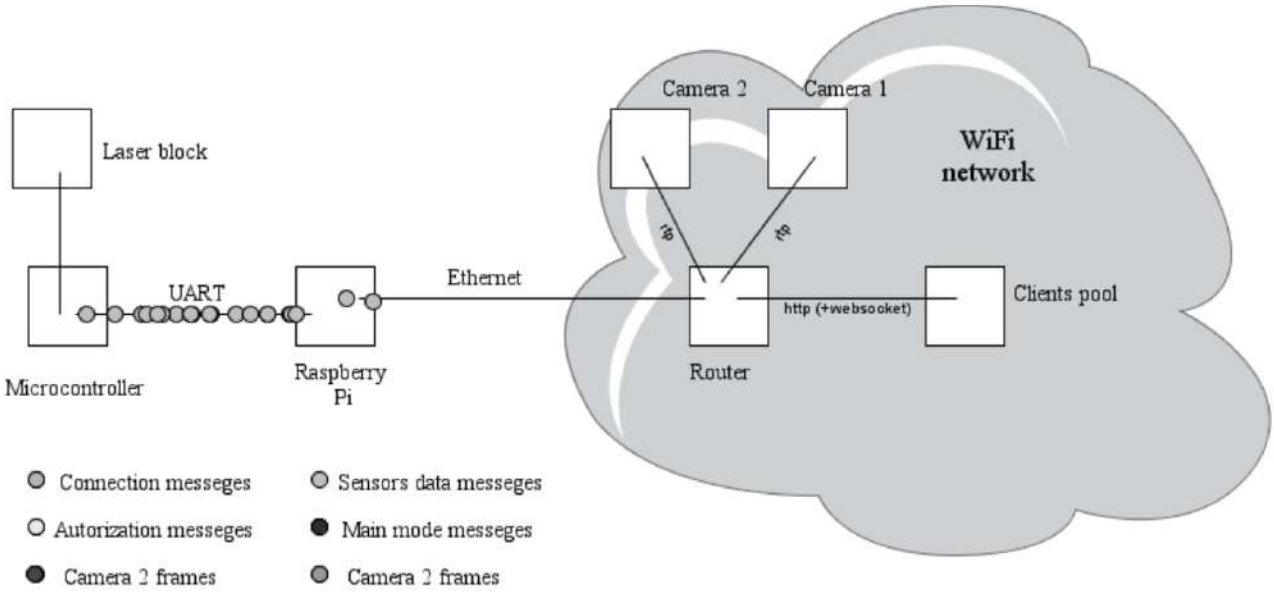

Fig. 5. AnyLogic model graphic representation

stream to display on the operator interface. BLLC and MVM interact by this algorithm: in front of the CNC program running on BLLC execution it asks for MVM coordinates of the reference point, MVM starts to scan the workspace (by sending control commands to BLLC) and, when the reference point is found, MVM turns into command waiting mode. MVM failure is critical only in case of searching a workpiece zero point process, in other cases, CM just alerts the operator that MVM is not available, but processing does not stop. The MVM also has internal storage to accommodate a database that keeps an event log, CNC programs and user interface components.

User Interface Module (UIM). It is a software module and is an aggregator for unit control elements (displays and widgets) to display them in the browser. Control elements are automatically loaded into it during the interaction with the modules. The basic idea is that the entire interface logic is written using a technology stack of HTML5 + ECMAScript + CSS, and stored directly in the database of each of the modules. During initialization of apparatus, all these static files are loaded to UIM, and then transmitted over the network for display in the browser. From the viewpoint of realization, UIM is a high-performance multi-threaded web-server written in Go programming language, and a display library of control units (buttons, switches, sliders, text fields, etc), widgets and displays. This server is physically placed on the same microcontroller with CM. However, it should be noted that due to microservice architecture, it could be located on a single computer or server.

Control Programs Preprocessing Module (CPPM). It is used for the preparation of G-codes directly in the browser. Like hardware modules interfaces, it is a web-application dynamically included in common equipment interface by UIM. It is physically placed on the same microcontroller as the CM and the UIM, and as the latter can be put on a separate server.

VII. Simulation and Numerical Results

It is well known that every distributed system has one significant disadvantage—its productivity depends on data throughput. In the case of complicated topology, bottlenecks can emerge, which in turn may limit the overall bandwidth of the network or at the very least will create load imbalance as a result of increased contention across the slow link. Obviously, the reviewed distributed microservice architecture is completely adequate for the demand of self-sufficiency for all modules. All operations connected with the manufacturing process are executed only in real-time mode and are not dependent on network delays. Nevertheless, all modules are connected to each other, hence when delays occur, the total time of operations such as initialization, jump from one control program to another or interacting with the operator can increase. Consequently, the network simulation model was designed and main modules load analysis was carried out.

The analysis considers that the reviewed network is heterogeneous, being composed of segments, that uses different data transfer mediums (Wi-Fi, or cable connection) and different transport protocols (TCP, RVDS). To calculate the network load metrics queueing theory was used. HTTP protocol also CPU-intensive load by most critical modules were used instead of absolute values of communication channel capacity and network delays.

This model is simplified, and not accounting for the CPPM and LBS parameters calculation, as soon these modules are working effectively in stand-alone mode and consume no network capacity. It should be noted that a Wi-Fi router is included in the model. This component is not a part of reviewed microservice architecture, but it has an influence on the network capacity. For the network architecture modelling simulation environment AnyLogic was used. The model graphic representation is pictured in Fig. 5. As the simulation of network latency in a decentralized network was the main

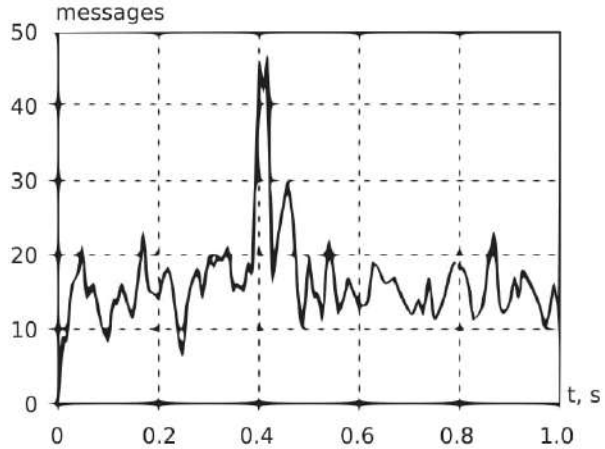

Fig. 6. Starting the system

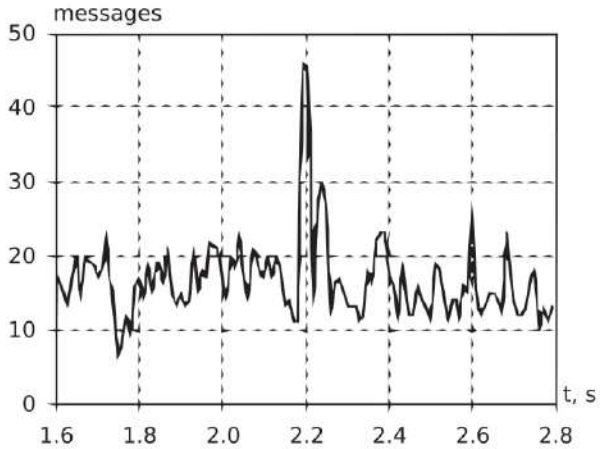

Fig. 7. Connecting the first client

purpose it is represented as entity flow with delays in the nodes. Therefore, discrete event simulation was the most appropriate. What is more, Anylogic allows to combine several approaches in one simulation, but in this case, discrete event approach was sufficient.

A number of experiments with various network parameters was carried out. Not surprisingly, the biggest network load takes place during control system initialization (turning the machine on). Initially, all modules statuses are polled and sensors data is transmitted. Experiments showed that main traffic is generated by two cameras (MVM and SLH), and a BLLC as well, which transmits all its current parameters (state and coordinates) all the time during the processing.

The network load can be increased when a number of users interacts with the system. Dealing with classical industrial management systems, it is hard to produce a case when more than one operator works with the apparatus. However, when equipment included in the integrated CPPS, this pattern changes. An operator as well as a dispatcher or technologist can work with the apparatus. It should be kept in mind, that the SCADA system can work as a client with a number of connections.

In Fig. 6 the diagram of the relation between time and network message amount is pictured. In this diagram, the first second of the control system activity is presented. From the diagram it is apparent that for 400 ms the number of messages in the network is smallest, then a short-duration peak occurs. This peak demonstrates that at this moment all modules have completed the self-checking procedures and the negotiation process is started. Upon completion, the network turned to the standby mode waiting for the client connections.

In Fig. 7 the process of connecting the first client is pictured. Obviously, a number of messages related to the webserver, which transmits to client static files, containing the user interface description. In Fig. 8 the process of reference point search is pictured. It can be noted that the MVM camera generates traffic during the whole workpiece reference point search. Fig. 9 demonstrates a change of the network loading during client connection to the camera, which translates the video stream from the working zone. Video traffic is transmitted from the camera and through the UIM module is delivered to users. The load on the network is not reduced as long as at least one user receives the signal from the camera.

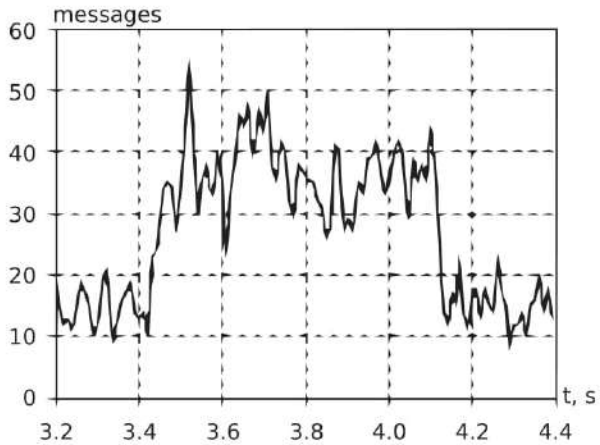

Fig. 8. Searching of the reference point

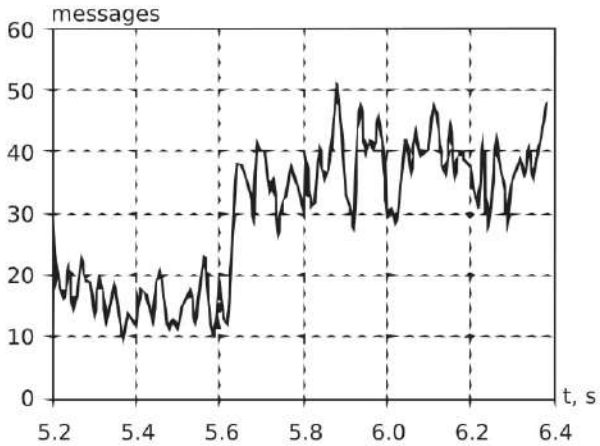

Fig. 9. Video stream from the working zone

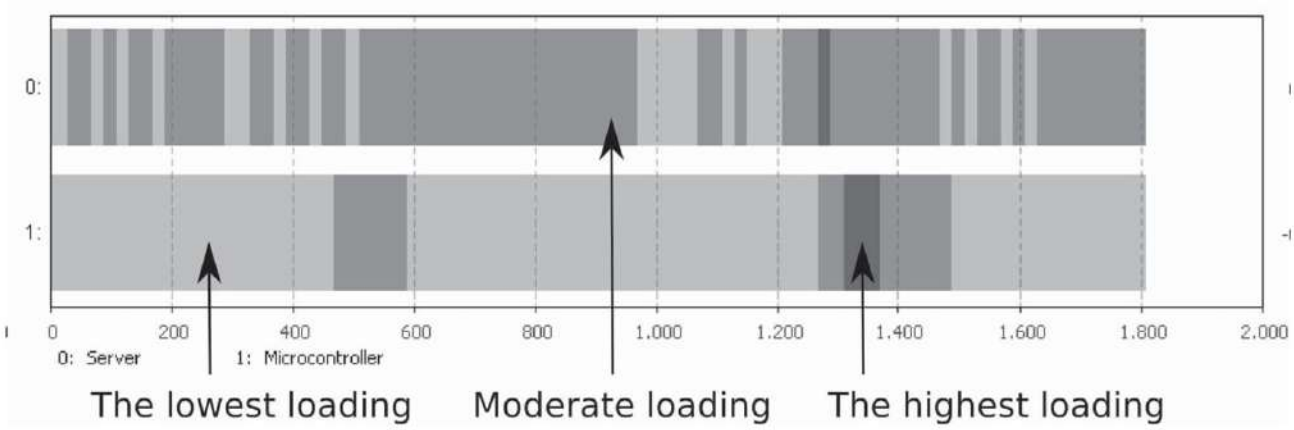

Fig. 10. Server and microcontroller loading

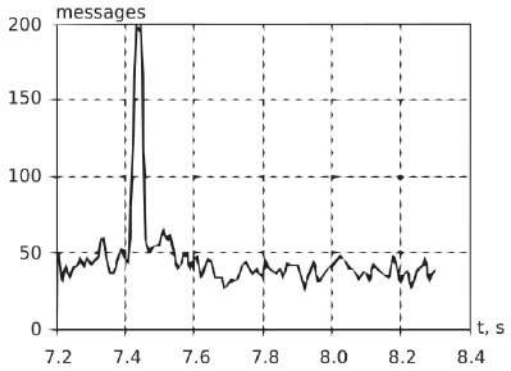

Fig. 11. Network maximum loading

TABLE I. CPU LOAD AS A FUNCTION OF A NUMBER OF CLIENTS

Number of clients	CPU load, %	
	Server node	Microcontroller node
1	64.5	13.6
2	68.5	15.2
3	71.6	30.8
4	69.1	39.8
5	74.9	43.7
6	73.0	50.3
7	74.9	50.8
8	75.7	56.5
9	80.2	57.3
10	83.3	59.8

Fig. 10 shows a time colour chart. There are two bars that represent server and microcontroller nodes and their states per time unit. The chart describes the BLLC CPU load and also Raspberry Pi microcomputer load, on which CM, UIM, and CPPM are installed.A network load test was carried out as well. We increased the number of clients to determine the maximum loads for the network, microcomputers CPU, and BLLC. The testing proved, that despite the increasing the number of clients, only a short peak of the network messages occurs (Fig. 11), and afterwards it turns to standby mode, while the microcomputer load increases up to $\sim 85\,\%$ and decreases no more (Table I). Thus, the number of clients simultaneously connected to the network, should not exceed ten.

VIII. CONCLUSION

In this paper a new approach to design architecture of modular-type CNC-driven industrial equipment was reviewed. As the basis of the developing system use of a distributed communication network where each node is a microservice is proposed. Due to its interoperability, such a system can be easily integrated into a cyber-physical manufacturing environment. Analysis of the distributed software with microservice architecture as well as system components communication and databases was carried out. As the main communication protocol, the system providing data transmission by using the message queue was chosen.

Such an approach makes the modules as autonomous as possible and simplifies the scaling, also providing an opportunity of network load leveling and provides a fail-safe feature by guaranteed message delivery. Protocol software implementation is based on the nanomsg library; broadcast messages and peer-to-peer messaging are used.

Data is stored in the system by using the distributed NoSQL database, UnQLite. This database system combines the storage advantages of the type "key-value" and document-oriented databases. The system is created as a useful library and using the similar data storage interface that easily integrates by the nanomsg protocol.

Realization of the proposed protocol is exemplified by an apparatus for selective polymers curing on a free-form surface by laser emission. All system modules and their interoperation algorithms are meticulously described. The advantages of utilizing a distributed microservice architecture in comparison to ordinary monolithic solutions are shown.

Simulation and numerical modelling of the distributed network supplies communication of all apparatus modules were done. The experiment showed that the highest network load occurs during system initialization as well as searching the reference point, but that these peaks are of temporary nature and do not significantly influence the modules during the polymer curing process. General modules load-testing was also carried out and shows that the number of clients connected to the system via HTTP protocol should not exceed ten.

REFERENCES

[1] S. N. Grigoriev and G. M. Martinov, "Research and development of a cross-platform CNC kernel for multi-axis machine tool," *Procedia CIRP*, vol. 14, pp. 517–522, 2014, 6th CIRP International Conference on High Performance Cutting, HPC2014.

[2] "A research on open CNC system based on architecture/component software reuse technology," *Computers in Industry*, vol. 55, no. 1, pp. 73–85, 2004.

[3] X. Ma, Z. Han, Y. Wang, and H. Fu, "Development of a PC-based open architecture software-CNC system," *Chinese Journal of Aeronautics*, vol. 20, no. 3, pp. 272–281, 2007.

[4] "Open-architecture system based on a reconfigurable hardware/software multi-agent platform for CNC machines," *Journal of Systems Architecture*, vol. 56, no. 9, pp. 407–418, 2010.

[5] N. Verba, K.-M. Chao, A. James, D. Goldsmith, X. Fei, and S.-D. Stan, "Platform as a service gateway for the Fog of Things," *Advanced Engineering Informatics*, 2016, to be published.

[6] C. Prazeres and M. Serrano, "SOFT-IoT: Self-organizing Fog of Things," in *2016 30th International Conference on Advanced Information Networking and Applications Workshops (WAINA)*, March 2016, pp. 803–808.

[7] "Chapter 2—service-oriented architecture and web services," in *Service Computing*, Z. Wu, S. Deng, and J. Wu, Eds. Boston: Academic Press, 2015, pp. 17–42.

[8] M. Rafighi, Y. Farjami, and N. Modiri, "Studying the deficiencies and problems of different architecture in developing distributed systems and analyze the existing solution," in *2015 2nd International Conference on Knowledge-Based Engineering and Innovation (KBEI)*, Nov 2015, pp. 826–834.

[9] A. W. Colombo, S. Karnouskos, J. M. Mendes, and P. Leitao, "Chapter 4—industrial agents in the era of service-oriented architectures and cloud-based industrial infrastructures," in *Industrial Agents*, P. Leitão and S. Karnouskos, Eds. Boston: Morgan Kaufmann, 2015, pp. 67–87.

[10] R. M. da Silva, F. Junqueira, D. J. S. Filho, and P. E. Miyagi, "Control architecture and design method of reconfigurable manufacturing systems," *Control Engineering Practice*, vol. 49, pp. 87–100, 2016.

[11] T. Vresk and I. Čavrak, "Architecture of an interoperable IoT platform based on microservices," in *2016 39th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, May 2016, pp. 1196–1201.

[12] B. Butzin, F. Golatowski, and D. Timmermann, "Microservices approach for the Internet of Things," in *2016 IEEE 21st International Conference on Emerging Technologies and Factory Automation (ETFA)*, Sept 2016, pp. 1–6.

[13] A. E. Rheddane, N. D. Palma, A. Tchana, and D. Hagimont, "Elastic message queues," in *2014 IEEE 7th International Conference on Cloud Computing*, June 2014, pp. 17–23.

[14] F. Reid, "Chapter 15—message queues," in *Network programming in .NET*, F. Reid, Ed. Burlington: Digital Press, 2004, pp. 419 – 452.

[15] N. Estrada and H. Astudillo, "Comparing scalability of message queue system: ZeroMQ vs RabbitMQ," in *2015 Latin American Computing Conference (CLEI)*, Oct 2015, pp. 1–6.

[16] V. C. Barroso, U. Fuchs, and A. Wegrzynek, "Benchmarking message queue libraries and network technologies to transport large data volume in the ALICE O system," in *2016 IEEE-NPSS Real Time Conference (RT)*, June 2016, pp. 1–5.

[17] V. M. Ionescu, "The analysis of the performance of RabbitMQ and ActiveMQ," in *2015 14th RoEduNet International Conference - Networking in Education and Research (RoEduNet NER)*, Sept 2015, pp. 132–137.

[18] K. Vandikas and V. Tsiatsis, "Performance evaluation of an IoT platform," in *2014 Eighth International Conference on Next Generation Mobile Apps, Services and Technologies*, Sept 2014, pp. 141–146.

[19] K. Maeda, "Performance evaluation of object serialization libraries in XML, JSON and binary formats," in *Digital Information and Communication Technology and it's Applications (DICTAP), 2012 Second International Conference on*, May 2012, pp. 177–182.

[20] S. Popić, D. Pezer, B. Mrazovac, and N. Teslić, "Performance evaluation of using protocol buffers in the Internet of Things communication," in *2016 International Conference on Smart Systems and Technologies (SST)*, Oct 2016, pp. 261–265.

[21] G. W. Poerwawinata and A. I. Kistijantoro, "Memcachedb persistent implementation using leveldb," in *2015 International Conference on Electrical Engineering and Informatics (ICEEI)*, Aug 2015, pp. 283–287.

[22] D. Stjepanovic, M. Savic, J. Jokić, and S. Marić, "Performance measurements of some aspects of multi-threaded access to key-value stores," in *2015 23rd Telecommunications Forum Telfor (TELFOR)*, Nov 2015, pp. 831–834.

[23] S. Wen, H. Zhang, and Y. Zhang, "A construction and implementation of high performance embedded database," in *2011 International Conference on Electrical and Control Engineering*, Sept 2011, pp. 4238–4241.

A Design of Cyber-physical Production System Prototype Based on an Ethereum Private Network

Maxim Ya. Afanasev, Anastasiya A. Krylova, Sergey A. Shorokhov, Yuri V. Fedosov, Anastasiia S. Sidorenko
School of Computer Technologies and Control
ITMO University
St. Petersburg, Russia
amax@niuitmo.ru, {ananasn94, stratumxspb}@gmail.com, {yf01, olesiia722}@yandex.ru

***Abstract*—The concept of cyber-physical production systems is highly discussed amongst researchers and industry experts, however, the implementation options for these systems rely mainly on obsolete technologies. Despite the fact that the blockchain is most often associated with cryptocurrency, it is fundamentally wrong to deny the universality of this technology and the prospects for its application in other industries. For example, in the insurance sector or in a number of identity verification services. This article discusses the deployment of the CPPS backbone network based on the Ethereum private blockchain system. The structure of the network is described as well as its interaction with the help of smart contracts, based on the consumption of cryptocurrency for various operations.**

I. Introduction

Blockchain is now one of the most actively discussed topics in various fields of human activity. Blockchain became widely known when the Bitcoin cryptocurrency first appeared in 2009. However, specialists started to reflect on the application of this technology not so long ago. Blockchain is a technology for storing and processing data in distributed computer networks and does not relate to any specific domain. All the blocks in blockchain can contain any data, which allows one to think about its use in production.

Development of the cyber-physical production systems (CPPS) and Industrial Internet of Things (IIoT) raises many problems related to the manipulation of data, such as storage, access, security, etc., that need to be solved. Moreover, there is currently a tendency to create distributed systems instead of centralized ones. One of the important properties of the Industrial Internet is the autonomy of its nodes and their ability to interact with each other. Such interaction is often based on the concept of service when certain nodes provide services to other nodes of the network. To ensure similar interaction, some implementations of a blockchain have a special mechanisms—smart contracts.

A smart contract is a self-executing scenario, which is stored in the blockchain as well as other data. Each smart contract has a certain algorithm implemented in a special programming language, which makes it possible to perform any actions automatically without involving third parties. Smart contracts monitor the fulfillment of certain conditions and makes decisions based on them in accordance with the specified algorithm. Since any network participant can sign the contract, interaction is applicable to smart things in the IoT. This approach can provide a trusted environment for interaction in the network and make the service providing mechanism transparent and unified. Furthermore, there is no need to create a separate register of services, since all contracts are already stored in the blockchain.

To date, there are many different implementations of smart contracts, but one of the first widespread implementations put into practice was the implementation of smart contracts in the peer to peer payment system, Ethereum in 2013. This implementation is currently the most advanced and has many features, like private blockchains, different protocols of consensus, etc. However, there have been other attempts such as Hyperledger [1].

These developments call for a more detailed study of smart contracts application in the Industrial Internet of Things. Additionally, it is necessary to consider not only theoretical features but also to carry out practical experiments to create a prototype.

The remainder of this paper is structured as follows. Section II gives the overview of studies dedicated to blockchain application in the context of IIoT and CPPS. Section III describes the possible role and place of blockchain in CPPS and IIoT. In Section IV, the architecture of the backbone network of the CPPS is proposed and, a created prototype of such a network is provided. The main limitations and drawbacks of the proposed approach are discussed in Section V. Finally, Section VI presents conclusions and suggests some future areas of research and development.

II. Related work

Despite the fact that the topic is relatively new, researchers and specialists have already published quite a large number of articles pertaining to the scope of the blockchain and production domain as well as adjacent production areas.

The discussion has mostly touched on the topic of the smart contracts. In the article [2], Nikolay Teslya considers the use of blockchain in IIoT. The article proposes the architecture of the Internet of Things, which is based on the Smart-M3 information sharing platform developed by them and the Hyperledger blockchain platform with smart contracts [3]. The authors note the drawback—restriction in the complex search

for information on the blockchain. As a solution, they created an additional information layer that performs search functions, data representation, additional checks, etc.

One of the neighboring areas to production domains are the supply chains. The place that blockchain can take in this sphere is discussed in the article [4]. The authors note that currently there is a lack of transparency in supply chains. Due to its transparency and immutability, as well as the mechanism of smart contracts, blockchain can help to automatically identify and record events in supply chains. A good example is in the article [5] where a cargo transition between points from the place of production to the desired destination is considered. Using radio tags and the smart contracts registration and tracking of the cargo are carried out automatically. Similar ideas are suggested by the authors of [6]. They propose the architecture of an information platform for the creation of the food supply chain traceability systems.

There are a huge amounts of blockchain implementations, which may be suitable or unsuitable for the IIoT. The author of [7] compared distributed ledgers by the most important characteristics for IIoT, such as the presence of smart contracts, the transaction time and the consensus protocol. In the comparison, the three most well-known distributed ledgers—IOTA, Ethereum, and Hyperledger—are involved. There are also publications where not only a qualitative but also a quantitative assessment of private blockchains is carried out [8], [9]. The authors of [9] created a special framework that allows you to analyze through special API private blockchains based on Ethereum, Parity and Hyperledger Fabric. The tool is open-sourced and can be really useful for blockchain analysis.

III. Place and role of blockchain in the structure of CPPS and IIoT

Blockchain technology is quite universal. To date, there are a significant number of implementations of blockchain used in various areas of human activity. In order to effectively use all the advantages of blockchain technology for building CPPS and IIoT, it is necessary to develop a network structure for the blockchain system that is optimal from the point of view of the solved tasks, and to choose the most suitable tools (software and hardware).

Many manufacturers of equipment, both in private conversations and publicly, have declared that they are engaged in the development of components and technologies for the industrial Internet of Things (and, consequently, for cyber-physical production systems) for many years. In their opinion, the only thing that has changed recently is the name of such systems. The essence has stayed unchanged since both the Industrial Internet of Things and cyber-physical production systems are nothing more than a set of industrial controllers, networked and communicating in one of the most common industrial protocols (Modbus, Profibus, CANbus, etc).

To confirm or disprove this statement, it is necessary to refer to the definitions of cyber-physical systems and the Internet of Things. Initially, the concept of "cyber-physical system" will be considered. In accordance with the standard NIST Special Publication 1500-201, "Cyber-physical systems are smart systems that include engineered interacting networks of physical and computational components." Such a definition itself makes it quite clear that the industrial cyber-physical system differs significantly from the usual industrial network of controllers.

Let's consider this issue in more detail. According to the definition, the main distinguishing feature of cyber-physical systems is the almost full "transparency" of the connections between the physical and logical components of the system. In the cyber-physical system, the distinction between real and virtual components is erased. In contrast, in traditional industrial networks the main focus is on physical devices—programmable logic controllers (PLCs). In accordance with this concept, only PLCs perform calculations in the network and are also responsible for the proper execution of the technological process and the information interaction of industrial equipment (both with each other and with the external environment). With that, the concepts of "information flow" and "material flow" are clearly separated.

It was the PLC that made it possible to create the first automated and automatic production lines, and it was the development and improvement of these basic nodes of industrial networks that focused the main efforts of manufacturers of such intelligent industrial equipment. Protocols and data transmitted over these protocols received much less attention as each manufacturer sought to create its own industrial protocol, which was incompatible with the protocols of other manufacturers.

In addition, we can add incomplete correspondence of the industrial protocols to the requirements for modern information and telecommunication technologies. For example, the widespread industrial protocol Modbus, despite its openness, uses approaches from the 1970s that manufacturers are trying to improve by using modern transport protocols, in particular TCP, which seems an extremely irrational approach. Also, it should be noted that the concept of "cyber security" in principle is not applicable to many of the currently used industry standards, again because of their strong lag behind modern information and telecommunications technologies.

For example, one implementation of the Modbus protocol—Modbus RTU—uses a physical connection based on the RS-485 standard, which is a slightly modernized UART. These protocols are among the first digital protocols for data transmission, and at the time of their widespread distribution there was no concept of cyber security. Implementing Modbus TCP simply allows you to put Modbus RTU messages in TCP/IP packets. It should be noted that Modbus is not only a protocol for receiving data from some passive devices (sensors, for example), but also a control protocol.

Consequently, this suggests that despite the external similarity, the industrial networks of the PLC (as well as the process control systems based on them) are not industrial cyber-physical systems. Let's turn to the comparison of industrial networks with the Industrial Internet of Things. Lets start with the general definition of the Internet of Things. Unfortunately, we could not find any industrial or other standards, in the list of terms of which would be a present definition of the concept of "Internet of Things". However, on the basis of the analyzed literature it is possible to give the following general formulation of the term: "The Internet of Things is a computer network consisting of physical devices equipped

with built-in electronics (and software), sensors, actuators and communication means that allow these devices to communicate with each other and to exchange data."

Such a definition seems to be similar to the definition of the concept of "cyber-physical system", but there is one significant difference between them. The Internet of Things is not designed to reach a certain common goal, that is, it is not a system in its classical meaning. According to the Oxford English Dictionary, a "System is a set of things working together as parts of a mechanism or an interconnecting network; a complex whole."

It should be noted that the cyber-physical system is holonic [10], that is, consisting of a set of "physical entities" and their "digital twins" connected together. "Digital twins" is a computational model of "physical essence", that is, it reproduces the behavior of a physical machine and gives an idea of how this machine reacts when various external influences occur. The connection can be ensured using sensors and actuators.

The holonic nature of cyber-physical systems implies the existence of both hierarchical and heterarchical links, which means that they can be combined into higher-order temporal entities (in turn, again performing one common function), that is, a "System-of-systems" [11]. At the same time, all communications are carried out via common and open to all participants network, which is the Internet of Things or the Industrial Internet of Things. Thus, the cyber-physical system forms the first level, and the Internet of Things is the second level of vertical digital integration.

It is clear that the standards of industrial networks based on PLCs, even on a set of properties, cannot be considered as either cyber-physical industrial systems or the Industrial Internet of Things. The last statement is due to the fact that such networks are not protocol-oriented or service-oriented solutions. They have a clear separation of material and information flows, do not have analogs of "digital twins" in their composition, and also have a rigid hierarchical control system with a single center, not assuming that there are heterarchical links and opportunities for flexible restructuring.

It should be noted that the problem of obsolescence of industrial protocols appeared a long time ago. In 1996, OPC Foundation developed the OPC (Open Platform Communications) family of protocols, which were intended to become the main technology, providing a single interface for managing automation objects and technological processes. The basic principle of this standard is "Open Communications on Open Protocols." However, for the most part, this protocol was based on Microsoft technologies and therefore was not widely used.

The logical evolution of the OPC protocol, called the OPC UA (OPC Unified Architecture) is more useful. This specification defines the transmission of data in industrial networks and the interaction of devices in them. The main advantages of this approach are openness, independence from any particular operating system or technology, scalability, and the implementation of its own system for ensuring cyber-security. The unified architecture of OPC is a service-oriented architecture (SOA) and is based on various logical levels.

However, like any other specification, striving to become an industry standard, the OPC UA is not without certain drawbacks. First, the OPC UA is a fairly complex protocol that attempts to accommodate a wide range of areas of industrial automation. Secondly, OPC UA is also focused on industrial PLCs. In general, this is another protocol for data transfer between industrial devices (machine-to-machine communication protocol), only based on TCP/IP and XML technologies. Nevertheless, OPC UA is the most elaborate protocol today and it was decided to make it the basis of the developed CPPS architecture.

Therefore, having examined in detail the concepts of IIoT and CPPS, we now consider the blockchain technology position in their structure. The first thing that needs to be addressed is the fact that there are two main subclasses of blockchain: global and private. The first are the most developed and are mainly used for solving global problems such as organizing international peer-to-peer payment systems or crowdfunding.

Global peer-to-peer networks are highly stable because of the large number of participants but they are not suitable for creating industrial networks similar to those described earlier. At the same time, the main limitation is the rigid linking of all data exchange operations to the cryptocurrency used in a global cryptocurrency blockchain. In other words, with this technology, any interaction between CPPS nodes will have a certain cost, and it will be directly related to the price of the cryptocurrency used. It is almost impossible to predict the change in the exchange quotations on the market of cryptocurrencies, which makes the predictable cost of ownership of the projected CPPS difficult. Consequently, the basis for the projected architecture will be a *private blockchain*.

Next, it is necessary to determine the main functions of blockchain libraries in the structure of the CPPS, as well as determine the main tools with which these functions can be implemented. The main functions of blockchain in the CPPS should be:

1) Organization of a common information space for the machine-to-machine interaction within CPPS.
2) Ensuring CPPS cyber-security.
3) Ensuring an easy scaling and CPPS restructuration.
4) Ensuring of redundancy of equipment and communication channels.
5) Ensuring common data storage.
6) Implementation of "digital twins" technology through the use of smart contracts.
7) Ensuring implementation of common tasks for CPPS through the use of smart contracts.

The requirements quite severely limit the choice of tools since the implementation of key functions of CPPS implies the possibility of working with smart contracts. At the moment, smart contracts are implemented in several public blockchain distributed computing platform. However, only Ethereum has a well-designed and debugged virtual machine (Ethereum Virtual Machine, EVM) that implements its own Turing-complete programming language, as well as the ability to work with many modern high-level programming languages by compiling them into the EVM bytecode. Thus, the private Ethereum platform will be used as the main platform for building the

described CPPS, and the Solidity programming language as a tool for the implementation of smart contracts.

IV. Proposed Architecture

A. Backbone network of the CPPS

As noted in the previous section, the CPPS backbone network will be built on the basis of a private blockchain of Ethereum. The following types of nodes should be included in the network (Fig. 1):

Full nodes—download and store all the network blocks in the internal memory, and are responsible for checking the validity of the blocks and their compliance with the rules of the Ethereum consensus protocol. They can act as miners of new blocks and it is assumed that general purpose PCs running Windows or Linux will be used as full nodes. For example, it can be client machines of CPPS, SCADA server, database server, etc.

The presence of full nodes is extremely important for the full functioning of the network since the stability and uniformity of generating blocks directly depends on the number of nodes. The functions associated with block validation and consensus-building do not require large expenditures of computing resources and RAM. The main requirement is that there is enough disk space to store the full chain of blocks.

Modern general-purpose PCs have a sufficient amount of internal memory, the amount of which is rarely less than 500 GB. At the same time, the volume of the entire Ethereum global network is about 385 GB, while 20–30 GB of free disk space is enough for maintaining a full node. Obviously, for a private network, such numbers are unattainable.

It is also possible to use the resources of general purpose PCs during their idle time for mining new blocks. In this case, the task responsible for mining can be started in the background mode. However, questions of the analysis of congestion

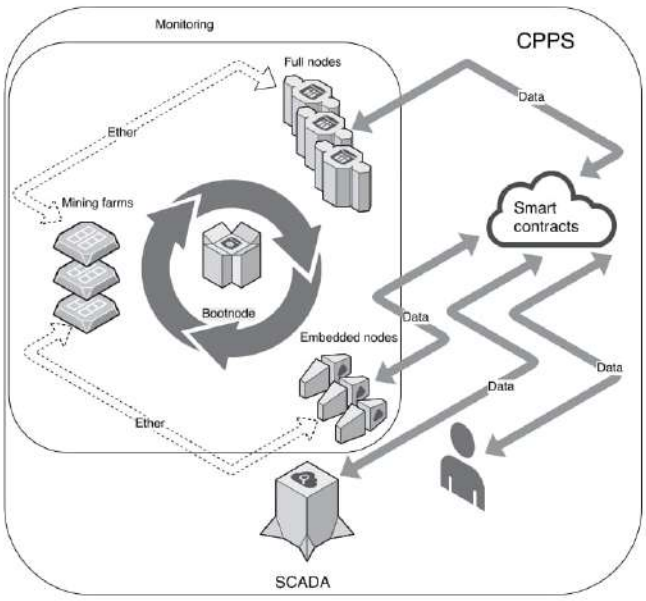

Fig. 1. Blockchain network architecture

and energy consumption of general purpose PCs operating in this mode require additional research. So, at this stage, special mining farms have been created as such farms can act the same general purpose PCs that are working constantly and perform only the mining task.

It should be noted that in a private network there is the possibility of managing the complexity of mining so the mining farms need not be high-performance, have specialized GPUs or memory with increased bandwidth. Experiments have shown that the minimum number of mining farms is 3. The network of the blockchain system operates even if there is only one mining node, however, to ensure greater reliability and redundancy, the number of miners should be larger.

For the architecture under consideration, there is no need to try to optimize the process of searching for new blocks, for example, by running multithreaded mining, it is much more important to have a sufficient number of "slow" miners. This allows for having greater stability in the appearance of new blocks (generating blocks too fast leads to conflicts and branching), as well as the stability of the system due to redundancy. Also important is that, in the absence of miners in the network, the generation of new transactions will not stop but transactions will stop receiving confirmation, and therefore spreading across the network. In addition to all the foregoing, the mining farms will be charged with the distribution of the Ether between the nodes as we discuss in more detail below.

Bootnodes—the special nodes that perform the initial network boot. Each new node that connects to the network first accesses this node to get a network peers list. Without this node, it is necessary to implement the manual procedure of adding peers. It consumes very few resources, so it can be installed on a microcomputer, for example, a Raspberry Pi. However, for more stability and network redundancy, one needs to install several more bootnodes in parallel with other services.

Nodestat and monitoring node—a special node that monitors the network, allows for obtaining statistics: number of nodes, their load, the frequency of new blocks, the number of blocks, transactions, the performance of miners, the difficulty of mining new blocks, etc. Availability and working capacity of this node does not affect the network as a whole, so this node can be either dedicated or installed as a service to any suitable network node.

Embedded nodes—nodes, working on a light client protocol, specially designed for embedded smart property environments. In the considering architecture, they are the main type of nodes connected directly to equipment, controllers, sensors, etc. Although full security is only possible for a complete node, the light client protocol allows the light nodes to receive about 1 kB of data every 2 min from the network. These data allow it to perform a partial check on the state of the network in the part that interests them, as well as monitor compliance with the consensus.

B. Hardware and software

To implement the network of test blockchain units the following types of equipment were used:

General purpose PC—a low-performance computer used to solve a wide range of CPPS tasks. Specifications: Intel Core i5 CPU 2.4 GHz, 4 GB RAM, AR9285 Wireless Network Adapter 802.11 b/g/n, Yukon Optima 88E8059 Gigabit Ethernet. Operating system: Ubuntu 16.10, GNU/Linux 4.8.0-41 x86-64.

Laptop—a portable computer that can be used by the CPPS operator as a terminal, or for remote management. Specifications: MacBook Pro, Intel Core i5 CPU 2.6 GHz, 8 GB RAM, Gigabit Network. Operating system: OS X 10.10.5, Darwin 14.5.0.

Server—a high-performance computer used for the most resource-intensive software components of CPPS. Specifications: 2x Intel Xeon E5620 2.4 GHz, 32 GB RAM, Intel 82575EB Gigabit Network. Operating system: Ubuntu 16.04.2 LTS, GNU/Linux 4.4.0-64 x86-64.

Embedded system—system-on-chip, used to implement low-level algorithms in real time. Specifications: Amlogic S905 Quad Core Cortex-A53 1.5 GHz 64bit ARMv8 CPU with Mali-450 GPU, 2 GB, Realtek RTL8211F Gigabit Network. Operating system: Ubuntu 16.04 LTS.

Virtual private server (VPS)—a cloud-based virtual machine used to optimize the processing power of CPPS by transferring part of the software components to dedicated servers. Specifications: Intel Xeon CPU E5645 2.4 GHz, 512 MB RAM. Operating system: Ubuntu 16.04.2 LTS, GNU/Linux 2.6.32-042stab120.18 x86-64.

As the Ethereum client, the console client "geth" version 1.8.1 [12] was used. With its help, all complete nodes, mining farms, a monitoring node and bootnodes were implemented. The GUI of the monitoring system was implemented on the platform "eth-netstats" version 0.0.1 [13]. To test the network, as well as the deployment of smart contracts, the browser "Mist" (version 0.9.3) was used [14]. IDE "Remix" [15] was used for writing and debugging smart contracts for "Solidity" [16]. For the initial network configuration, the "puppet" manager, which is part of the "geth" distribution, was used.

For embedded systems on the ARM platform, the "EthEmbedded" client [17] was used-a Ubuntu OS special assembly for Raspberry Pi, Odroid microcomputers, etc. It should be noted that the implementation of the Ethereum client on the basis of PLC is possible. However, to date, the authors have not found a single mention of such projects.

C. Network deployment

The kernel of the private Ethereum network was deployed on the basis of the local network of the laboratory CPPS of Instrumentation Technologies Department. Also, several nodes located in the global network were connected to the network (several territorially separated cloud VPSs were used). At the first stage of the deployment of a private Ethereum network, bootstrap nodes were created. As the main bootnode, a Raspberry Pi microcomputer was used, which has a permanent connection to the local network and Internet access. To configure a bootnode, the key must be generated with the command:

```
bootnode -genkey nodekeyfile
```

After that, the bootnode can be started by:

```
bootnode -nodekey nodekeyfile
```

Then the bootnode address with a form like `enode://<node-name>@[::]:30301` will be displayed. In the future, this address will be used to configure other nodes, and the expression in square brackets is replaced by the IP address of the bootnode. To organize access from the external network, the address was additionally translated to the "white" address, as well as port forwarding 30301.

Next, the genesis block is set up the zero block of the blockchain. All nodes belonging to the same private network must have the same genesis. Genesis is a plain text file in JSON format. The file contents are as follows (Listing 1):

```
{
    "config": {
        "chainId": 73655,
        "homesteadBlock": 0,
        "eip155Block": 0,
        "eip158Block": 0
    },
    "alloc"      : {},
    "coinbase"   : "0x0...0",
    "difficulty" : "0x400",
    "extraData"  : "",
    "gasLimit"   : "0x2fefd8",
    "nonce"      : "0x0000000000000076",
    "mixhash"    : "0x0...0",
    "parentHash" : "0x0...0",
    "timestamp"  : "0x00"
}
```

Listing 1. Genesis file

The most important parameters in this file are:

`chainId` is a random number that is a unique identifier of the chain of blocks. It is necessary to prevent relay attacks when an attempt is made to record and then reproduce previously sent correct messages or their parts.

`nonce` is a "number that can only be used once", a one-time code selected in a random or pseudo-random manner, which is used for secure data transmission, and is also used to prevent relay attacks.

`difficulty` is a scalar value corresponding to the difficulty level of finding a new block. It determines the purpose of the mining, which can be calculated based on the complexity and time of generation of the previous block. The higher the complexity, the more calculations must be made by the miner to discover a new valid block. This value is used to control the block generation time in the blockchain, keeping the block generation frequency in a predetermined interval. For a private network, this value should be small in order to minimize the load on the mining farms.

`gasLimit` is a coefficient that is used to calculate the final cost of a transaction. At the initial stage, this value should be large enough so that smart contracts can be tested without hindrance. In the future, it can be changed.

After the genesis of the block is created, the first full node on the network is initialized with the following command:

```
geth
    --datadir .
    --bootnodes
        "enode://<name>@<ip>:30301, ..."
    --networkid 10777
init genesis.json
```

As one can see from the presented record, the start flags followed after the command, and then the client mode. If the mode is not specified, the client runs in background mode and waits for connections. The `--datadir` flag specifies the full path to the directory where the blockchain will be stored, the `--bootnodes` flag specifies the addresses of the bootstrap nodes, and the `--networkid` flag specifies the unique identifier of the private network, which must not coincide with the known global block IDs. The init mode starts the procedure for initializing the block library structure in the directory specified by the `--datadir` flag according to the genesis of the block specified in the `genesis.json` file, which should be located in the same directory.

To start the monitoring node, it is necessary to specify the `--rpc` flag, which allows access to the geth client API via the JSON-RPC protocol (port 8545). Geth uses console (accesses the node and the console at the same time) and attach (connects to the already-running node) modes to access the geth client console. At least one account must be created at each node through which it can receive funds from other sites and access smart contracts.

Mining is started by the command:

```
geth --mine --minerthreads=1
```

or from the console using the command:

```
miner.start(1)
```

As mentioned above, multithreaded mining is not applied to a private network, because it does not give any advantages, but only increases the load on the mining farms.

Initialization of embedded nodes does not differ from the initialization of the full nodes. The launch of embedded node occurs using the same command, only with the additional flag `--light`. The rest of the nodes are configured in the same way. It is important that, for the correct operation of the private network, all nodes must be synchronized in time.

The "Mist" browser is installed only on the nodes where it will be used. The setup of the statistics collection and monitoring node is well described in the documentation and is of no interest since it is reduced to a simple installation of the "Node.js" framework and deploying application.

D. Smart contracts

The first thing that is needed to pay attention to when considering smart contracts is the lack of the possibility of changing the smart contract after deployment. This is due to the special abilities of blockchain technology—it is only allowed to add data to the blockchain as removing something from the database, without violating the structure of the whole chain of blocks, is denied. Therefore, all contracts are deployed by the CPPS operators as necessary, and each contract has a self-destruct function, which can be started only from the address of the operator that carried out the deployment. Obviously, this creates a certain vulnerability, but it will not be possible to block a function that disrupts the operation of all CPPS without this function, and this is much more dangerous for its operability.

At the current stage of implementation of the CPPS under consideration, smart contracts are used to solve the following tasks:

- Implementing the "digital twins" of CPPS' physical components
- Provide failsafe and redundancy of CPPS
- Providing a persistent store of process data
- Redistribution of funds between nodes.

The last task is the first thing that is needed to be discussed. Since the developed CPPS uses the Proof-of-Work consensus protocol, network nodes require a cryptocurrency (Ether), which they can convert into so-called "gas", which is necessary for providing work in the network. In the Ethereum blockchain, "gas" is used to execute smart contracts, is consumed during the passage of transactions, and when data is stored in the blockchain.

As previously mentioned, mining on embedded nodes does not seem to be advisable because of their limited computing abilities. Because of this, there are mining farms in the network, which constantly generate new blocks, for which, according to the Proof-of-Work protocol, they get a reward—Ether. This reward is distributed amongst the CPPS nodes according to the following principle: all the nodes of the CPPS are ranked by their importance. The significance is determined by the influence of the node on the technological process realized by the CPPS.

For example, machine tools and industrial robots are considered "rich" customers and they are allocated the maximum amount of cryptocurrency, which gives them instant and unhindered access to all CPPS' resources. Also the "middle class" is highlighted—CPPS' components related to the planning and logistics of the production process such as automated ground vehicles. Units that are not directly related to the production process but are necessary for general observing and monitoring are considered "poor", and they are given limited access to CPPS resources. For example, temperature and humidity sensors in the workshop are "poor", because there is no need to receive data from them too often. Accordingly, a small amount of cryptocurrency allocated to such devices automatically will not allow them to use the resources of the network and distributed storage of CPPS too aggressively.

The Ether extracted by the mining farms is transferred to the address of the smart contract that carries out the redistribution procedure. The ranking algorithm determines the time intervals and automatically transfers funds from the mining farm wallet to the corresponding node.

The next task of smart contracts is the "digital twins" equipment implementation. At the current stage, it is possible to collect data from sensors used in CPPS. An example of such a contract with comments is presented in Listing 2.

```
pragma solidity ^0.4.18;

contract TempSensor {
    int256 temp = 0;
    address holder = msg.sender;
    event tempChanged(int256 temp);

    function
    setTemp(int256 newTemp)
    public {
        temp = newTemp;
        tempChanged(temp);
    }

    function
    getTemp()
    public constant returns (int256) {
        return temp;
    }

    function
    kill()
    public {
        if (msg.sender == holder) {
            selfdestruct(holder);
        }
    }
}
```

Listing 2. Smart contract for temperature sensor.

What is important is that this small and, at first sight, very simple smart contract also solves the task of creating a "digital twin". It is necessary to understand that the smart contract is a separate independent entity that is located in the blockchain. All data that the smart contract receives is automatically stored in the database of the blockchain, for example, technological process data, states, addresses of the senders of the transaction, transaction time, etc. That is, blockchain in general, and smart contracts in particular are excellent tools for collecting and analyzing statistics about the technological process, which is implemented in CPPS. At the same time, *these data cannot be changed or destroyed without completely removing all the nodes of the blockchain*. This behavior opens up new prospects for using blockchain technology such as to investigate emergencies, or the integration of CPPS on the basis of a blockchain system with a SCADA system for monitoring the technological process in real time. In addition, all the same contracts can provide both redundancy and fault tolerance of CPPS. Suppose that there are several identical temperature sensors in the workshop. Each of them is a smart thing, that is, it has built-in computing resources, a control program, and is connected to the blockchain network. The received data is sent to the blockchain unit via a function call of the smart contract, while the smart contract is a "digital-twin" of this array of sensors, thereby providing redundancy. The failure of one of the sensors will not disrupt the operation of the system. since the smart contract will continue to receive data.

The only problem that can arise when implementing such clusters of physical devices is their performance monitoring, because there may be a situation when all the sensors fail. But even this can be envisaged. It should be noted that the sensors are nodes of the block and receive a cryptocurrency from the mining farms. In accordance with the algorithm, they should spend it on access to the resources of the blockchain system, and if it turns out that one of the nodes ceases to create transactions and to work with smart contracts, it means, perhaps, this node has failed.

E. Test results

The network was tested for three weeks. The network had three mining farms and, in total, 161,000 blocks were found, with each miner receiving approximately 265,000 ETH at a gas price of 18 gwei. On average, a new block appeared on the network every 15.13 s. With a target value defined in the "geth" client code of 15 s. The time interval distribution corresponds to the classical Pareto distribution, which can be shown in Fig. 2. It should also be noted that in spite of the fact that the interval for the appearance of a new block in the blockchain is 15 seconds, the time of the block propagation through the network is 250 ms, that is, the process of distribution of new transactions has practically no effect on the performance of the network of blockchain. During the testing, more than 10,000 transactions were performed, related to the redistribution of the cryptocurrency and the implementation of smart contracts. The total database volume of the block database for the full node was 139 MB.

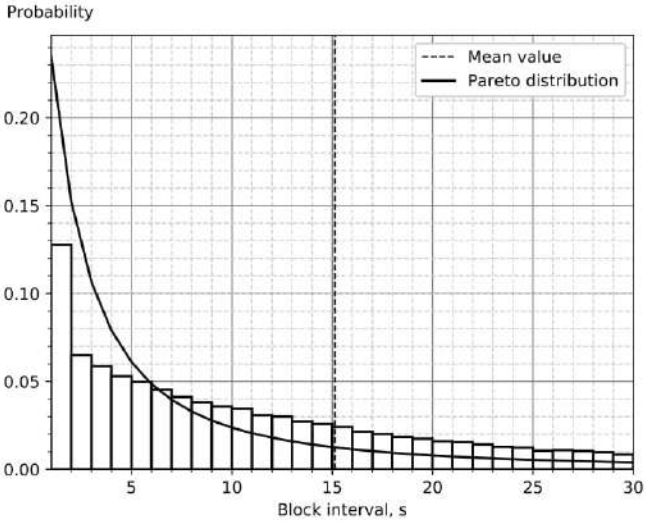

Fig. 2. The time interval distribution

V. Discussion

A. Performance issues

According to the authors, the main restriction of using blockchain to create CPPS for today is the low latency of the network, that is, the transaction does not pass until a new block appears. Since the generation of blocks depends on the complexity and complexity changes by the total network hashrate, the periodicity of the changes in each block, nevertheless, the average time for the appearance of a new block remains practically unchanged. This is especially noticeable if only one miner is working on the network.

Nevertheless, a cyber-physical production system based on a blockchain network can be referred to real-time systems, because it generally meets the requirements of such systems. Experiments have shown that it is able to react to events in an external environment, or to influence the environment within the required time constraints, and to process information for a certain finite period of time in order to maintain a constant and timely interaction with this environment.

Thus, such a system can at least fulfill the role of a real-time database, storing information about the technological process. At the same time from the point of view of data analysis, the frequency of collecting the readings from the sensors is not limited by the speed of appearance of the blocks in the network. In blockchain, the block is a storage unit, not a data transfer unit. Each block can have an unlimited number of transactions, each of which contains data from sensors arriving at the speed with which they are polled. Before the new block appears, all of this data is stored on the sending node, and in the block it is recorded and distributed over the network.

The uncontrolled frequency of the occurrence of blocks can make it difficult to transmit signals for logical control of equipment, that is, using smart contracts as "digital twins". However, according to the authors, such a control is still possible. It should be noted that the implementation of the algorithm for automatically modifying the complexity of block generation implies that the average frequency tends to some pre-programmed value. Testing the blockchain network showed that with a sufficient number of nodes, the average generation time of a new block stays unchanged within a few hundredths of a second, which is quite acceptable for many real-time production networks.

It should be noted that in any industrial networks there are delays in the transmission of control signals. These delays are caused by many factors, in particular obsolete physical layer protocols, which just are not able to provide an acceptable data transfer speed at large distances between network nodes. For example, the still used RS-485 protocol for a segment length of 1200 m provides a transfer rate of only 62.5 kbps. And this is only a physical level, delays can be on the channel, and even on the application layers.

Nevertheless, analyzing many years of experience in the use of industrial networks, it can be concluded that with proper networking, as well as the correct implementation of the management process, these delays do not affect the operation of industrial equipment. Also, it should not be forgotten that devices that generate emergency stop signals ("mushroom" head pushbuttons, automatic fuses, limit switches, etc.) are never connected to the industrial network by the safety rules, only directly to the equipment. Also, it must not be forgotten that the blockchain technology in general and the Ethereum network in particular are developing very rapidly. The latest versions of the "geth" client implement the consensus protocol Proof-of-Authority, which firstly allows one to almost completely abandon the mining, and secondly, it allows one to specify the exact interval for the appearance of a new block during the initial configuration of the block.

Also worth mentioning the Raiden technology [18] is a lightweight network within the global Ethereum network. It allows to execute transactions between predefined nodes without having to write to the main blockchain, while using all the advantages of the core network associated with security and anonymity. The main advantage of using a lightweight protocol is a reduction in commission and a shorter transaction time. Can be used for micropayments, in particular in applications of Internet of Things. As for mining, first, there is a clear tendency to abandon mining and the transition from the Proof-of-Work consensus model to Proof-of-Stake, and secondly, unused resources can be used for mining. In particular, the previously considered microcomputers such as Odroid or Raspberry Pi have built-in video processors that are not used to solve CPPS tasks, so they can be used for blocks mining. For example, GPU Raspberry Pi (Broadcom Videocore IV) supports OpenCL technology. The same applies to Mali video processors from T6xx version too (ODROID-XU4 microcomputer) and PowerVR SGX544 (microcomputer CubieBoard6). Support of this technology means that on these microcomputers it is possible to run the "ethminer" program—Ethereum miner with OpenCL, CUDA and stratum support, which allows to mine blocks with a hashrate comparable to the general purpose PCs, with only a much lower energy consumption.

B. *Security issues*

Undoubtedly, like any other open network, a blockchain may be subject to a hacker attack. The most dangerous type of attack for blockchains is a "51 % attack", in which the attacker can block all transactions on the network [19]. However, in order to carry out such an attack on private blockchain, the attacker will not only need a physical connection to the network, but also the availability of the genesis-block for registering the node, as well as the resource costs for mining. Of course, with low computational complexity, there is a potential threat of hacking the network by temporarily connecting a high-performance cluster or even an ordinary high-performance PC.

Thus an interesting situation arises. The fact is that in order to carry out any malicious activity in a private detachment, a hacker needs an ether. The only way to get it is mining, that is, the generation of new blocks. Thus, even at the moment of infiltration, an attacker is forced to perform work to ensure the operability of a private network. From this, we can also conclude that the denial-of-service attack is extremely difficult to implement, because any transaction on the network entails the need to pay a commission, so simply "flood" the network with transactions will not work.

The reason for using blockchain in corporate networks on the basis of which CPPS can be built is the need to ensure security and openness simultaneously. The classical approach used to ensure the security of geographically distributed networks is the organization of virtual private networks (VPNs). Geographically distributed VPN nodes are combined in some kind of local network, all traffic is encrypted and passes through a public network along the tunnel organized in it. This approach proved its reliability and effectiveness, but it has a number of significant drawbacks. First, truly reliable VPN solutions can be implemented only with the use of specialized software and hardware systems that have embedded encrypting tools. Deployment of such complexes is a non-trivial task requiring the participation of highly qualified

specialists, and the cost of such solutions is quite high. It is also quite a challenge to change any structure and topology of the network. Even a simple extension by adding a new node (for example, when organizing a new unit) requires the purchase and configuration of additional equipment. At the same time, the reliability of a VPN is determined by the reliability of its most vulnerable node, because hacking one segment of the network theoretically can compromise the entire network.

Blockchain allows to simplify work with global and territorially distributed networks, as well as the possibility of simple integration with cloud solutions, which can be very useful in the organization of CPPS. Blockchain allows you to work your sites in an unfriendly environment without the need to "fence off" from the global network with tunnels, firewalls, etc. A potential attacker can connect to the network while remaining unnoticed. However, this connection will give the attacker not so many opportunities to intercept data, as well as interference in the network. The structure of the blockchain system is such that it is only allowed to add data to the decentralized database, the attacker will never be able to delete or change the data already stored in the blockchain. At the same time, any actions of the attacker aimed at destabilizing the work of the blockchain network require an active connection, that is, they will also remain in the database. This will not allow you to "clean up tracks" to hide the fact of infiltration.

VI. Conclusion

To implement the CPPS core network, it is not enough to design a protocol for exchanging data between nodes and sensors; it is necessary to take into account network security, data access mechanisms, and performance of hardware components. Considering the fact that many implementations of CPPS on production lines rely on the technologies of the 70s and 80s, an important task is to find an alternative solution. The authors suggest that the actively developing technology of the blockchain network can serve as a platform for a distributed decentralized network, not least thanks to a special method of communication between the network nodes of the blockchain system—smart contracts.

Based on the tests conducted by the private Ethereum network as a CPPS backbone network, it can be concluded that this implementation is possible, but has a number of problems that need to be resolved. First, the time of block generation directly depends on the complexity of the network algorithms, and without the appearance of a new block, the transaction cannot be passed. Secondly, the susceptibility of the network of blockers to a number of hacker attacks, which can block the passage of transactions in the network. Finally, the frequency of the occurrence of blocks is uncontrollable, which can cause undesirable delays in the network, however, these delays are rather small in time and are not critical. Nevertheless, the Ethereum blockchain is a good alternative to the current CPPS technologie, and given that the blockchain technology is currently actively developing and being finalized, many of the shortcomings can be eliminated in future work.

VII. Acknowledgements

This work was carried out under project no. 617026 "Technologies of cyber-physical systems: management, computing, security."

References

[1] D. T. T. Anh, M. Zhang, B. C. Ooi, and G. Chen, "Untangling blockchain: A data processing view of blockchain systems," *IEEE Transactions on Knowledge and Data Engineering*, vol. PP, no. 99, pp. 1–20, 2018.

[2] N. Teslya and I. Ryabchikov, "Blockchain-based platform architecture for industrial IoT," in *Proceedings of the 21st Conference of Open Innovations Association FRUCT*, ser. FRUCT'21. Helsinki, Finland, Finland: FRUCT Oy, 2017, pp. 42:321–42:329. [Online]. Available: http://dl.acm.org/citation.cfm?id=3176190.3176232

[3] J. Honkola, H. Laine, R. Brown, and O. Tyrkk, "Smart-M3 information sharing platform," in *The IEEE symposium on Computers and Communications*, June 2010, pp. 1041–1046.

[4] J. H. Lee and M. Pilkington, "How the blockchain revolution will reshape the consumer electronics industry [future directions]," *IEEE Consumer Electronics Magazine*, vol. 6, no. 3, pp. 19–23, July 2017.

[5] K. Christidis and M. Devetsikiotis, "Blockchains and smart contracts for the Internet of Things," *IEEE Access*, vol. 4, pp. 2292–2303, 2016.

[6] F. Tian, "A supply chain traceability system for food safety based on HACCP, blockchain and Internet of Things," in *2017 International Conference on Service Systems and Service Management*, June 2017, pp. 1–6.

[7] V. A. Red, "Practical comparison of distributed ledger technologies for IoT," pp. 10 206–10 206–6, 2017. [Online]. Available: https://doi.org/10.1117/12.2262793

[8] S. Pongnumkul, C. Siripanpornchana, and S. Thajchayapong, "Performance analysis of private blockchain platforms in varying workloads," in *2017 26th International Conference on Computer Communication and Networks (ICCCN)*, July 2017, pp. 1–6.

[9] T. T. A. Dinh, J. Wang, G. Chen, R. Liu, B. C. Ooi, and K.-L. Tan, "BLOCKBENCH: A framework for analyzing private blockchains," in *Proceedings of the 2017 ACM International Conference on Management of Data*, ser. SIGMOD '17. New York, NY, USA: ACM, 2017, pp. 1085–1100. [Online]. Available: http://doi.acm.org/10.1145/3035918.3064033

[10] M. Y. Afanasev, Y. V. Fedosov, A. A. Krylova, and S. A. Shorokhov, "Modular industrial equipment in cyber-physical production system: Architecture and integration," in *2017 21st Conference of Open Innovations Association (FRUCT)*, Nov 2017, pp. 1–9.

[11] A. W. Colombo, T. Bangemann, and S. Karnouskos, "A system of systems view on collaborative industrial automation," in *2013 IEEE International Conference on Industrial Technology (ICIT)*, Feb 2013, pp. 1968–1975.

[12] A. Beregszaszi, F. Vogelsteller, A. Maiboroda, I. Matias, J. Pitts, and J. Steiner. Go ethereum official repository. [Online]. Available: https://github.com/ethereum/go-ethereum/releases

[13] M. OANCΞA. Ethereum network stats official repository. [Online]. Available: https://github.com/cubedro/eth-netstats/release

[14] A. Beregszaszi, F. Vogelsteller, A. Maiboroda, I. Matias, J. Pitts, and J. Steiner. Mist browser official repository. [Online]. Available: https://github.com/ethereum/mist/releases

[15] ——. Remix IDE official repository. [Online]. Available: https://github.com/ethereum/remix

[16] ——. The Solidity contract-oriented programming language official repository. [Online]. Available: https://github.com/ethereum/solidity

[17] J. Gerryts. The EthEmbedded official repository. [Online]. Available: https://github.com/EthEmbedded

[18] Raiden network–fast, cheap, scalable token transfers for ethereum. [Online]. Available: https://github.com/EthEmbedded

[19] C. Natoli and V. Gramoli, "The blockchain anomaly," in *2016 IEEE 15th International Symposium on Network Computing and Applications (NCA)*, Oct 2016, pp. 310–317.

Mesh Networking in Cyber-Physical Production Systems: Towards Modular Industrial Equipment Integration

Maxim Ya. Afanasev, Anastasiya A. Krylova, Sergey A. Shorokhov, Yuri V. Fedosov, Kseniia V. Zimenko
ITMO University
Saint-Petersburg, Russia
amax@niuitmo.ru, {ananasn94, stratumxspb}@gmail.com, yf01@yandex.ru, zksenia@yahoo.com

***Abstract*—Ensuring uninterrupted interaction of modular industrial equipment units is one of the most important engineering tasks. The concept of Cyber-Physical Production Systems (CPPS) assumes that the distributed network should correspond to the current industrial process and be able to quickly reorganize it when changes occur. If composition of the equipment becomes more complicated, a standard topology with one central control node might get ineffective. This article describes the application of mesh-network technology to ensure the interaction of industrial devices and sensors included in the modular equipment. Virtual deployment of the network and a description network nodes interaction including new node registration in the dispatcher registry are given.**

I. Introduction

Within the concepts of Cyber-Physical Production Systems (CPPS) modular architecture is getting more appropriate and gradually phasing out devices with a single monolithic architecture. Current trends require increased production flexibility in order to respond to the product range changes and customer requirements in time. It is assumed that modularity is not only a division into blocks in a physical way but also a modular software. In particular, the authors have already implemented the software architecture of modular equipment using the microservice approach [1]. This article brings up a question of creating a network infrastructure that uses modular equipment and special features of data transfer protocols.

Upon that, a division into unified blocks to simplify the architecture of industrial equipment leads to the inevitable complication of the connections between the equipment components located in a common decentralized network. However, at the moment there is no uniform standard for industrial distributed networks and the organization of network elements interaction among different companies is done in its own way.

On this basis, one of the main problems of existing CPPS is network implementation, which must meet the requirements of fault tolerance, security, and operation speed. The more production equipment is located on the production site, the more difficult to organize the interaction of network participants with each other. The reason behind this is a large number of connections central node is forced to work with within typical "star" network. In addition, the topology of devices and sensors distribution in the production area can be significantly different depending on the type of production process. Based on the foregoing, the network architecture being developed must have the capacity for self-organization and self-recovery.

One of the options for implementing the network architecture is a technology of mesh networks. Mesh network equipment allows choosing optimal frequencies and data transmission routes in automatic mode. The standard industrial network implies that each device is a closed unit of equipment consisting of interacting modules on the network, which are controlled by a special control module (dispatcher) [2].

Obviously, for this type of architecture, the classic "star" is more suitable. It allows quick subordinate control of each module without the need to search for it on the network and build an optimal route. Thus, each unit must stay closed and monolithic in order to solve directly the production tasks, because simple control model makes it more reliable and fault-tolerant. On the other hand, CPPS is not just a set of industrial equipment and sensors of the production environment. The second important entity of CPPS is a complete production model, also known as a "digital twin".

The presence of a "digital twin" is the main difference between CPPS and ordinary industrial sensor or control network. The control effect on the model should lead to a change in the physical component of CPPS and vice versa. At the same time, the system as a whole should stay absolutely transparent to the user e. g., replacing the physical components of CPPS with virtual ones should not affect the system functions. Obviously, to implement such a concept, the model should be as detailed as possible. For that reason, it is proposed not to use a monolithic approach to the industrial equipment design but instead a modular one with the possibility of combining all CPPS components into one decentralized network.

This results in a variation of the so-called holonic approach when the same system can be both hierarchical and heterarchical depending on working conditions. It should be explained with the example of the designing system. As it was already noted earlier, each unit of industrial equipment is a set of modules controlled by a common dispatcher over a network. Each module is an autonomous entity with its own internal operation logic, actuators, sensors, etc. With a decentralized network architecture, data from each module comes not only to the direct dispatcher connected with it but directly to the CPPS model.

An abnormal situation or a failure of any of the equipment modules would be immediately translated through the network. With that, the "digital twin" based on its own logic can immediately decide on the possibility of continuing the operation of an equipment unit in the module where failure

occurred (e. g., a failure occurred in an unresponsive or not involved at the moment module), or on the operational change of the production process and the transfer of the task to another (workable or unoccupied) unit of equipment. Additionally, this information will be distributed throughout the network (it should not be forgotten that CPPS is also integrated with cloud services), and this will work even if the failure occurs in the dispatcher module.

With this approach, the most reasonable way is to switch to the wireless organization of data transmission channels. In the case when the topology of the network is constantly changing, the installation of wire connections will be complex and expensive, besides there may be places with difficult access on the site. Wired connections will inevitably be damaged and rubbed in places of bends, which directly affects the overall reliability of the network infrastructure.

Summarizing all the above, the following requirements for the mesh network aimed to integrate modular equipment into CPPS can be formulated:

- The usage of the wireless method of data transfer.
- Openness and the possibility of integration with other production support systems: SCADA (Supervisory Control And Data Acquisition), cloud services, etc.
- High resiliency and the ability to self-repair.
- An embedded security system for data transmission.
- The usage of open standards.
- The network presentation as a digital production model, being the basis of the "digital twin".
- The unnecessity of network nodes manual configuration (zero configuration principle).

The remainder of this paper is organized as follows. In Section 2 the previous work on wireless production networks is presented including examples of MANET and SDN technologies. Section III focuses on the current progress in wireless production networks area as well as on the existing standards and their pros and cons.

Section IV describes the proposed approach to implement a wireless network for modular equipment. Section V discusses the future work and the existing challenges which the proposed approach does not solve. The conclusion is provided in Section VI.

II. Related Work

Dealing with challenges in network infrastructure is of current interest, that is underpinned by gradual enabling of Industry 4.0 principles. Obviously, the future of production is behind wireless networks. Speaking of wireless production networks, first of all, coverage and interference issues are usually raised [3]. One of the solutions—mesh networks application. In [4] authors consider such a network, serving for interaction and management of mobile vehicles. They developed a combined approach, where an existing production network infrastructure and a mesh network are used together. Authors in their research [5] employ the combined approach as well. They suggest IEEE 802.15.4 channel based mesh network that employed for point to point communication, while broadcast traffic is transmitted by the sub-GHz band in one hop. This channel is intended to determine an emergency or the case of communication loss. However, we reject the idea of using a combined approach. In our case, an emergency has to be resolved at the level of modular equipment via a dispatcher. This means that interaction will be executed in one hop without any additional channels. Along with that, elimination of the emergency at the level of the production network can be carried out with some delay, since this will not lead to fatal consequences.

In addition to the above, production network managing and configuration is a hot topic of a discussion. One of the solutions to this problem is to use the so-called self-organizing ad-hoc or MANET networks, where no additional configuration is required when adding nodes to the network topology. In the study [6] authors are coming up with an ad-hoc sensor network for work under extreme conditions, such as power plants or warships. For these purposes, they develop OCARIN technology based on ZigBee standard. However, to meet harsh requirements, ZigBee was extended. Authors add deterministic MAC layer for time-constrained communication and a more sophisticated algorithm for saving energy, allowing to enter the energy-saving mode not only for end devices but also controllers. Moreover, unlike traditional sensor networks, where nodes are stationary, OCARIN supports mobile nodes. We are trying to contribute in ad-hoc networks as well and propose our own version of the organization of a mesh ad-hoc network. However, unlike the studies above, this article is focused on the integration of modular technological equipment, that introduces its own features into the architecture of the network and requires a separate consideration.

In parallel with ad-hoc networks, such technologies as SDN and SON [7] are getting known. This approach doesn't allow one to achieve "zero configuration" principle but essentially simplifies network managing by making the controlled panel into a separate network node. The article [8] discusses its possible implementation. The authors note that in practice, many industries are already abandoning the use of fieldbus technologies, such as Modbus, Profibus, etc. in favor of industrial Ethernet. In such case, SDN application gives some benefits, e. g. remote update of PLC firmware and traffic management for interlevel communication between field and plant. The authors developed the architecture employing SDN technology and demonstrated a testbed based on such network and Raspberry Pi controllers.

Some authors are trying to combine the main principles of wireless ad-hoc networks and SDN to manage traffic manually. Thus, the authors of [9] research the possibility of creating a self-organizing network for tactical military purposes. The authors note that the application of general MANET is not enough, e.g. for coalition operations, where sensitive information has to circulate only between a certain group and never reach other network nodes. Thereby, the authors try to achieve benefits of centralized control as well as the ability to autonomously rebuild a network topology by enabling SDN technology. SDN obvious drawback is the centralized control and the presence of a single point of failure. Moreover, this technology, unlike the approach considered in this article, does not contribute to the principles of "zero configuration"

and "plug and produce", which are important for modular equipment and flexible re-configurable production.

The community is actively discussing the choice of the channel type for wireless data transmission in production. Currently, it is almost impossible to achieve control in a closed loop in a large production network due to delays and lack of reliability. However, with the advent of 5G networks, this will become realizable. There is an active discussion on the development of telecommunication standard 5G for production purposes [10]. The currently developed 5G standard will make it possible to get rid of all drawbacks mentioned above. In particular, Audi and Ericsson have recently launched a collaborative project [11] intended to introduce advanced communication systems based on 5G for industrial purposes that meet the requirements of the Industry 4.0. Moreover, the questions of integrating 5G networks with other types of networks existing on the plant are being considered, since it is obvious that 5g will not be able to displace them instantly [12]. However, in the current realities for the network described in this article, we focused on the architecture of the mesh based on the 802.15.4 channel and the OpenThread technology.

III. State-of-the-Art of Mesh Networking

Currently, there are four basic approaches to industrial mesh networks design:

- based on the set of standards IEEE 802.11 (Wi-Fi),
- based on the IEEE 802.15.1 standard (Bluetooth 4+, at usual Bluetooth Low Energy—BLE),
- based on cellular networks,
- based on the IEEE 802.15.4 standard (different variations of high-level protocols on the basis of special radio-channel with low bandwidth).

Each of the existing technologies is considered below in details.

A. Mesh networks based on IEEE 802.11 standard

The set of IEEE 802.11 standards defines the communication of devices in wireless computer networks in the frequency ranges from 900 MHz to 60 GHz. In practice, such networks are called Wi-Fi, so this term will be used in the future.

In accordance with the standard, Wi-Fi networks have an architecture consisting of at least one access point and one client. The access point transmits its network identifier (SSID) using special signal packets. Knowing the network SSID, a client can determine whether a connection to this access point is possible. If two access points with identical SSIDs fall within the coverage zone, the receiver can choose between them based on the signal level data. It is also possible to connect two clients on an ad-hoc principle, which theoretically makes it possible to say that the Wi-Fi standard is oriented towards decentralized networks, but this is not quite so.

In fact, the standard does not describe the mechanism for switching the network node from the access point mode to the client mode. That is, if two or more clients are on a network with the same SSID, in most cases they will not be able to communicate in peer-to-peer mode, even if it is more efficient for them in terms of data transfer speed. Such an assertion is easy to verify in any home Wi-Fi network: removing clients from the access point inevitably leads to loss of data transmission speed, even if they are in close proximity to each other. If one wants to increase the data transfer speed, one needs to move both clients closer to the access point.

Nevertheless, the standard does not explicitly prohibit the creation of decentralized mesh networks not from clients, but from access points (in particular, it is described in the IEEE 802.11s standard), which makes it possible to use this technology when designing the architecture of the Industrial Internet of Things. To date, such solutions are implemented at both hardware and software levels.

Analysis of hardware solutions for the creation of cellular networks based on Wi-Fi technology displayed that most of these devices belong to the class Small Office, Home Office (SOHO). Only a small number of manufacturers position their network devices as industrial. Among the most famous industrial solutions, it is possible to identify specialized Cisco devices, for example, Cisco Industrial Wireless 37xx and Cisco Aironet, many devices from Mikrotik, Juniper Networks, etc.

Such devices have proven themselves on the market, but according to the authors, their use to create an Industrial Internet of Things within the proposed concept of modular equipment is not entirely justified. First, such devices will not allow creating a completely decentralized network, where each client can directly communicate with everyone. Secondly, these solutions are quite expensive and often carry excessive functionality that will not be claimed in the area of the projected CPPS. Thirdly, all these solutions are proprietary, which calls into question the possibility of their modernization and improvement through the use of equipment from different manufacturers or open source solutions.

Next issue is the consideration of software implementations of mesh networks based on Wi-Fi technology. It should be noted that most implementations are based on the already mentioned IEEE 802.11s standard, but do not have a binding to any particular hardware vendor.

First of all, it is necessary to mention the project "open80211s" (https://github.com/o11s/open80211s). The goal of this project is to create a full and open implementation of this standard for use on devices running operating systems of the Linux family. Using `open80211s` allows overcoming the limitation on the use of mesh mode on devices of different manufacturers at the firmware level. The project has a good repository structure with source codes, a large number of developers, but, nevertheless, it is practically not being developed or used today. In particular, it should be noted that the last commit to the main branch (master branch) of the repository on `github.com` was made five years ago.

In conclusion, it should be noted that the idea of using open source and hardware solutions based on ESP devices family seems pretty promising. Such devices are based on special ESP microchips (like ESP8266 or ESP32) with embedded Wi-Fi interface and ability to execute programs from external flash memory. Considering it as a basis for a mesh network we find several projects—"painlessMesh" (https://github.com/gmag11/painlessMesh) and "easyMesh" (https://github.com/Coopdis/easyMesh). Both

projects provide firmware and API that allows one to build an ad-hoc wireless mesh network. The network implemented with these projects has no IP communication instead of a simple unique chipid identification is used. Moreover, messaging is based on a simple JSON format that seems useful for easy integration with external systems. However, practical usage of these projects appears to be incompatible with real production cases.

First of all, there are performance issues that constrain the number of messages per minute limited by processing power, a memory of the controller and messaging format. Along with that, there is no protection from overloading and self-stabilization. Due to the security issue, a password is used that requires to configure a node before its integration and goes against ad-hoc networking principles. Giving the above, such solutions are more suitable for home automation rather than the industrial field.

B. Cellular Internet of Things

It is obvious that the structure of mobile cellular networks is very similar to the structure of mesh networks. This similarity has led to the emergence of a whole new direction in the development of decentralized networks, called the Cellular Internet of Things. To date, there are at least three technologies related to this class: EC-GSM-IoT, NB-IoT, and LTE-M.

The first of them—EC-GSM-IoT (Extended Coverage–GSM–Internet of Things)—is a standard for low-speed data exchange in 2G, 3G, and 4G mobile networks. EC-GSM-IoT regulates communication of various smart things with the purpose of data transfer between them. The standard allows utilizing all the advantages of GSM-networks available today: device authentication, confidentiality of data transmission, data integration, etc.

In addition to that, this standard allows overcoming the limitations associated with increased power consumption of devices that work as classic mobile GSM devices (phones, smartphones, tablets, etc.). The latter is due to the specifics of smart things, as a rule, they all do not need to maintain a continuous connection and do not require high data rates.

The other two technologies are practically the same as the EC-GSM-IoT. LTE-M is the logical development of LTE networks for IoT while maintaining backward compatibility with the LTE standard. While NB-IoT uses a different type of modulation and is not compatible with LTE networks.

In general, all presented approaches are quite interesting and promising, but not suitable for the implementation of industrial networks. Despite the fact that the developers of standards speak about the possibility of using their technologies in the industry, in fact, there is some shift in the direction of logistics, tracking, diagnostics, supply chain management, etc. And the main areas where these protocols can be applied are the oil and gas industry and food production. The use of such technologies in instrument engineering and machine building is extremely doubtful, according to the authors.

The main disadvantages of mobile networks application in a subject area considered in this paper:

- the low latency of the network, which is permissible for solving of sensors' data receiving problems but not permissible for control (even non-real time one)
- binding to external data networks: failure of the mobile network will lead to the failure of the entire network of the enterprise.
- binding to the equipment of a specific manufacturer.
- the necessity of using physical SIM-cards for access to the network (the Embedded SIM standard has not yet become widespread).

C. Mesh networks based on the IEEE 802.15.1 standard

The IEEE 802.15.1 standard defines the physical layer parameters for various devices joined by a wireless personal network (Wireless Personal Area Network, WPAN). Usually, it is denoted by the term Bluetooth (or Bluetooth Low Energy—BLE). Bluetooth combines the inexpensiveness of equipment, independence from any specific equipment manufacturer, the simplicity of the protocol, sufficient performance, the ability to connect nodes at distances up to 100 m, the presence of an integrated mechanism for ensuring data confidentiality, and sufficiently low delays in transmission in the network.

Since 2017, the Bluetooth standard officially supports mesh network mode, that makes it extremely promising for Industrial Internet of Things (IIoT) systems. There is already a sufficient number of successfully implemented projects for the IIoT, where BLE was chosen as the basis for building a network. All of the above speaks for BLE application in a CPPS core network as it meets all the requirements stated in Section I.

However, this technology still has drawbacks. Firstly, only the latest versions of the Bluetooth 5+ standard implements the mesh mode. Thus, it is still difficult to find ready-made solutions (or even systems on the Crystal, SoC) for the latest version. Secondly, there is no common standard for the transmission of IPv6 traffic via Bluetooth 5+. There is a standard RFC 7668 and a number of researches devoted to the standard description and examples of mesh-networks created on its basis [13] but it has not yet been widely spread.

D. Mesh networks based on the IEEE 802.15.4 standard

As well as the Bluetooth standard IEEE 802.15.4 defines the physical layer and media access control (MAC) for wireless personal networks. The remaining levels of the OSI model are not described. There are several implementations of this standard to specify higher levels. The most common technologies based on the IEEE 802.15.4 standard are ZigBee, ISA100.11a, WirelessHART, MiWi, Z-Wave, and OpenThread.

In general, the differences between networks based on IEEE 802.15.4 and IEEE 802.15.1 are the following:

- The IEEE 802.15.4 protocol is designed for a lower throughput. It is about 1 Mbitps for BLE and no more than 260 kbitps for IEEE 802.15.4
- BLE uses one frequency band within 2400–2483.5 MHz while IEEE 802.15.4 allows network nodes to communicate on others non-licensed frequency bands—868.0–868.6 MHz and 902–928 MHz

- Initially, Bluetooth was oriented to the "star" type topology, point-to-point connections with the ability to organize mesh networks appeared only in the latest version of the standard. In the IEEE 802.15.4 specification, both topologies were originally described.
- For a long time, Bluetooth was focused on consumer electronics, and only recently projects that use this technology in industry and production began to appear. Whereas, The IEEE 802.15.4 specification was originally designed for industrial use.

The most promising implementations of the IEEE 802.15.4 specification to create a cyber-physical system for the integration of modular technological equipment are ZigBee and OpenThread. These technologies are very similar, but have the following slight differences:

- ZigBee is more developed because it appeared earlier. It has larger community, more detailed documentation, and more usage examples.
- OpenThread is a more open technology because it was originally developed to attract as many developers as possible.
- OpenThread has more portability. It supports both system-on-chip (SoC) and network co-processor (NCP) designs.
- OpenThread is focused on the transfer of IPv6 traffic and uses the standard 6LoWPAN (IPv6 over Low power Wireless Personal Area Networks). Therefore, each node can be addressed with its IP address. ZigBee was not originally IP-based technology. To transfer traffic to the IP-network, you either need to use edge router or a new separate protocol ZigBee IP, which, however, has not yet found wide distribution.

In conclusion, it should be noted that it is extremely difficult to choose one of these three technologies: BLE 5+, ZigBee, and OpenThread. The search for information in open sources showed that manufacturers of equipment are of the same opinion—many of the solutions on the market support several protocols at the same time.

Nevertheless, in our opinion, the most interesting and promising technology is OpenThread. Thus, it will be used in the future description of the infrastructure of the developed modular equipment interaction network.

IV. Proposed approach

The main feature of the modular industrial equipment is the ability to quickly re-configure any unit of equipment when changing a technological process. Re-adjustment is carried out due to the change of processing tools and external modules. This is achieved by maximizing the autonomy of each processing head or module. Each of them has its own functioning algorithms, as well as a set of sensors and actuators. All processing heads and modules can communicate with each other via a unified protocol, with a set of commands and data minimized, all commands are strictly regulated. The main module is a three-axis chassis that moves the processing heads in three-dimensional space. Naturally, from the control point of view, such a system can not be completely decentralized—with each equipment unit its virtual dispatcher ("digital twin") is connected, which is a model of equipment and coordinates modules, heads and chassis. It is the presence of the dispatcher that allows quick adjustment of equipment. All heads and modules are configured in the same way, they all know their capabilities and protocol of interaction, so when they are physically connected they find their dispatcher in the network, register their services, and then they are ready to work right away.

This concept works fine if it is imagined that only one unit of equipment exists, but it is not. CPPS can consist of many hundreds and thousands of devices, and the vast majority of these devices is precisely the technological equipment. In this case, as already mentioned above, each piece of equipment, each processing head, each module, and each sensor is connected to a common decentralized mesh network, because they are parts of the same CPPS.

At this moment the problem is traced: if all modules are located to a decentralized network and have the ability to find their dispatcher, how can they determine that this particular controller is physically connected to this module. A usual way solution: to assign this duty to the operator. However, this will break the "zero configuration" principle, that is, the operator should only think about the technological process, not about the configuration or reconfiguration of the equipment. Here an analogy with conventional universal equipment can be given. A worker who performs operations on a lathe, for example, should not think about the information compatibility of the tool and the machine. He has enough physical compatibility. For example, the cutter is physically placed in the tool post. It is this simplicity of readjustment that we want to achieve in our work, not just for "dumb" universal equipment, but for "smart" equipment integrated into CPPS.

From all of the above, it can be concluded that the organization of a CPPS mesh network, in which modular equipment is used, is not a trivial task. Therefore, the rest of the section will be devoted to the deployment of the OpenThread test network, the description of the general CPPS architecture built on it, as well as principles of the modular equipment included in the CPPS.

A. *OpenThread testbed*

As already noted in the previous section, OpenThread technology was chosen as the core CPPS network by its set of merits and demerits. This wireless data transfer protocol is a BSD licensed open-source implementation of the Thread network protocol originally developed by Google Nest. A distinctive feature of OpenThread is maximum portability, while according to the specification, various designs can be implemented, both purely hardware and software-hardware. Such flexibility allows one to first create a support network structure using virtualization technologies, test and debug it, and then move it to physical devices.

To deploy the virtual network, a set of debugging tools was used, located in the OpenThread repository (https://github.com/openthread/openthread). In particular, the assembly of a network client for the x86 architecture was used with the emulation of the radio module via a network

card with messages transmission using the UDP protocol. The primary tasks were the construction of a simple network and working out, on its example, an algorithm for the nodes interaction. To simplify the task, the original network structure consisted of three independent nodes (Fig. 1). Each node of the network was deployed the same way, which allowed to automate this process in a virtual environment.

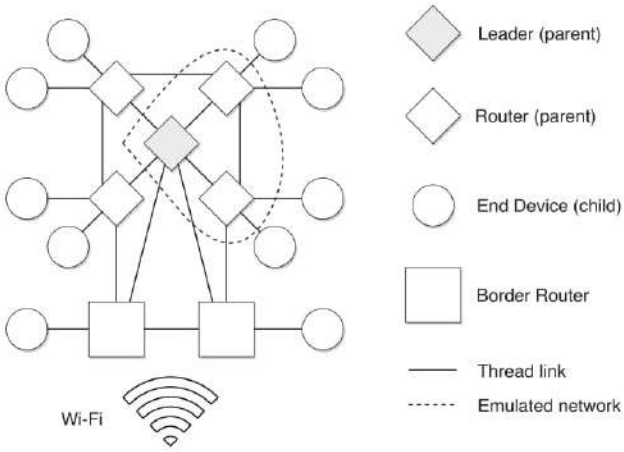

Fig. 1. OpenThread architecture

In accordance with the protocol specification, there are two types of nodes:

- A Router that never turns off the transceiver, participates in packet transmission over the network, and is also a security node for all new connected devices.
- End Device, which mainly interacts only with its router, does not transfer data over the network and can turn off its transceiver to save power.

In addition to this, all devices in OpenThread mesh network can be divided into the Full Thread Device and the Minimal Thread Device. The Full Thread Device never turns off its radio transmitter, subscribes to the all-routers multicast address, and maintains IPv6 address mappings. Full Thread Devices consists of Router, Full End Device (can not self-elect as a router) and Router Eligible End Device. The latter works as an End Device, but it can become a Router if the number of Routers on the network is less than 16, or if a Full End Device appears next to it.

A Minimal Thread Device does not subscribe to multicast traffic and forwards all messages to its Parent. There are two types of Minimal Thread Devices: Minimal End Device—radio transceiver always on, does not need to poll for messages from its parent and Sleepy End Device—normally disabled, wakes on occasion to poll for messages from its parent. Obviously, the role of Minimal Thread Device is mainly used for nodes that are autonomous devices with battery power source. In scenario considered in this paper the using of Minimal Thread Devices is not planned.

Of the entire set of routers, one of them (usually the first turned on) is assigned as Leader. Leader aggregates and distributes information about the configuration of the entire network. There can only be one Leader in one PAN (Personal Area Network). In order to ensure reliability and fault tolerance, the Leader's role dynamically moves from one node to another, that is, any router can self-elect as a Leader. Also, one Border Router can be assigned to the network, which is responsible for the interaction and transfer of IPv6 traffic between the mesh network of OpenThread and other networks (Wi-Fi, for example).

Thus, the test network will consist of one Router, with the assigned role of Leader and two Router Eligible End Devices, which will demonstrate the ability of changing the role when disconnecting/connecting nodes.

As a test environment for virtualization, a server with the following technical specifications is used: 2 x Intel Xeon E5620 CPU @ 2.40 GHz, 32 GiBRAM, Intel 82575EB Gigabit Network. Operating system: Ubuntu 16.04.2 LTS, GNU/Linux 4.4.0-64 x86-64. As a hypervisor, Oracle VirtualBox 5.2.16 and the Vagrant 2.1.2 virtual environments management system were used.

The virtual machine configuration file is downloaded from the official github repository using the command

```
$ git clone https://github.com/openthread/openthread
```

After this, deploying of the guest operating system and configuring of OpenThread client are executed:

```
$ cd ~/openthread/etc/vagrant
$ vagrant up
```

During the installation process, the configuration of the basic guest machine running under the operating system Ubuntu 14.04 LTS is performed. This machine is a template by which an arbitrary number of OpenThread nodes can be created. To connect to a virtual machine, one should use the following command:

```
$ vagrant ssh
```

Then creating the first node using the console client of the network can be started. It should be noted that in this example all actions for node configuration and interaction over the network will be done manually by entering commands into the shell client. In the future, all these actions will be automated due to the fact that each shell command is just a call to a function C Thread API. The console client launch:

```
$ cd ~/src/openthread
$ ./output/x86_64-unknown-linux-gnu/bin/ot-cli-ftd 1
```

The first parameter indicates that the transmission will use radio emulation via the UDP protocol. Specifically, this parameter point on the port number to be used for communication ($9000 + 1 = 9001$ port). For subsequent nodes this number must be different. Naturally, in a real network, this configuration is not required. All devices are configured in advance in a uniform manner and operate on one of the predefined radio channels IEEE 802.15.4. Non-reserved values are in the range of 11–26. First nodes are configured with these commands:

```
> panid 0x1234 # to set the unique network id
> ifconfig up  # to start network IPv6 interface
> thread start # to start Thread protocol
```

After Thread protocol was started, the first node checks the network for the presence of other nodes and, since there are no other nodes in the network, assigns the role of Router and Leader. The second and third nodes are configured using the same commands. The only difference is that during the initial scan of the network, the nodes see that there is already one Router on the network, therefore they initially assign the role of Child, which automatically changes to the Router after a two-minute timeout. This is due to the fact that the Thread network tries to maintain the number of Routers in the network in the range 16–23, respectively, until the minimum value is reached, each newly connected node will change its role on the Router.

The demonstrated example of the configuration of the OpenThread test network clearly shows that it is possible to perform a complete preliminary configuration of devices without the need to manually connect the new device to the network. At the same time, it can be noted that the configuration is a very simple procedure, the connection to the network is very fast, and after the launch, any network node immediately identifies all the neighboring nodes and can interact with them via IPv6 protocol.

B. Communication protocol

In the previous section, the procedure for physically deploying the decentralized backbone CPPS network was considered. The result is a ready-made backbone network in which nodes can exchange raw data. However, it should not be forgotten that this system is not an ordinary sensor network, where only the simplest protocols of the application layer are used, for example, the MQTT message queue.

Of course, the presence of all possible sensors monitoring the production process is implied, but the basis of CPPS is a modular industrial equipment. On the one hand, it consists of autonomous and independent modules located in a common mesh network. On the other hand, all these modules are able to organize stable hierarchical formations to perform specific tasks of production. From the last statement it follows that for the effective operation of such a heterogeneous network, a more complex protocol for two-way interaction is needed.

This protocol should to be considered in more detail. Consolidation of modules is carried out around the base module, which acts as a dispatcher. Since it is assumed that to ensure the flexibility of equipment, it must be maximally unified, each piece of equipment includes in its composition a mandatory module—a 3-axis chassis that moves the processing heads in workspace. Of course, many types of processing require the possibility of moving the working element to 4 or 5 coordinates, but, as practice shows, in many even professional 5-axis machining centers, the 3-axis chassis is used as the basis, and additional coordinates are executed as a separate module. Accordingly, from a control point of view, it is the 3-axis chassis that acts as the dispatcher around which one or another unit of equipment is configured.

Each dispatcher has a registry that is part of a distributed JSON-based repository. The registry consists of slots and only one module can be registered in each of them. In its turn, the totality of all registers is a "digital twin" of CPPS and is stored in the cloud. For uniformity in the same scheme, a digital model of devices that are not industrial equipment is created (different sensors of the production process, at first). The sensor is a dispatcher with one slot without the possibility of re-registration (Fig. 2).

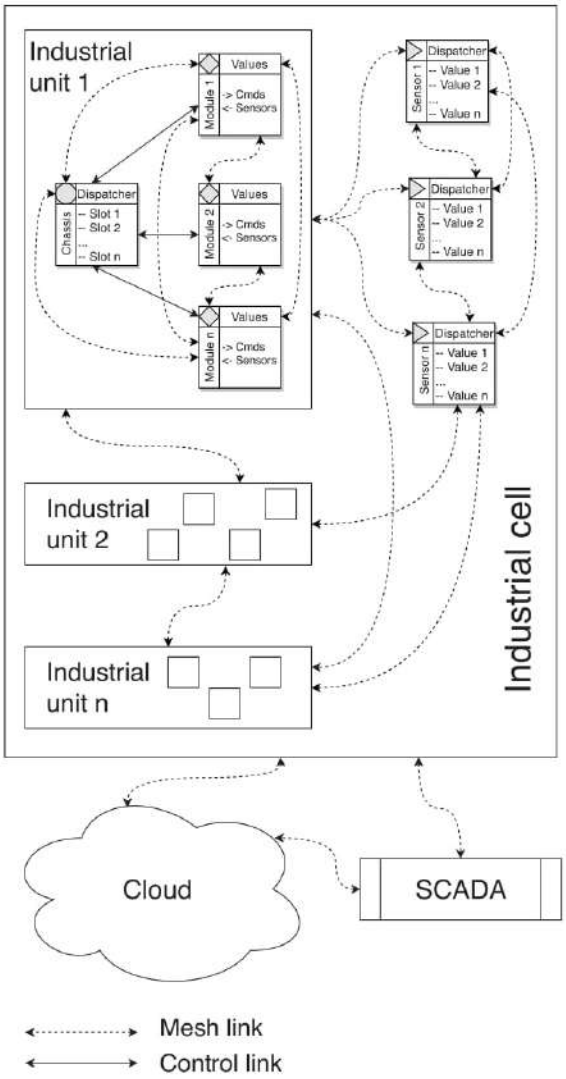

Fig. 2. General CPPS architecture

Slot is a "key-value" type record from which dispatcher retrieves the required data. Slot consists of from:

- an address;
- a module/sensor name;
- functions;
- a return value;
- a range of the return value.

The address is the IPv6 address and port that are required to communicate with the module. The name must be in string or integer format. Available functions are described by a set

```
{} JSON
  address : "[fdde:ad00:beef:0:0:ff:fe00:fc00]:9001"
  caption : "Laser module"
  [] functions
    0 : "M3"
    1 : "M5"
    2 : "S"
  cmd : "M3S500"
  [] sensors
    {} 0
      {} temp
        type : "int"
        [] limits
          0 : 18
          1 : 120
        [] values
          {} 0
            20180609T183100 : 45
          {} 1
            20180609T183200 : 48
          {} 2
            20180609T183200 : 47
          {} 3
            20180609T183400 : 46
```


Fig. 3. Example of a slot

of G-codes and M-functions in accordance with ISO 6983-1 and ISO/TR 6983-2. For example, the milling head will be able to work with the commands M3 and M4 (starting the spindle clockwise and counterclockwise), M5 (spindle stop) and S (rotation speed); for the laser head—this will be the commands M3 (turn on the laser), M5 (turn off) and S (set the power in percents); for the tool store—M6 (change tool) and T (choose the tool from the store by number). In this case, all the commands related to moving the head in three basic coordinates are determined by the chassis module.

There can be more that one return value. The description of each of them includes a type, a range, and an array of values with timestamps. The dispatcher controls the output of each value beyond the permissible limits, and in the event of such a situation, analyzes the error and makes a decision whether or not the equipment can continue to operate. An example of a slot is shown in Fig. 3.

The procedure for registering the module in the dispatcher's registry needs to be explained. From the protocol point of view, there are no problems here: simply transferring the binary JSON message through the message queue and then writing it to the distributed store [14]. However, it should be noted that all dispatchers and all modules are in a single self-organizing mesh-network. The question arises—how the module will determine to which specific dispatcher it is connected physically.

The following is proposed. Each module is physically connected using 3-wire interface, and two wires are used as power supply and ground lines while third wire is used as a safety line. The safety line connects all equipment units modules as wired and logic circuit of an AND type (Fig. 4). The safety line utilizes the pull-up resistor. Due to huge resistance between power supply line and ground and relatively low resistance, that equals connected resistor between safety line and power supply, the voltage on safety line equals to power supply voltage. Thus the low and high logic levels are formed. As it was spoken before, all modules are connected to safety line and can pull it to ground. Consequently, the high logic level on the safety line will be presented only when all the modules switch the outputs to high logic level. If any of the modules ground the safety line, the low logic level on the safety line will be presented and none of modules will be able to switch it back to high.

Thus, the main task of the safety line is to register emergency situations. In case of an accident in any of the modules, it simply connects the safety line to the ground, which is a signal for other modules to stop working and get into recovery mode after a failure. It should be noted that the emergency stop button is also connected to the same line, which is mandatory for any industrial equipment, as well as all safety doors, if they are provided by the design. However, during the initialization of modules, the safety line can be used to register new modules. The newly connected module firstly connects to the OpenThread network, then sets the safety line to a low logical level. The dispatcher detects this, then gets a list of all neighbors closest to it, chooses those that have the status "not connected" among them, then asks the first of them to set the line back to a high logical level. If the level has changed, then the dispatcher and the module are connected to the same line, therefore, the module can be registered in the registry. Otherwise, the dispatcher moves to the next module in the list. At the same time, it is known that all dispatchers can communicate with each other, so one needs to establish a rule that regulates the order of initialization of dispatchers, i. e., if one of them spends filling the slots of its registry, the others should be in the standby mode and do not accept connections from the modules. It is also clear that the process of connecting a new module can be carried out only in the case, when the equipment is in standby mode.

V. Discussion

Currently, the proposed approach is intended for small-batch production companies. It is really convenient for such companies to have a set of universal and easy to reconfigure equipment that can be joined into various chains according to a technological process. The suggested idea allows one to solve shop floor organization and equipment placement issues, as each equipment unit based on the universal chassis only depend on the electrical network. Moreover, the ability to remove unused modules optimizes equipment storage space and gives the opportunity to move quickly in case of company relocation. This concept can be called "a production office" due to the ease and speed of production deployment at a new placement like simple office equipment. Upon that, there is practically no equipment on the market that would have such properties

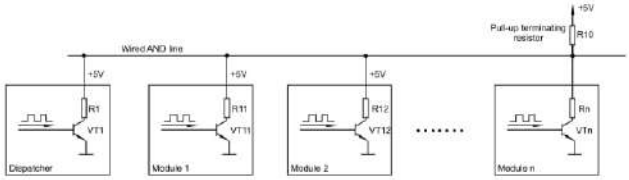

Fig. 4. Wired AND line circuit

nowadays. The only exception is industrial 3D printers, which can be put into operation in a matter of hours. They do not have special requirements for deployment though. Moreover, 3D printers are connected to the wireless network and can work on ordinary office premises.

On the other hand, the idea to extend this approach for large production companies makes sense. Indeed, features of wireless self-organizing mesh-networks can be useful in mass production. Mesh networks are already partially applied in the form of sensor networks covering large production areas, where laying additional cable lines is unprofitable. The benefits, in this case, are 50–90 % compared to the cost of a wired network [15]. However, we believe that having a common network for all industrial devices is just as important as collecting data from sensors. There are three problems that have to be solved when implementing a common mesh network that joins not only sensors but also industrial equipment. First, the IEEE 802.15.4 standard does not imply the possibility of data transmission over sufficiently long distances without significant loss of speed. It is especially important for automated (automatic) guided vehicles or mobile robots used in production. To solve this issue stationary retransmitters located at the node points of workshops and other production facilities can be applied. Secondly, the security issue is still not solved. IEEE 802.15.4 protocol describes a physical layer of the OSI model, but it is not enough to prevent complex attacks. Currently, we are carrying on the research on the development of a two-level protocol. In this protocol, a message queue is applied to transfer data from autonomous sensors to a common network and for equipment modules communication. At the CPPS level blockchain technology is employed as well as smart-contracts [16]. Thirdly, the issue of radio frequency range choice still exists. The majority of modern wireless networks uses a frequency of 2.4 GHz. Obviously, this frequency is very congested and noisy. Globally, this problem can only be solved by allocating exclusive frequency ranges for industrial automation. However, for small production, it is possible to use local screening of production facilities.

VI. Conclusion

The main issue in organizing the work of CPPS is a distributed network of modular equipment that can include thousands of devices, sensors and controllers, which must communicate with each other and react to events in a timely manner. The composition complication of the production equipment leads to the connection complication between the equipment units. At some point, the standard "star" topology is obviously inefficient for such networks, since there is a risk of overloading the central node. In addition, this topology directly contradicts the concept of decentralized management. This paper briefly presented the network architecture based on the application of mesh-networks, in particular, on the OpenThread technology. The possible approaches to the implementation of mesh networks for industrial use have been described. Further, the mesh-network deployment architecture is proposed and its virtual model is tested. The description of interaction of network nodes and the procedure for registering a new module in the register of the dispatcher are given.

Obviously, this paper shows the first stages of development and implementation of the proposed architecture. Of course, there are a number of architectural limitations related both to the technical features of the technologies used and to the security of the described interaction, which will need to be solved consistently in further research.

VII. Acknowledgements

This work was carried out under project no. 617026 "Technologies of cyber-physical systems: management, computing, security".

References

[1] M. Y. Afanasev, Y. V. Fedosov, A. A. Krylova, and S. A. Shorokhov, "An application of microservices architecture pattern to create a modular computer numerical control system," in *2017 20th Conference of Open Innovations Association (FRUCT)*, April 2017, pp. 10–19.

[2] ——, "Modular industrial equipment in cyber-physical production system: Architecture and integration," in *2017 21st Conference of Open Innovations Association (FRUCT)*, Nov 2017, pp. 1–9.

[3] A. Varghese and D. Tandur, "Wireless requirements and challenges in industry 4.0," in *2014 International Conference on Contemporary Computing and Informatics (IC3I)*, Nov 2014, pp. 634–638.

[4] I. M. J. Haxhibeqiri, E. Jarchlo and J. Hoebeke, "Flexible Wi-Fi communication among mobile robots in indoor industrial environments," in *Mobile Information Systems*, 2018, p. 19.

[5] F. Kauer, E. Kallias, and V. Turau, "A dual-radio approach for reliable emergency signaling in critical infrastructure assets with large wireless networks," *International Journal of Critical Infrastructure Protection*, vol. 21, pp. 33–46, 2018.

[6] T. Dang and C. Devic, "OCARI: Optimization of communication for ad hoc reliable industrial networks," in *2008 6th IEEE International Conference on Industrial Informatics*, July 2008, pp. 688–693.

[7] P. D. Wegener, *GERMAN STANDARDIZATION ROADMAP Industrie 4.0 Version 3*. DIN e. V., 2018.

[8] K. Ahmed, N. S. Nafi, J. O. Blech, M. A. Gregory, and H. Schmidt, "Software defined industry automation networks," in *2017 27th International Telecommunication Networks and Applications Conference (ITNAC)*, Nov 2017, pp. 1–3.

[9] K. Poularakis, G. Iosifidis, and L. Tassiulas, "SDN-enabled tactical ad hoc networks: Extending programmable control to the edge," *CoRR*, vol. abs/1801.02909, 2018. [Online]. Available: http://arxiv.org/abs/1801.02909

[10] U. Graf, R. Heidel, and D. G. Kadel, *DISCUSSION PAPER. Network-based communication for Industrie 4.0*. BMWi, 2016.

[11] L. Bösch. Spokeswoman Audi IT. Audi and Ericsson to pioneer 5G for automotive manufacturing. [Online]. Available: https://www.ericsson.com/en/press-releases/2018/8/audi-and-ericsson-to-pioneer-5g-for-automotive-manufacturing

[12] A. Neumann, L. Wisniewski, R. S. Ganesan, P. Rost, and J. Jasperneite, "Towards integration of industrial ethernet with 5G mobile networks," in *2018 14th IEEE International Workshop on Factory Communication Systems (WFCS)*, June 2018, pp. 1–4.

[13] P. Narendra, S. Duquennoy, and T. Voigt, "Ble and ieee 802.15.4 in the iot: Evaluation and interoperability considerations," *Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, LNICST*, vol. 170, pp. 427–438, 2016, cited By 4.

[14] M. Y. Afanasev, Y. V. Fedosov, A. A. Krylova, and S. A. Shorokhov, "Performance evaluation of the message queue protocols to transfer binary json in a distributed cnc system," in *2017 IEEE 15th International Conference on Industrial Informatics (INDIN)*, July 2017, pp. 357–362.

[15] J. Colpo and D. Mols, "No strings attached," *Hydrocarbon Engineering*, vol. 16, pp. 47–52, 2011.

[16] M. Y. Afanasev, Y. V. Fedosov, A. A. Krylova, and S. A. Shorokhov, "An application of blockchain and smart contracts for machine-to-machine communications in cyber-physical production systems," in *2018 IEEE Industrial Cyber-Physical Systems (ICPS)*, May 2018, pp. 13–19.

A Concept for Integration of Voice Assistant and Modular Cyber-Physical Production System

Maxim Ya. Afanasev, Yuri V. Fedosov, Yuri S. Andreev, Anastasiya A. Krylova, Sergey A. Shorokhov, Kseniia V. Zimenko, Mikhail V. Kolesnikov
Faculty of Control Systems and Robotics ITMO University, St. Petersburg, Russia
amax@niuitmo.ru

***Abstract*—Voice interaction is helping to increase efficiency in many areas of human activity. However, most studies either research its general aspects or domestic applications. On the contrary, there is almost no research on using this technology in production field, although, the idea of retrieving data this way can be a successful solution for large production areas, where it is difficult to find the right display. The paper aims to develop the method of voice assistant's implementation as a part of Cyber-Physical Production System (CPPS) for production processes. The study includes creating a prototype of the voice assistant module based on Alice cloud voice service API from Yandex and Raspberry Pi 3 microcomputer. Main aspects of using voice assistants for accessing production data from a CPPS were described. A possible approach to the introduction of this technology is shown. And a working prototype has been successfully designed. The research shows the possibility of using voice control in production conditions and its potential effectiveness in data access automation.**

***Index Terms*—Voice assistant, cyber-physical production system, modular equipment**

I. Introduction

Historically, humanity in its activities seeks to make the work easier by delegating the processing operations to more and more improving tools up to complex automated machines. Modern cyber-physical systems are focused on minimizing the human factor in a production process. Generally, they include not only the equipment itself, but also the interaction channels with the virtual environment, where huge amounts of data can be collected and processed. Wide use of information technologies, local area networks and data aggregation systems are all aimed primarily at automating routine actions, reducing the number of errors and failures in production, and improving product life cycle support. For the same purpose a "digital twins" of production line are being developed. A digital twin is a virtual model of a physical device that simulates internal processes, technical characteristics and behavior of a real object under the influence of interference and environment.

Each product manufacturing company is faced with a large amount of data on the production process, such as: machine parameters, surface treatment parameters, sensor data and more. Different systems of data aggregation cope with the task of collecting and organizing information very well, but the task of timely delivery of necessary data to an operator is just as important. It is also necessary to emphasize the fact that data should be prompt, videlicet, the request should be processed as soon as possible, and data should be presented in a convenient for an operator form.

Although data is often displayed on monitor screens and machine displays, there are often situations where it is not enough. For example, to get the required number, a person has to first find the desired display, approach it, find the target number in the complex interface of the program and go back to their place. It is obvious that the larger the production area, the more time will be spent on such a trivial routine in terms of manufacturing process. To solve the problem of process optimization to simplify data reading, the so-called "augmented reality"—the integration of virtual objects into a real world image—is being used more and more often. For the first time, the experience of augmented reality in production was obtained in 1992 by Boeing Company: the image of wired circuits was displayed on an aircraft body, which significantly reduced the number of errors in production.

Currently, augmented reality systems are widely distributed in both entertainment and industry. Generally, such systems imply operator-wearable equipment in the form of glasses, but there are options that use the built-in camera of mobile devices like tablets. One of the most famous solutions is the German RE'FLECT company, which specializes in producing the appropriate equipment for leading German enterprises. Another brand, EON Reality, chose a simpler way and focused on developing software for smartphones and tablets. Among Russian developments Itorum MR is the most well-known.

Another way to get information quickly is to use voice assistants. According to literary sources, the visual channel gives 80–90% of all information [1], and the second most important, auditory, is practically not involved in production. For example, if a colleague stands in a workshop next to the desired screen, then it is much easier to ask him than to go and look for the needed parameter yourself. In addition, operators are often being deeply involved in working process so that voice control remains the only possible way of interacting with an information system.

The benefits of voice assistants were assessed back in the 70s, when the first prototypes of speech recognition devices appeared. Lately, almost all leading IT-companies have among their developments a voice assistant: a voice input by Google, Alexa by Amazon, Siri by Apple, Cortana by Microsoft, Alice from Yandex and others. At the moment, their capabilities have significantly expanded from a simple search for information to

978-1-7281-2927-3/19/$31.00 ©2019 IEEE

Authorized licensed use limited to: St Petersburg Natl Uni of Info Tech Mech & Optics. Downloaded on January 31,2021 at 10:04:20 UTC from IEEE Xplore. Restrictions apply.

the full voice control of various household devices. Of course, the use of voice assistants in a production environment has its own nuances like noise pollution of manufacturing premises, simultaneous interaction with several operators, and security issues.

Another important task is the integration of a voice assistant system in a structure of a modular architecture. Modularity is a modern concept of organizing a decentralized control system for production equipment aimed at increasing flexibility and scalability of a system. Plainly, such architecture implies not only physical components, but also its own software, which simplifies the integration of new components into a control system. In the platform architecture described in previous papers, each piece of equipment can be represented as a set of modules, otherwise stated, managed by a dispatcher unit.

Most modules are closed autonomous entities working on their tasks, but dependent system components may also exist. Dependencies mainly include sensors that provide one-way data sending to a request. Naturally, the voice assistant module will be autonomous, since its task is only to communicate with CPPS directly, download data from a "digital twin" and issue it to an operator. Next the organization of the voice assistant in industrial equipment will be discussed.

The article is organized as follows. Section II is dedicated to an overview of related work. In Section III the reader can find a description of the prototype module of the voice assistant module. Section IV describes the limitations of the technology and how to overcome them at this stage of development. The final section is dedicated to an analysis of the results.

II. Related work

Despite the fact that in recent years, interest in voice assistants has been steadily growing, today there are not many implementations for its industrial and manufacturing use. The most interesting work in this area is the initiative of connyun GmbH companies to introduce Alexa voice assistant from Amazon to industry [2]. The platform I_Station Optimizer, developed by the company, will connect with Alexa voice assistant in production processes. This will allow workers and employees of enterprises to conduct a direct dialogue with the IIoT platform, for example, to request data, set tasks and solve problems in dialogue with the voice assistant.

There are also a sufficient number of theoretical studies related to the analysis of existing cloud services and prospects for the speech recognition technologies and voice assistants' introduction and use in various areas of production for solving the widest spectrum of problems.

In the study [3], in particular, a detailed analysis of the automatic voice recognition applicability in medicine, industrial robotics, forensic science, arms industry, and aviation was conducted. It has been shown that with the current level of technology development, automatic speech recognition can be a worthy alternative to the classical tools of human-machine interaction. Nevertheless, the authors come to the conclusion that a success of speech recognition applications requires the limitations elimination in modern technology, the use of new specialized hardware, in particular. However, it should be noted that this work is primarily aimed at implementing an autonomous speech recognition system that does not use cloud service resources.

Hoy [4] provides a comparative analysis of popular cloud services that provide voice assistants. The paper discusses basic principles of modern voice assistants' work and its common features, and also discusses some confidentiality and security issues, as well as some potential possibilities of using it in the future.

Bradley et al. in his work [5] addresses the applicability problems of interactive agents, made to simplify the work of software developers. As an example, a voice assistant prototype called Devy, based on the Amazon cloud service from Amazon, is described. The implemented prototype demonstrates that developers can successfully launch complex workflows without interrupting their current work. Based on the study, the authors conclude that in the future, voice assistants will be able to increase the productivity and/or efficiency of software developers, allowing them to focus on their core tasks.

In general, it can be noted that recently there has been a steady increase in the number of works devoted to voice assistants' introduction in various areas of human activity. However, most of studies are either of research nature and assess the potential use of voice assistants in the future, or are devoted to information security of this approach, or are focused on the consumer sector and are associated with home automation and mobile applications. There exists virtually no work on the integration of voice assistants to manufacturing production associated with the fourth industrial revolution. In this regard, it could be assume that the research in this direction is relevant and appropriate.

III. Proposed Approach

A. Digital twin of modular CPPS

A digital twin is the basis of the described modular CPPS. An important feature of a digital twin is that the information from sensors of a real device operating in parallel with a twin is used as input influences for it. Work is possible in both online and offline modes. After that, it is possible to compare the information of virtual sensors of a digital twin with sensors of a real device, identifying anomalies and their causes. Differently put, a real production process is combined with its digital model, and any impact on the model leads to a change in the physical process and vice versa.

The modular CPPS is based on a module. It can be a standalone device or a part of a production equipment unit. The concept of modular equipment is discussed in more detail in [6]. According to this concept, the core of a CPPS digital twin is a distributed registry, which is a decentralized database. From a technical point of view, this registry is a binary JSON tree that stores all data on a production process and also allows modules to register in the system and merge to solve common tasks (within an industrial equipment unit), in other words, it is an aggregator of data and services of CPPS [7]. The interaction

Authorized licensed use limited to: St Petersburg Natl Uni of Info Tech Mech & Optics. Downloaded on January 31,2021 at 10:04:20 UTC from IEEE Xplore. Restrictions apply.

between modules is carried out through a mesh network based on the OpenThread protocol [8], [9].

A registry is a collection of nodes, each of which is either a unitary or a modular device. In both cases, all the functions of an equipment unit are determined by a dispatcher, in which modules can be registered. If only one module is registered in a dispatcher, this is a unitary device; if several, it is a modular unit of industrial equipment. Otherwise stated, a dispatcher can be configured to connect a certain number of modules, which can vary from 1 to 32, which is set at the design stage of new equipment integrated into CPPS.

A module is registered in a controller when it is physically connected to it. Each module is allocated to a separate slot. A slot is a structural information unit of a registry that stores information about services, received commands, and data that can be transmitted to a CPPS network. The slot includes the following fields:

- address;
- the name of the module/sensor;
- functions;
- return value;
- the limits of the return value.

A data set from all dispatcher modules determines the parameters of a piece of equipment, the data set from all dispatchers defines CPPS parameters in general.

Robotic systems (warehouses, automated guided vehicles, multi-axis industrial robots), conveyors and production lines, industrial equipment with computer numerical control can be modular. On the other hand, various sensors of a production process, as well as equipment for indication and notification, are unitary devices, i.e. from software implementation point of view, they combine the functions of a dispatcher and one module connected to it.

As an example, modular equipment with computer numerical control (CNC) is considered. Such equipment can be conditionally represented as a set of multi-axis chassis, moving in the space of any module that determines the purpose of an equipment unit, as well as a set of certain external units. For example, equipment for the selective photopolymers curing consists of three modules: a chassis, a laser head and an external source of laser radiation, as well as a set of sensors that determine the return value.

In addition, each of the modules registers its functions as available G-codes and M-functions according to standards ISO 6983-1 and ISO/TR 6983-2. The return value does not have to be just one. The description of each of them includes a type, limits and an array of values with timestamps. The dispatcher controls the output of each value beyond the permissible limits, and in the event of such situation, it analyzes the error that has occurred, and the decision is made whether it is possible or impossible to continue operating the equipment.

B. Voice assistant Integration into modular CPPS

According to the concept presented in the previous section, a voice assistant will be a unitary device operating at the CPPS level. Voice assistant's main function is to receive commands, transfer them to a cloud service for recognition, and also output the result in the form of an audio message. A general architecture of modular CPPS using voice assistants is shown in Figure 1.

As noted earlier, in the registry each module is associated with a set of sensors and the limits of acceptable values for each of them. Therefore, the first scenario of the voice assistant's work is an alert when a parameter is out of acceptable values. This task is multilevel, since the registry structure is flat, in other words, data from all sensors is always available at the CPPS level. Values that exceed permissible range are a violation of CPPS and spread throughout the network as soon as one of the controllers detects this event.

Therefore, the question at what point to carry out the notification of any parameter violation arises. The following solution is proposed. All sensors in the system are ranked according to a degree of their influence on the production process. According to these ratings, a separate custom alert policy has been implemented, which includes the following levels:

- General production level (critical error, all modules are synchronized in time and the message is transmitted via speakerphone).
- Local error of an equipment unit (the message through the closest module, the OpenThread protocol allows determining the neighboring nodes in the mesh network).
- Minor error (the message is not pronounced, but entered into the general log; a signal lamp is lit or a warning is displayed on a remote control).

The ranking is carried out by experts using the analysis of the subject area at the CPPS design stage. From the general registry point of view, an additional level field is added for each sensor with values ranging from 0 to 2 (low, medium, above average, high).

The second scenario of using voice assistant in the presented CPPS structure is to obtain data on the current equipment parameters. According to the accepted model of data representation in the registry, each unitary module and each equipment unit has a unique machine-readable identifier that allows you to search for it in the registry. For the voice assistant, an additional identifier has been added, which is displayed on the screen of the voice assistant module (see the next section). The identifier in this case is a four-digit number unique to the production facility.

Commands which help an operator to request from the assistant a list of all modules or equipment closest to him, if it is out of sight of the operator, are implemented in the system. Other sampling commands for any parameters can also be implemented. For each unitary module or equipment unit, the voice assistant can list all available parameters. As a result the following voice command is built {**keyword for accessing the voice assistant**} {**identifier**} {**parameter**}, for example "Alice, what's the temperature on the 4078 sensor?". Human memory works in such a way that after some time the operator will remember production units identifiers important for him

Authorized licensed use limited to: St Petersburg Natl Uni of Info Tech Mech & Optics. Downloaded on January 31,2021 at 10:04:20 UTC from IEEE Xplore. Restrictions apply.

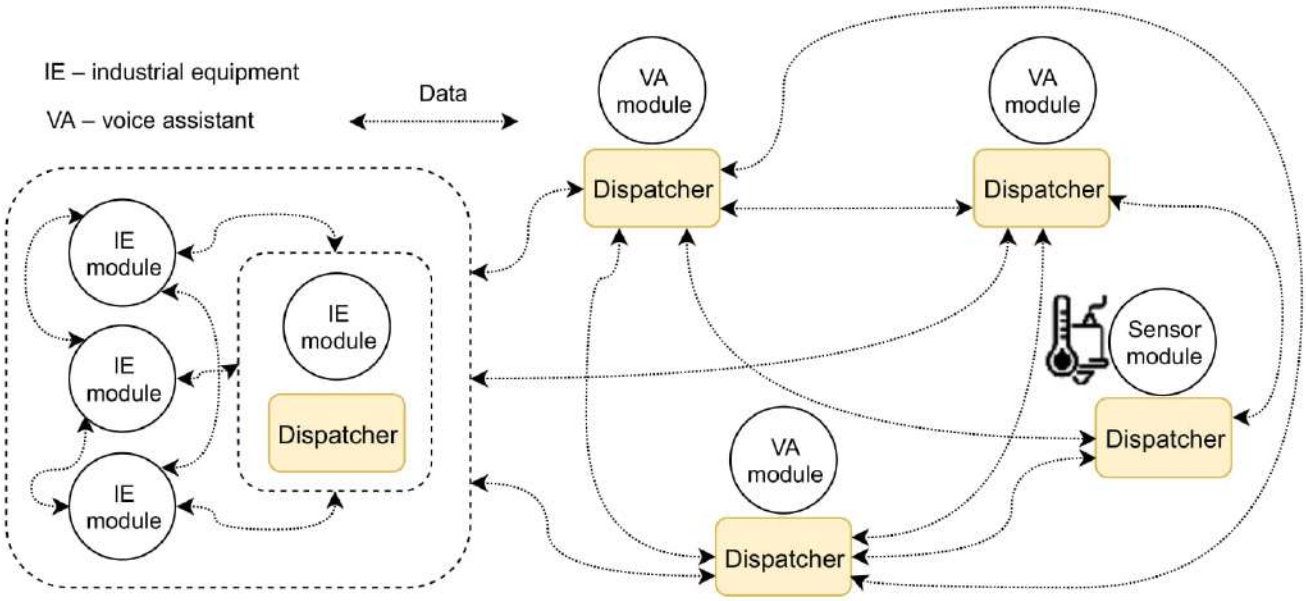

Fig. 1. General architecture of the CPPS with the voice assistant modules.

and the names of parameters that interest him most often and that will significantly increase the speed of data access.

C. Software and hardware implementation of the voice assistant module

A voice assistant module includes a control unit and peripheral devices: a microphone, a dynamic head (speaker) and a display. The module prototype is built on the basis of the Raspberry Pi 3 microcomputer and the nRF52840 Nordic Semiconductor module, which is responsible for the radio channel communication using the OpenThread protocol. An appearance of the described prototype is presented in Figure 2.

The prototype uses ReSpeaker Core v2.0 voice recognition board (https://respeaker.io/rk3229_core/) and the Alice cloud voice service API [10] from Yandex, which allows user to interact with the cloud service using a software library written in the Python programming language. The microcomputer is responsible for receiving data from the microphone, sending messages to the built-in speaker and integrating into the general structure of the CPPS, as well as accessing the Alice cloud service (Figure 3). Also in the structure there is a service for receiving commands from Alice and translating it into requests used in the modular CPPS under consideration (generating a message using the *nanomsg* protocol to fetch data from the registry).

IV. Discussion

Despite a number of obvious advantages of using voice assistants for the operator's operational interaction with CPPS, built based on a modular approach, a range of problems can be identified that need to be solved at this stage of development. Such problems may primarily include:

1) Noise pollution of manufacturing premises.
2) Interaction with several operators simultaneously.
3) Control or monitoring, security issues.

A. Noise pollution of manufacturing premises

It is not a secret for anyone that any industrial production is associated with an increased noise environment, the source of which is primarily industrial equipment. There are standards that regulate a permissible ambient noise in manufacturing. Different maximum allowable noise levels have been established in different countries, but mostly this parameter ranges from 70 to 90 dB [11].

Fig. 2. Appearance of the voice assistant module.

Authorized licensed use limited to: St Petersburg Natl Uni of Info Tech Mech & Optics. Downloaded on January 31,2021 at 10:04:20 UTC from IEEE Xplore. Restrictions apply.

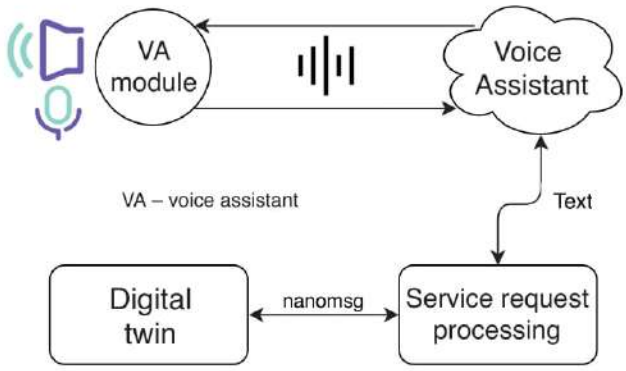

Fig. 3. Integration with Yandex cloud service Alice.

It is obvious that such noise values have a negative impact not only on the hearing organs of operators, but also on microphones used in modules of the considered voice guidance system. Certainly, the correct solution to this problem could be a complete rejection of the background analysis of incoming commands from an operator. Indeed, various technologies of active noise cancellation, filtering and signal amplification are implemented in modern acoustic headsets. The most illustrative in this regard are special solutions used for military purposes but household appliances are often able to provide acceptable performance as well, because the noise level of modern megalopolises often exceeds that in manufacturing.

However, this decision imposes a number of significant restrictions on voice assistants' usage in the workplace. Firstly, the use of headsets reduces the working convenience for an operator, because its constant wearing is associated with a number of disadvantages. People often do not feel comfortable wearing such devices for a long period of time; it reduces the ability to communicate with other operators, which, with industrial noise background added, can lead to emergencies or accidents. Apart from that headsets require constant maintenance, recharging etc. Lastly, headsets must be wireless and interact with CPPS via a separate radio channel (now, this is usually bluetooth). The developed CPPS modular architecture uses the OpenThread wireless protocol [12], which is not designed to transmit audio data, and it is highly undesirable to simultaneously deploy a bluetooth network operating in the same frequency band.

Thus, it becomes obvious that the use of acoustic headsets for interaction between an operator and CPPS is impractical. Instead, it is proposed to add a certain redundancy to the system, which can be implemented with the modular approach used.

As noted earlier, all CPPS modules are located in a common decentralized network, therefore they can communicate with each other and exchange information on commands received and the signal level at the receiving device (microphone) used. Accordingly, if one of the modules detects a voice command from an operator but cannot decipher (understand) the signal, it informs the nearest modules that a signal, different from the background noise, has been received but cannot be recognized. At the same time, it is possible to select one or several modules that will determine the parameters of the background noise of a production room as a whole or its individual parts and record this data into the CPPS database so that other modules can filter the background noise and extract voice commands.

Thus, if one of the modules recognizes a command and can process it, all its neighbors are informed that the command has been received and its processing begins now. Consequently, all voice accompaniment modules become a kind of omnidirectional microphone that perceives commands from the entire production room. At the same time, the quality of commands recognition can be significantly increased with the correct selection of the number of required modules and their location.

It should also be noted that the problem of increased noise exposure is receiving new solutions, and the overall noise level of production equipment is reduced due to improved noise insulation and the use of modern technical tools. All the mentioned tendencies should have a positive impact on the voice assistants implementation in various fields of industrial production in the future.

B. Simultaneous interaction with several operators

The considered approach to voice assistants' introduction in production suggests that such a system of interaction with an operator will not be individual. The voice assistant, in this case, is not an analogue of an operator's console, but a part of CPPS, which any of the operators can interact with by giving commands in the background mode. Clearly, such a system has to effectively recognize voice commands from several people, and it is necessary that the following requirements will be met:

1) The system must be self-taught, i.e. a preliminary recording voice samples from different operators for later recognition is unacceptable.
2) The system should be able to receive commands from different operators in an arbitrary order, otherwise stated, be able to quickly switch from one operator to another.

This problem is partially solved due to the modularity described above, when all voice tracking modules communicate with each other and can control the signal level of each neighboring module. In this case, all the assistants work within a single model (digital twin). In fact, it is difficult to imagine a situation in production process when several operators, being in close proximity to each other, try to interact with the voice assistant. A traditional production facility is divided into independent zones or areas, occupied by only one person.

Nevertheless, work on the recognition and separation of commands from several operators is still being conducted and there are certain positive results. In particular, Hershey et al. [13] describes an approach called deep clustering. The use of this method makes it possible to effectively separate and reconstruct the speech of two people speaking into one microphone, with an accuracy of up to 90%, three—up to 80%. At the same time, an approach similar to the one under consideration makes it possible to achieve an accuracy of 51%

Authorized licensed use limited to: St Petersburg Natl Uni of Info Tech Mech & Optics. Downloaded on January 31,2021 at 10:04:20 UTC from IEEE Xplore. Restrictions apply.

with the use of two microphones connected to the same voice recognition system without measuring the signal level [14].

C. Control or monitoring, security issues

An important issue that needs to be addressed when using voice assistants in production is a safety issue. In this regard, it is necessary to immediately determine which CPPS functions can be implemented in the voice assistant interface, and which functions cannot. It is obvious that, like any other human-machine interface, the voice assistant can be used for both input and output of information. The implementation of the voice output system, in response to the command given by an operator, seems to be quite safe. For the vast majority of industries, the leakage of data related to the production process is not critical, especially if it is related to the current parameters of production facilities and equipment. Much more serious are commands for managing the production process, the parameters of production premises or equipment. Here safety issues come to the foreground. On the one hand, access to the voice assistant is collective and does not require authorization. That is what makes its use so convenient and allows one to reduce the time needed to obtain operational information about a production process. Any kind of authorization would again require the presence of wearable devices associated with an operator that will control access rights to execute certain commands. The obvious solution could be the storage of voice samples of each operator, in other words, the use of biometric parameters, but here rises the question about whether or not the system understands if a command is really being pronounced at the moment or the recording is being played, that is, authentication is needed then. Based on the analysis of the sources, it can be concluded that with the current quality level of the sound reproducing equipment, this problem cannot be solved by software. On the other hand, the issues concerning security and access control at any enterprise are addressed centrally and include, among other things, the control of personnel access to certain premises. In other words, security issues concerning the use of voice assistants could be resolved in a complex and have a multi-level structure. Physical access to equipment should be carried out only through the operator's console using the appropriate authorization mechanisms and commands to control certain non-responsible parameters. In this case an apropriate example is the scenario of using an automated guided vehicle (AGV), when an operator of a production site calls an AGV using a voice command to load finished products. Since each unit of the equipment is associated with its own voice tracking module, AGV automatically obtains the necessary coordinates and moves to a given point. After that, loading takes place and an operator gives a voice command to move the AGV to the next point, to an automatic warehouse for example.

V. Conclusion

Since the advent of studies on speech recognition technology, voice control has been one of the attributes of the life in the future, however, currently, these devices have already become part of the present time. The input audio information processing algorithms continue to improve its quality and recognition speed; there are already a large number of commercial products with extensive functionality, the ability to filter out ambient noise, tune in to the voice of an operator and execute fairly complex commands.

In this paper, aspects of the use of voice assistant were considered as a channel for accessing production data from a CPPS. The introduction of such technology can be justified in large production areas, where it is difficult to quickly find a display with necessary information and get to it. The approaches to the introduction of technology into modular equipment were analyzed, possible operational difficulties and tasks requiring further research were identified.

Acknowledgment

This work was financially supported by Government of Russian Federation, Grant 08-08.

References

[1] N. Lindgren, "Organy chuvstv zhivotnyh i ih ehlektronnye analogi [The organs of sense of animals and their electronic counterparts, in Russian]," *Electronika*, vol. 35, no. 7, pp. 22–27, 1962.

[2] K. Tubbesing. (2018) Alexa is engaged as voice assistant in industry. [Online]. Available: https://www.hannovermesse.de/en/news/alexa-is-engaged-as-voice-assistant-in-industry-90434.xhtml

[3] J. Vajpai and A. Bora, "Industrial applications of automatic speech recognition systems," *Int. Journal of Engineering Research and Applications*, vol. 6, no. 3 (Part 1), pp. 88–95, March 2016.

[4] M. Hoy, "Alexa, Siri, Cortana, and more: An introduction to voice assistants," *Medical Reference Services Quarterly*, vol. 37, pp. 81–88, 01 2018.

[5] N. C. Bradley, T. Fritz, and R. Holmes, "Context-aware conversational developer assistants," in *Proceedings of the 40th International Conference on Software Engineering*, ser. ICSE '18. New York, NY, USA: ACM, 2018, pp. 993–1003.

[6] M. Y. Afanasiev, Y. V. Fedosov, A. A. Krylova, and S. A. Shorokhov, "Modular industrial equipment in cyber-physical production system: Architecture and integration," in *Proceedings of the 21th Conference of Open Innovations Association FRUCT*, November 2017, pp. 3–9.

[7] ——, "An application of microservices architecture pattern to create a modular computer numerical control system," in *Proceedings of the 20th Conference of Open Innovations Association FRUCT*, April 2017, pp. 10–19.

[8] C. Garrido-Hidalgo, D. Hortelano, L. Roda-Sanchez, T. Olivares, M. C. Ruiz, and V. Lopez, "Iot heterogeneous mesh network deployment for human-in-the-loop challenges towards a social and sustainable industry 4.0," *IEEE Access*, pp. 1–1, 05 2018.

[9] M. Y. Afanasiev, Y. V. Fedosov, A. A. Krylova, S. A. Shorokhov, and Z. Ksenia V., "Mesh networking in cyber-physical production systems: Towards modular industrial equipment integration," in *Proceedings of the 23rd Conference of Open Innovations Association FRUCT*, April 2018, pp. 3–11.

[10] (2018) Yandex dialogs documentation. [Online]. Available: https://tech.yandex.ru/dialogs/alice/doc

[11] L. S. Goodfriend and F. M. Kessler, *Industrial Noise Pollution*. Boston, MA: Springer US, 1973, pp. 572–586.

[12] (2018) Openthread node roles and types. [Online]. Available: https://openthread.io/guides/thread-primer

[13] J. R. Hershey, Z. Chen, J. Le Roux, and S. Watanabe, "Deep clustering: Discriminative embeddings for segmentation and separation," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2016, pp. 31–35.

[14] J. C. Murray, S. Wermter, and H. R. Erwin, "Bioinspired auditory sound localisation for improving the signal to noise ratio of socially interactive robots," in *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Oct 2006, pp. 1206–1211.

Authorized licensed use limited to: St Petersburg Natl Uni of Info Tech Mech & Optics. Downloaded on January 31,2021 at 10:04:20 UTC from IEEE Xplore. Restrictions apply.

Interpolation Algorithm for High-Speed Processing of Complex Curvilinear Trajectories

Kseniia Zimenko, Maxim Afanasev, Yuri Andreev,
Anastasiya Krylova, Sergey Shorokhov, Yuri Fedosov, Mikhail Kolesnikov
ITMO University, St. Petersburg, Russia
zksenia@yahoo.com, amax@niuitmo.ru, ysandreev@itmo.ru, {ananasn94, stratumxspb}@gmail.com,
yf01@yandex.ru, km@hexaxis.ru

***Abstract*—The rising popularity of complex product design leads to frequent use of curvilinear toolpaths during machining. Along with growing accuracy requirements for processing the complication of part geometry leads to the importance of developing efficient interpolation methods for processing along complex toolpaths. In this case, to obtain an online interpolation algorithm for complex trajectories the combination methods, such as the use of parametric curve representation for complex paths, approximation into equal linear segments during interpolation and application of curve on trajectory corners, is considered. The paper contains the description and analysis of the developed interpolation algorithm for complex trajectories. The proposed curve approximation into equal segments allows to achieve constant speed and minimal contour error. The comparison between the conventional and proposed segmentation methods have been obtained via simulation and analyzed in the present paper. The developed algorithm allows to significantly reduce the segment length and velocity deviations. The application of an acceleration/deceleration control with S-shape feedrate profile for curve interpolation is also described in the paper. The proposed method allows the calculation of an optimal parameter increment with low computational load, and thus can be applied to real-time machining.**

I. Introduction

The growing popularity and availability of personalized manufacturing, when the design of a product and its functions and features is made strictly according to customer requirements, can be noticed in the field of modern industrial production. More and more enterprises offer a complex and unique product design for their customers.

This tendency leads inevitably to the complication of devices geometry. So despite the fact that during processing a toolpath still mostly consists of linear segments and arcs, complex trajectories such as curves, spirals and involutes are used increasingly more often. Combined with the high required accuracy, the complication of trajectories leads to the problem of developing efficient algorithms for processing along complex toolpaths – curves.

In Computer Numeric Control (CNC) software the axial tool movement generation, based on data obtained from the control program code (G-codes), is performed by the interpolation module [1]. At this stage the toolpath is formed. The processing accuracy depends on how much the initial path obtained from G-codes coincides with the toolpath obtained during interpolation. Therefore, the aim of this project is to develop an efficient interpolation algorithm for complex trajectories that will produce minimal contour error and stable velocities, which thus will ensure high precision processing and high quality of a machined surface.

Most of CNC machines presented on the modern market are equipped with the necessary software, allowing it to process complex trajectories at high tool speeds with minimal machining error. However, these solutions are proprietary and are revealed neither for third-party developers nor for the public. In light of the above, one of the aims of the present project is to create an efficient open-source solution.

Most of the open-source CNC algorithms available imply curve segmentation into unequal path units which leads to inconsistent velocity and, hereby, to the deterioration of a part surface obtained [2]-[4]. Apart from the mentioned above most of the proposed methods are not applicable for high-speed machining due to limitations that implies a machining error which is in direct proportion of a tool speed.

The way to perform segmentation and calculate a curvilinear path so that obtained equal segments ensure the constant feedrate value is proposed in the present paper. Parametric representation of curves is used in the solution. The described interpolation method is combined with the acceleration/deceleration control that generates an S-shaped speed profile and thus provides smooth changes of velocity during processing.

Together with curves, arcs interpolation is also considered in the paper with application of the same approach. The developed interpolation method for trajectories at corners will also be described. This paper presents a description of the methodology, implementation and simulation results.

Among the criteria for the development is maximum simplicity and universality of calculations. Firstly, all proposed calculations are suitable for application in a real-time interpolator and do not have a heavy computational load. Secondly, as opposed to most of the interpolators that are based on a specific type of a curve, the present research works with parametric representation in general, so that the developed method could be applicable to any type of curvilinear trajectory.

The paper describes the algorithm application for machines that are stiff dynamic systems. The influence of dynamic constraints is compensated by the negligible mass of a tool (laser head, extruder) compared to a machine itself.

The paper is organized as follows. Section II is dedicated to the overview of related researches. In Section III the main problem concerning complex trajectory interpolation,

ISSN 2305-7254

Authorized licensed use limited to: St Petersburg Natl Uni of Info Tech Mech & Optics. Downloaded on January 31,2021 at 10:09:56 UTC from IEEE Xplore. Restrictions apply.

segmentation of a curve, is discussed and the proposed solution is introduced. The opportunity to use the developed interpolation algorithm for arc processing and corner smoothing technique are also considered in the paper. Particular properties that circular processing imposes on the proposed algorithm application, are described in Section IV. Section V is given for highlighting the specifications for curves generation at corners. Section VI includes the description of the acceleration/deceleration control developed and the example of its application for curve interpolation. The analysis of the algorithm performance, including possible drawbacks and directions for further improvement and research, is given in Section VII.

II. Related Work

Two main method groups used in interpolation are:

1) Reference-Pulse Interpolators.
2) Reference-Word Interpolators [5].

In the first group mentioned, a computer generates reference pulses as an external interrupt signal. The produced pulses are relegated directly to the machine drive. It can achieve high accuracy but as the velocity on each axis depends on external interrupt signal frequency, high speed machining cannot be obtained and a high performance Central Processing Unit (CPU) is required [5].

The Reference-Word Interpolation implies calculating the next reference point based on the equal sample time T also called interpolation time. The trajectory is divided into small segments that are processed in one T in accordance with the required feedrate. This method is more widely used in modern CNC with servo motors and does not impose limitations on a processing speed [6]. It will be considered in the present paper.

The main idea of interpolation lies in dividing a path into small units, individual for each axis that are processed during the sample time T.

While this method of processing does not cause machining errors along the linear path, its application for complex trajectories brings certain problems including the optimal segmentation of a curve and the contour error value. Main attention in this paper is dedicated to developing an effective interpolation algorithm for complex trajectories that solves this problems.

The step of generating a toolpath and velocity profile based on a given trajectory is a key stage of CNC performance. Therefore, the question of developing an effective motion planning algorithms in both CNC machining and robot manipulators has always received a lot of attention among researches [7]-[9].

Parametric representation of curves finds application in a large number of CNC software algorithms for curve trajectories [9]-[12]. However, the proposed solutions are mostly applicable to specific type of a curve. For example, Daoshan Du et al. [10] described the adaptive parametric curve interpolator which results in high accuracy, designed specifically for NURBS curves. Lijuan Chen et al. in [11] took into consideration the shape preserving interpolation with the use of cubic NURBS curves. Kaan Erkorkmaz and Yusuf Altintas [12] proposed a trajectory generation algorithm that produces continuous position profile and reduced jerk value during processing, based on a quintic spline curve. The present research, on the other hand, does not work with a specific type of curves but aims to provide universality in terms of applying different types of curves. The developed algorithm should be efficient for any curvilinear representation including Bezier, B-spline, NURBS etc.

Most of papers include solutions for parametric path generation during interpolation based on the uniform increment of the curve parameter. Burak Sencer et al. [13] propose the algorithm for corner smoothing with Bezier curve, however the calculating is performed by gradual addition of the curve parameter. Bingran Li et al. [14] described a corner smoothing technique based on the allowable contour error, the solution does not guarantee the equal size of an obtained curve. Other solutions require complex calculations of a curve parameter that are only suitable for offline interpolations. The proposed solution provides the curvilinear approximation into equal segments and the calculations are suitable for the real-time generation.

Lu L. et al. [15] proposes the algorithm of feedrate scheduling for parametric curves processing. However, the method implies its application in an offline mode. One of the aims of the present project is the development of a solution for real time performance. H. Li et al. [16] presents a method for acceleration/deceleration control that allows smooth changes of velocity and creates an S-shape feedrate profile. Although the proposed algorithm could be used in a real time interpolation procedure it is only applied for processing of linear trajectories.

To summarize, after analysis of related researches the task remains develop an interpolation algorithm for curvilinear trajectories that provides uniform path segments and thus stable velocity during processing that does not impose limitations on the type of curve used and is applicable for real-time CNC algorithms.

III. Curve segmentation

The most common way to represent a curve in a CNC is the parametric form, when it is described by the Eq. 1.

$$x = x(u); y = y(u) \tag{1}$$

where u is a parameter that, in most cases, changes in range between 0 and 1. Most CAD/CAM systems represent curves using these equations.

Coordinates of each axis are calculated directly using different equations, which facilitates the transition from two-dimensional to three-dimensional representations and higher. For simplicity, all examples shown in the paper are presented in two dimensional (XY) plane.

When working with complex trajectories, traditional interpolation algorithms produce a preliminary approximation of a curve by linear sections. Each of these sections than in turn is further divided into smaller segments. It leads to an increased contour error. In this case, to avoid inaccuracies, the initial segmentation of a curve should produce length units as close as possible to the initial trajectory. However, this process requires significant computational resources and memory. To solve the problem, in this paper the algorithms that

Authorized licensed use limited to: St Petersburg Natl Uni of Info Tech Mech & Optics. Downloaded on January 31,2021 at 10:09:56 UTC from IEEE Xplore. Restrictions apply.

perform direct calculations of a parametric curve form with single approximation are considered. This way the number of iterations is reduced and processing accuracy is maintained.

The curve segmentation is carried out by successively increasing the parameter u value and calculating the next point to which a tool should arrive per sample time T. To construct a curve in this manner adequately, it is necessary determine the optimal u increment or Δu. The Δu value has to comply with three requirements. Firstly, parameter u must be calculated independently of the geometric parameters of the curve. Secondly, the representation of $x(u)$ and $y(u)$ must be explicit to allow direct calculation of $x(u)$ and $y(u)$. And the last rule by mention, but one of the most important in processing is that each next increment of u should produce an equal in length segment Δl. Since the tool passes one segment in equal sample time T, this rule should ensure that a constant feedrate $V = \Delta l \ / \ T$ is obtained.

Most of the traditional interpolation algorithms meet the first two requirements, but the third condition is usually ignored and presents the notable difficulty. This rule holds only in the case of linear paths and arcs.

In this paper two main strategies to calculating the u increment for curve interpolation are analyzed.

A. Segmentation with equal Δu

The first approach for curve calculation suggests dividing it by incrementing the u parameter gradually to define each next point $[x(u_i), y(u_i)]$. The tool then moves straight from the point $[x(u_{i-1}), y(u_{i-1})]$ to $[x(u_i), y(u_i)]$. And segment Δl is then calculated using the Eq. 2:

$$\sqrt{(x_i - x_{i-1})^2 + (y_i - y_{i-1})^2} \tag{2}$$

It can be concluded that the smaller the increments of parameter u, the bigger number of segments will be generated. This leads to less divergence between to the initial and generated curve. The described segmentation method is used in most interpolation algorithms. However, despite its obvious calculation simplicity, this method has two main drawbacks that limit its use in a CNC system.

The first problem is related to a tool speed during processing. Equal increments Δu do not guarantee that each next segment of a curve Δl will have a equal length. However, as according to the processing standards, a tool passes each segment in an equal period of time (sample time T), therefore it will result in different processing speed during each segment, which results in irregular feedrate and, and as a consequence, to deterioration in the quality of the obtained surface.

The second disadvantage is the problem of finding the optimal parameter increment Δu. There is no specific equation or rule that could determine the exact needed value for each curve. There are convincing arguments in favor of both maximizing, and minimizing increment Δu [17].

On the one hand, Δu should be value as little as possible for the following reasons. Firstly, as was stated above it will ensure more accurate approximation of the original curve. Secondly, it will minimize the segmentation effect, which results in poor surface quality.

However, on the other hand this leads to a large number of curve segments that require significant computation and memory capacity. Apart from that, since the CNC system perceives each segment as a separate processing unit, this leads to a decrease in the average tool speed and consequently to an increase in machining time [17]. In this regard, the number of segments of the curve should be minimized (increase of the Δu value). Between maximization and minimization it is necessary to find the average value, which differs for each next curve. The possible equation for Δu should consider the curvature of a path, the feedrate and the sample time value and should be calculated for each T.

B. Segmentation with equal Δl

The main idea of the second algorithm for curve segmentation is to obtain equal segments of the curve Δl, but not equal Δu. At each interpolation step, the optimal Δu is calculated so as to achieve Δl segments equal in length. This will guarantee an even feedrate during processing. The method does not require much computation time and can be used in real-time interpolators.

The equation for determining the increment at $i - th$ step is calculated based on the given feedrate value, sample time and curve characteristics. The velocity V of the tool moving along the curve can be determined by Eq. 3 with respect to the increment of u:

$$V(u) = \frac{ds}{dt} = \frac{ds}{du}\frac{du}{dt} \tag{3}$$

Then the increment Δu equals to:

$$\frac{du}{dt} = \frac{V}{ds/du} = \frac{V}{\sqrt{(x')^2 + (y')^2}}$$

The final equation for the next u value can be obtained with the use of Taylor's expansion. The first order approximation is possible since the interpolation time usually has small value, the assumption takes the form as shown (Eq. 4).

$$u_{i+1} = u_i + T\dot{u}_i = u_i + \frac{VT}{\sqrt{(x')^2 + (y')^2}} \tag{4}$$

So, to calculate the parameter increment Δu on the step $i + 1$, Eq. 5 can be applied.

$$\Delta u_{i+1} = \frac{VT}{\sqrt{(x')^2 + (y')^2}} \tag{5}$$

Since the algorithm allows obtaining a stable feedrate and solves the problem of determining the optimal increment Δu, the use of this calculation method for interpolating curvilinear trajectories seems appropriate.

To analyze the difference in obtained values after applying the mentioned segmentation methods the example of a curve is considered below.

Authorized licensed use limited to: St Petersburg Natl Uni of Info Tech Mech & Optics. Downloaded on January 31,2021 at 10:09:56 UTC from IEEE Xplore. Restrictions apply.

C. Simulation system characteristics

All given examples have been implemented using the Python programming language version 3.6.5, charts are made with the Matplotlib plotting library version 2.2.2 and calculations are made using Numpy library version 1.14.2. CPU's clock speed is 2.20 GHz, Random Access Memory volume is 4 Gb, operation system is 64-bit Windows 10.

The tool trajectory is represented in the parametric form by the Eq. 6 and is depicted in Fig. 1.

$$\begin{aligned} x &= -90u^3 + 5u^2 + 85u \\ y &= 10\sin(u) \end{aligned} \tag{6}$$

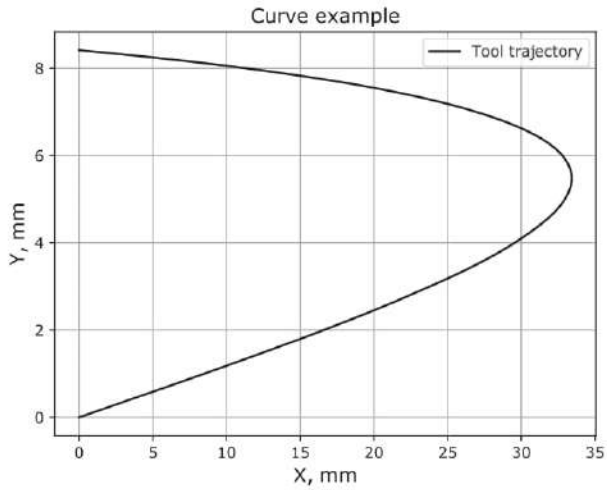

Fig. 1. Tool path example

The initial processing specifications for the simulation are given below.

```
start velocity V = 25 mm/sec;
sample time T = 0.01 sec.
```

In the case of conventional method the length of obtained segments goes from 0.35 to 0.64 mm. This significant change may cause variations in speed up to 98%. The contour error reaches 11.4 μm with the number of segments equal 100.

With the application of the proposed method the contour error stays within 8 μm, the discrepancy between the segment length does not exceed 3%. This results in a stable velocity during processing. The number of linear segments in this example equals 270. Bigger number of segments leads to greater correspondence of the initial given trajectory. It should be noted that in the case of applying traditional segmentation method with the same number of linear units (270), due to the segments length inequality, the contour error is higher than obtained by the proposed algorithm.

Summarizing the obtained results, the proposed approximation method allows to achieve minimal deviations in size of segments obtained during interpolation, which leads to stable velocity during processing.

As a part of the development for complex paths interpolation, the algorithm for arcs processing and corner rounding technique has been modified according to proposed segmentation method to satisfy the feedrate stability and are described in the following sections.

IV. Circular interpolation

Circular interpolation is performed using the proposed above algorithm for curves. The arc is represented in a parametric form determined by the Eq. 7.

$$\begin{aligned} x(u) &= a\cos(u), \\ y(u) &= a\sin(u), \\ 0 &\le u \le 2\pi \end{aligned} \tag{7}$$

where a is a real number. However, in this case the range of the parameter u differs from other types of curves. It depends on the quadrant, where the arc is in, and the direction of the tool movement. Applying the parametric form of an arc in calculation of Δu increment with Eq. 8 could easily show that Δu value depends only on the velocity V and sample time T.

$$u_{i+1} = u_i + \frac{VT}{\sqrt{(-asin(u))^2 + (acos(u))^2}} = u_i + \frac{VT}{a} \tag{8}$$

So on condition that tool speed remains the same, equal segments are obtained with equal u increments. Since the uniform increments of u results in equal segments and the feedrate maintains stable, it is reasonable to calculate the Δu value only once on the first step of arc interpolation process.

For example, the equation for the tool motion along the arc in the first quadrant when moving counterclockwise is described by the Eq. 9 and shown in Fig 2:

$$\begin{aligned} x(u) &= 10\cos(u), \\ y(u) &= 10\sin(u), \\ 0 &\le u \le \frac{\pi}{2} \end{aligned} \tag{9}$$

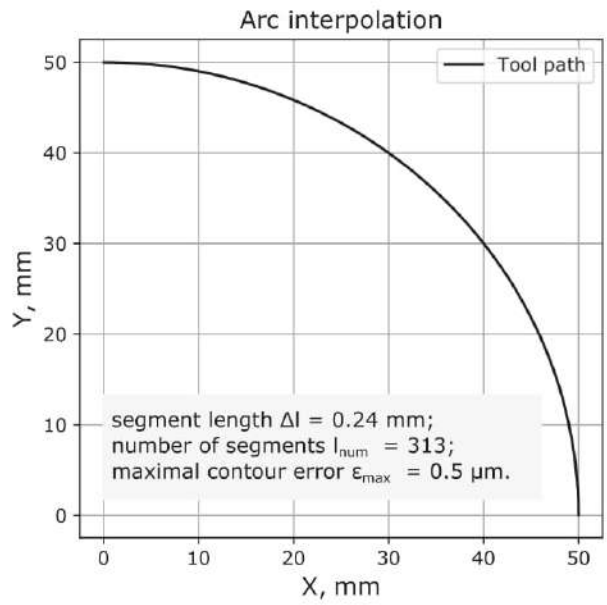

Fig. 2. Arc example

Authorized licensed use limited to: St Petersburg Natl Uni of Info Tech Mech & Optics. Downloaded on January 31,2021 at 10:09:56 UTC from IEEE Xplore. Restrictions apply.

The input simulation parameters are set as follows:

```
feedrate V = 25 mm/sec;
sample time T = 0.01 sec.
```

The simulation of the defined arc results in the uniform sized segments 0.24 mm. The contour error lays within 0.5 μm. As has been mentioned above, the arc interpolation with equal increments Δu results in equal segments. However, the proposed algorithm is still preferable, as it defines the number of segments depending, besides velocity and sample time, on the size of an arc. For instance, in this case the segments quantity equals to 313. On the other hand, if an arc radius equaled to 10 mm instead of 50 mm, the number of segments decreases to 61 with the same contour error value.

V. Trajectory generation at corners

A corner rounding technique was also developed as a part of the proposed interpolation algorithm for complex trajectories. It suggests using parametric Bezier curves, which allows processing without complete tool stops at corners. The possibility to control the size and shape of rounding to achieve the required accuracy of processing and higher flexibility was considered. The calculation of a rounding is carried out based on 6 control points Bezier curve. The detailed rationale for the choice of the curve type and the number of control points for corner smoothing is given in the previous paper [18].

The curve is calculated based on the Eq. 10:

$$B(t) = \begin{bmatrix} B_{x(u)} \\ B_{y(u)} \end{bmatrix} = \begin{Bmatrix} (1-u)^5 P_0 + \\ +5u(1-u)^4 P_1 + \\ +10u^2(1-u)^3 P_2 + \\ +10u^3(1-u)^2 P_3 + \\ +5u^4(1-u)P_4 + u^5 P_5 \end{Bmatrix} and \tag{10}$$

$$P_0 = \begin{bmatrix} P_{0,x}(u) \\ P_{0,y}(u) \end{bmatrix}, P_1 = \begin{bmatrix} P_{1,x}(u) \\ P_{1,y}(u) \end{bmatrix}, ..., P_5 = \begin{bmatrix} P_{5,x}(u) \\ P_{5,y}(u) \end{bmatrix}$$

where $P_0, P_1, \ldots, P_5$ are control points and u changes in range between 0 and 1. The algorithm for corner processing has been modified by applying the segmentation method with equal segments. Partial derivatives x' and y' are calculated as (see Eq. 11):

$$B'(t) = \begin{bmatrix} B_{x'(u)} \\ B_{x'(u)} \end{bmatrix} = \begin{Bmatrix} 5(1-u)^4(P_1 - P_0) + \\ +20u(1-u)^3(P_2 - P_1) + \\ +30u^2(1-u)^2(P_3 - P_2) + \\ +20u^3(1-u)(P_4 - P_3) + \\ 5u^4(P_5 - P_4) \end{Bmatrix} and$$

$$P_0 = \begin{bmatrix} P_{0,x}(u) \\ P_{0,y}(u) \end{bmatrix}, P_1 = \begin{bmatrix} P_{1,x}(u) \\ P_{1,y}(u) \end{bmatrix}, ..., P_5 = \begin{bmatrix} P_{5,x}(u) \\ P_{5,y}(u) \end{bmatrix} \tag{11}$$

Δu is calculated based on the proposed algorithm to achieve equal segment length, according to Eq. 12:

$$\Delta u = \frac{VT}{(B_{x'(u)})^2 + (B_{y'(u)})^2} \tag{12}$$

A trajectory segment with an acute corner has been generated to display the application of the proposed segmentation method for corner rounding (see Fig. 3). Two curves of different shapes, with ratio of n (the ratio of the distance between control points) 0.95 and 0.4, are interpolated using Eq. 11-12. The initial processing parameters are listed below:

```
Trajectory p = [
[0, 0],
[50, 50]
];
feedrate V = 25 mm/sec;
sample time T = 0.01 sec.
```

Results show that the size and number of segments is different depending on the length and curvature of trajectory. However, the contour error is low in both cases and stays within 5.5 μm and the segment length is uniform which allows to achieve stable feedrate and low machining error.

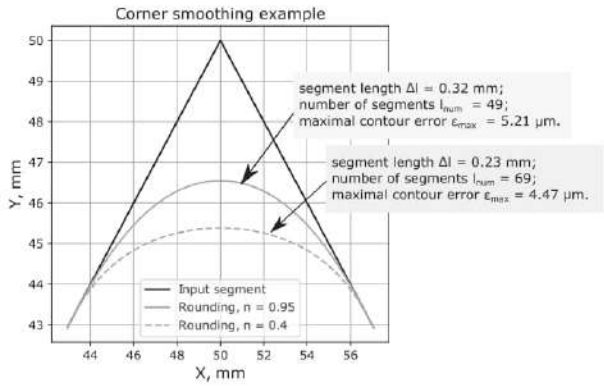

Fig. 3. The example of curves at corner

VI. Acceleration/deceleration control application

It should be noted that due to possible need for acceleration or deceleration of the tool during curve processing, segments size will not remain equal. In this case it is suggested to apply S-shape feedrate profile control to achieve smooth transition of velocity and for vibration and jerk reduction. The detailed description and rationale on the feedrate profile choice as well as the developed algorithm of acceleration/deceleration control has been given in the previous publication [18].

The S-shape feedrate scheduling is performed to minimize the velocity deviations and avoid the exceeding of the allowable acceleration A_{max} and deceleration D_{max}. The velocity on every sample time T is calculated based on the given feedrate F as the first derivative of acceleration A with $T \in [t_0, t_1]$. The following Eq. 11:

$$A = \frac{A_{max}}{2} \cdot \left(1 - \cos\left(\frac{2\pi A_{max}}{2F}\right) \cdot T\right) \tag{13}$$

and deceleration D with $T \in [t_2, t_3]$ as in Eq. 14:

$$D = -\frac{D_{max}}{2} \cdot \left(1 - \cos\left(\frac{2\pi D_{max}}{2F}\right) \cdot (T - t_2)\right) \tag{14}$$

For example, if the tool needs to accelerate from 20 to 25 mm/sec while processing the curve from the previous example

Authorized licensed use limited to: St Petersburg Natl Uni of Info Tech Mech & Optics. Downloaded on January 31,2021 at 10:09:56 UTC from IEEE Xplore. Restrictions apply.

(see Fig. 1) with the allowable acceleration and deceleration $A_{max} = D_{max} = 5$ mm/sec, the feedrate profile will look as shown in Fig. 4. The obtained velocity values on every sample time could be used for interpolation procedure. The results show that segment length in this case will vary from 0.2 to 0.3 mm, and contour error will be within 7 μm.

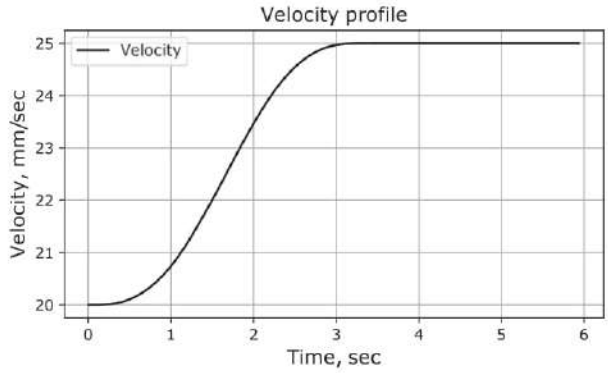

Fig. 4. The obtained feedrate profile

To summarize, the acceleration/deceleration algorithm with S-shape feedrate profile can be applied for complex trajectories, creates a smooth changes of velocity and provides a high quality surface at high speeds. Its application for curve interpolation does not affect the contour error that remains minimal.

VII. Analysis of the Algorithm Performance

The following aims have been achieved during the present research. The proposed interpolation algorithm allows to perform the segmentation of curvilinear paths resulting in segments of equal value by calculating the optimal increments of a parametric value u. It ensures the stable velocity value during processing. The obtained equations for calculations do not have heavy computational load and thus are applicable for the real-time interpolators. The proposed method is universal can be used for interpolation of any parametric curve, including NURBS, B-splines etc. The application of acceleration/deceleration algorithm prior to the interpolation results is possible to ensure smooth transitions of velocity.

To simplify the calculations it is reasonable to perform linear interpolation with the use of Reference-Word method and apply traditional representation method of linear trajectories instead of possible parametric representation, since the contour error does not occur in this type of path.

One of the main drawbacks of the proposed interpolation algorithm is the fact that the segment length is calculated independently of the parameters of a CNC system. Since every machine has different maximum accuracy, basic length-unit (BLU) size and other specifications the equation should be modified so that it would be possible to adjust the minimal segment length according to the actual hardware requirements.

Another problem which arises during the usage of the segmentation method is related to its dependency on the geometric parameters of a curve. It means that the number of segments are directly proportional to the size of a toolpath. This may cause low accuracy when working with small scale trajectories. The reduction of geometric characteristics impact on the process of segmentation may serve as a direction for further research.

The algorithm is implemented in the Python programming language (link to repository: `github.com/stwinter2014/Interpolation-algorithm-for-complex-trajectories`).

To sum up, the developed algorithm allows to obtain uniform sized curve segments during interpolation, which results in stable feedrate and ensures minimal contour error. The application of the feedrate control creates smooth changes of velocity on the acceleration and deceleration stages where needed. However, several improvements should be considered like the decrease of geometrical parameters (e.g. curve size) impact on segmentation and the opportunity to adjust the segmentation process according to CNC characteristics and required accuracy.

VIII. Conclusion

To summarize, an efficient interpolation algorithm for curvilinear trajectories has been developed. The use of parametric curves with the proposed segmentation method allows to calculate the optimal parameter increment resulting in equal curve interpolation units and thus maintain a stable velocity during processing. The use of the acceleration/deceleration control based on the S-shape feedrate profile creates gradual change of velocity during processing and could be applied for curve interpolation.

The performed simulation of the algorithm implementation showed its effectiveness and compliance with the initial requirements. The deviations of segment length do not exceed 3% and produce minimal contour error during simulations. The work is carried out as a continuation of a research published in Proceedings of the 23rd Conference of Open Innovations Association FRUCT in 2018 [18].

It is planned to integrate the present algorithm with the linear interpolation and feedrate control stages. A modification of the proposed segmentation method, so that the increment value could be restricted according to the CNC machine characteristics and accuracy requirements for any specific processing, is also taken into consideration. At the end of development, tests on a CNC machine will be carried out.

Acknowledgment

This work was carried out under the project no. 619296 "Technologies of cyber-physical systems: management, computing, security" conducted at the Faculty of Control Systems and Robotics, ITMO University.

References

[1] P. Smid, *CNC Programming Handbook: A Comprehensive Guide to Practical CNC Programming*, ser. Industrial Press eBooks. Industrial Press, 2003.

[2] Smoothieware official website, Smoothieboards, Web: http://smoothieware.org/smoothieboard.

[3] Linux CNC official website, Lunix CNC documents, Web: http://linuxcnc.org/documents/.

[4] Grbl, GitHub Grbl repository, Web: https://github.com/grbl/grbl.

Authorized licensed use limited to: St Petersburg Natl Uni of Info Tech Mech & Optics. Downloaded on January 31,2021 at 10:09:56 UTC from IEEE Xplore. Restrictions apply.

[5] S.-H. Suh, S.-K. Kang, D.-H. Chung and I. Stroud, *Theory and Design of CNC Systems, 1st ed*. Springer Publishing Company, Incorporated, 2008.

[6] A. Hughes, *Electric Motors and Drives: Fundamentals, Types and Applications*. Elsevier Science, 2005.

[7] L.-P. Luo, C. Yuan, R.-J. Yan, Q. Yuan, J. Wu, et al., "Trajectory planning for energy minimization of industry robotic manipulators using the Lagrange interpolation method", *International Journal of Advanced Manufacturing Technology*, vol. 16, issue 5, May 2015, pp. 911–917.

[8] A. Gasparetto and V. Zanotto, "A New Method for Smooth Trajectory Planning of Robot Manipulators", *Mechanism and Machine Theory*, vol. 42, no. 4, 2007, pp. 455-471.

[9] J. Wu, H. Zhou, X. Tang, J. Chen, "Fast NURBS interpolation based on the biarc guide curve", *The International Journal of Advanced Manufacturing Technology*, vol. 58, issue 5-8, Jan. 2012, pp. 597–605.

[10] D. Du, Y. Liu, C. Yan, C. Li, "An accurate adaptive parametric curve interpolator for nurbs curve interpolation", *International Conference on Information Computing and Applications*, vol. 32, issue 9-10, April 2007, pp. 999–1008.

[11] L. Chen, X. Zhang, M. Li, "Cubic nurbs interpolation curves and its convexity", *The International Journal of Advanced Manufacturing Technology*, 2010, pp. 488-495.

[12] Kaan Erkorkmaz and Yusuf Altintas, "High speed cnc system design. Part I: jerk limited trajectory generation and quintic spline interpolation", *International Journal of Machine Tools and Manufacture*, vol. 41, issue 9, July 2001, pp. 1323-1345.

[13] B. Sencer, K. Ishizaki, and E. Shamoto, "A curvature optimal sharp corner smoothing algorithm for high-speed feed motion generation of nc systems along linear tool paths", *International Journal of Advanced Manufacturing Technology*, vol. 76, no. 9-12, 2015, pp. 1977–1992.

[14] B. Li, H. Zhang, P. Ye, "Error constraint optimization for corner smoothing algorithms in high-speed CNC machine tools", *International Journal of Advanced Manufacturing Technology*, vol. 99, issue 1-4, Oct. 2018, pp. 635–646.

[15] L. Lu, L. Zhang, S. Ji, Y. Han, J. Zhao, "An offline predictive feedrate scheduling method for parametric interpolation considering the constraints in trajectory and drive systems", *International Journal of Advanced Manufacturing Technology*, vol. 83, no. 9-12, April 2016, pp. 2143–2157.

[16] H. Li, Z. Gong, W. Lin, and T. Lippa,"Motion profile planning for reduced jerk and vibration residuals", *SIMTech technical reports*, vol. 8, no. 1, 2007, pp. 32–37.

[17] Y. Koren, C. C. Lo, M. Shpitalni, "CNC interpolators: algorithms and analysis", *Manufacturing Sciem:e and Engineering*, vol. 64, 1993, pp. 83-92.

[18] K.V. Zimenko, M.Y. Afanasev, A.A. Krylova, S.A. Shorokhov, and Y.V. Fedosov "Motion Profile Control Algorithm and Corner Smoothing Technique for Trajectory Optimization of High-Precision Processing", *Proceedings of the 23rd Conference of Open Innovations Association FRUCT*, 2018, pp. 425-431.

Authorized licensed use limited to: St Petersburg Natl Uni of Info Tech Mech & Optics. Downloaded on January 31,2021 at 10:09:56 UTC from IEEE Xplore. Restrictions apply.

УНИВЕРСИТЕТ ИТМО

НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ
сентябрь–октябрь 2020 Том 20 № 5 ISSN 2226-1494 http://ntv.ifmo.ru/
SCIENTIFIC AND TECHNICAL JOURNAL OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS
September–October 2020 Vol. 20 No 5 ISSN 2226-1494 http://ntv.ifmo.ru/en/

НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ

УДК 681.51 doi: 10.17586/2226-1494-2020-20-5-729-738

ПОМЕХОЗАЩИЩЕННОСТЬ БЕСПРОВОДНЫХ ПЕРСОНАЛЬНЫХ СЕТЕЙ В УСЛОВИЯХ ЦИФРОВОГО ПРОИЗВОДСТВА

М.Я. Афанасьев[a], Ю.В. Федосов[a,b], А.А. Крылова[a,c], С.А. Шорохов[a], К.В. Зименко[a]

[a] Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация
[b] ОАО «Российский институт мощного радиостроения», Санкт-Петербург, 199048, Российская Федерация
[c] ООО «ЛАР Технологии», Санкт-Петербург, 197342, Российская Федерация
Адрес для переписки: ananasn94@gmail.com

Информация о статье
Поступила в редакцию 16.06.20, принята к печати 07.08.20
Язык статьи — русский

Ссылка для цитирования: Афанасьев М.Я., Федосов Ю.В., Крылова А.А., Шорохов С.А., Зименко К.В. Помехозащищенность беспроводных персональных сетей в условиях цифрового производства // Научно-технический вестник информационных технологий, механики и оптики. 2020. Т. 20. № 5. С. 729–738. doi: 10.17586/2226-1494-2020-20-5-729-738

Аннотация

Предмет исследования. Исследовано влияния факторов производственной среды на беспроводные персональные сети. Представлена их классификация, определена помехозащищенность беспроводных персональных сетей на примере реального производства. **Метод.** Предложен способ оценки помехозащищенности на основе показателя уровня принимаемого сигнала, значения которого можно получить нативно с любых приемников и передатчиков, что делает этот способ доступным по сравнению с использованием сетевых анализаторов и другого специализированного оборудования. В проводимом эксперименте приемник и передатчик располагались друг от друга на разных расстояниях в диапазоне 0,5–25 м. Акт передачи проводился поочередно под влиянием каждого из факторов производственной среды. Измеренные значения показателя уровня принимаемого сигнала анализировались и преобразовывались в максимально допустимое расстояние между приемником и передатчиком в соответствии с предложенным способом. **Основные результаты.** Получены данные о влиянии производственной среды на помехозащищенность беспроводных персональных сетей, к которым относятся: опоясывающие толстостенные стальные препятствия, сварочные аппараты и сети аналогичного частотного диапазона. Сделан вывод о том, что влияние не настолько значительное, чтобы отказываться от применения беспроводных персональных сетей, так как воздействие многих факторов может быть нивелировано применением ячеистой топологии и плотным расположением приемников и передатчиков. **Практическая значимость.** Результаты работы особенно интересны в контексте цифровизации производства, где из-за требования к гибкости и мобильности производственного процесса беспроводной способ передачи данных с датчиков полевого уровня становится предпочтительнее проводного.

Ключевые слова
цифровое производство, производственная среда, беспроводные персональные сети, оценка помехозащищенности, показатель уровня принимаемого сигнала, силовое оборудование

doi: 10.17586/2226-1494-2020-20-5-729-738

NOISE IMMUNITY OF WIRELESS PERSONAL AREA NETWORKS UNDER DIGITAL PRODUCTION CONDITIONS

M.Ya. Afanasiev[a], Yu.V. Fedosov[a,b], A.A. Krylova[a,c], S.A. Shorokhov[a], K.V. Zimenko[a]

[a] ITMO University, Saint Petersburg, 197101, Russian Federation
[b] Russian Institute for Power Radioengineering (JSC RIPR), Saint Petersburg, 199048, Russian Federation
[c] LLC Lar Technologies, Saint Petersburg, 197342, Russian Federation
Corresponding author: ananasn94@gmail.com

Article info
Received 16.06.20, accepted 07.08.20
Article in Russian

For citation: Afanasiev M.Ya., Fedosov Yu.V., Krylova A.A., Shorokhov S.A., Zimenko K.V. Noise immunity of wireless personal area networks under digital production conditions. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2020, vol. 20, no. 5, pp. 729–738 (in Russian). doi: 10.17586/2226-1494-2020-20-5-729-738

Abstract

Subject of Research. The paper considers the effect of working-environment factors on wireless personal networks. A classification of such factors is given and the noise immunity of wireless personal networks is determined on a real production example. **Method.** A method for noise immunity evaluation was proposed based on a received signal strength indicator (RSSI). RSSI values can be obtained natively from almost any receiver and transmitter, that makes this method affordable compared to application of network analyzers and other specialized equipment. In the carried out experiment the receiver and transmitter were located at a distance ranging from 0.5 to 25 m. The act of signal transfer was carried out alternately under the impact of each working-environment factor. Then the measured RSSI values were analyzed and converted into the maximum permissible distance between the receiver and the transmitter in accordance with the proposed method. **Main Results.** Data on the working-environment effect on the noise immunity of wireless personal networks is obtained. The most significant factors are: girded thick-walled steel obstacles, welding machines and similar frequency range networks. Nevertheless, it is concluded that the effect is not significant enough to decide against the application of wireless personal networks, since the exposure of many factors can be offset by the use of mesh topology and dense arrangement of receivers and transmitters. **Practical Relevance.** The results are of particular interest in the context of production digitization, where the wireless method of data transmission from the field level sensors becomes preferable to the wired one due to the requirement for flexibility and mobility of a production process.

Keywords

digital production, industrial environment, wireless personal area network, noise immunity evaluation, received signal strength indicator, power electronic equipment

Введение

Промышленное производство неизменно считается наукоемким и высокотехнологичным направлением деятельности человека. Начиная с появления паровых машин в начале XVIII века, производство прошло через три промышленные революции, вбирая в себя новые технологии. Современная промышленность находится у следующей ступени развития (Индустрия 4.0) и претерпевает цифровую трансформацию. Развитие технологий промышленного интернета вещей и индустриальных киберфизических систем привело к появлению новой концепции — цифровое производство. Основной идеей цифрового производства является глубокая интеграция компьютерных технологий в производственный процесс, позволяющая создавать цифровые модели (или «цифровые двойники») и воздействовать на них в реальном времени. Однако для проведения подобного моделирования необходимы полные и актуальные данные обо всех элементах и участниках производственного процесса. В настоящее время активно разрабатываются системы класса MDC/MDA (Machine Data Collection/Machine Data Acquisition), позволяющие собирать данные такого рода. Вместе с ними развиваются и другие системы, которые позволяют не только осуществлять сбор данных, но и обеспечивать межмашинное взаимодействие (Machine-to-Machine, M2M) на нижних уровнях производства. Внедрение вышеуказанных систем на производстве сопровождается созданием соответствующей сетевой инфраструктуры.

Развертывание сетевой инфраструктуры влечет за собой значительные финансовые, трудовые и временные издержки, поэтому многие компании не торопятся переходить на цифровое производство, ожидая подходящего решения, которое позволило бы минимизировать эти издержки. Вместе с этим меняется рынок товаров и услуг, и подходы, которые использовались раньше при организации производственного процесса, устаревают [1]. Производство становится более персонализированным, что приводит к появлению нового подхода — «производство как услуга». В результате возникает необходимость обеспечения гибкости и мобильности производственного процесса. Исходя из этого, применение беспроводных сетей оказывается предпочтительным, поскольку приводит к меньшим издержкам [2], а также обеспечивает выполнение новых требований, предъявляемых к организации производственного процесса.

На производстве существует множество факторов, которые могут сказываться на помехозащищенности беспроводного сигнала, поэтому проводятся различные исследования влияния производственной среды на беспроводной сигнал, и определяются критерии его оценки. Так, например, ученые Линчепингского университета и университета Евле в Швеции оценили влияние электромагнитного шума в цехах для различных частотных диапазонов беспроводного сигнала [3, 4]. В своем исследовании они определили профиль задержки мощности в различных помещениях на производстве, отличающихся по характеру отражения и поглощения сигнала. Похожие исследования проводились и в Пекинском университете Цзяотун [5], где определялись амплитудные и временные характеристики электромагнитного шума в диапазонах 315, 433 и 916 МГц, возникающих в ходе эксплуатации сварочного робота. В работе[1] приводится описание установки для проведения измерений параметров беспроводного сигнала диапазона 2,4 ГГц, а также методика измерений. Используя данную установку, ученые получили зависимость стабильности беспроводного сигнала от времени, необходимого на передачу пакета стандарта[2]. Работа [6] затрагивает вопрос зашумления диапазона 2,4 ГГц другими сетями и источниками помех. В ней предлагается математическая модель электромагнитного шума в этом диапазоне, которая может использовать-

[1] Girs Svetlana et al. Design of Channel Measurement Guidelines for Characterization of Wireless Industrial Environments [Электронный ресурс]. Режим доступа: https://www.diva-portal.org/smash/get/diva2:638002/FULLTEXT02.pdf, свободный. Яз. англ. (дата обращения: 31.05.2020).

[2] [Электронный ресурс]. Режим доступа: http://www.rovdo.com/ieee-802-15-4-networks, свободный. Яз. рус. (дата обращения: 31.05.2020).

ся для предварительной оценки помехозащищенности на производстве.

Беспроводные персональные сети

Производство принято представлять как иерархическую систему, состоящую из уровней (рис. 1). На каждом уровне присутствуют элементы сетевой инфраструктуры, обеспечивающие как горизонтальную, так и вертикальную передачу данных. При этом от уровня к уровню типы применяемых вычислительных сетей меняются. Это происходит из-за того, что в зависимости от положения в иерархической системе выдвигаются различные требования к скорости передачи данных, защищенности, ширине канала, топологии, надежности и энергоэффективности. Более того, в зависимости от выполняемых задач, на одном уровне могут сосуществовать различные вычислительные сети. Из этого следует, что сетевая инфраструктура предприятия является гибридной.

Для того чтобы понять на каких уровнях функционирует беспроводная персональная сеть, нужно определить «цифровых двойников», которые оперируют передаваемыми в сети данными. Например, данные, получаемые от станка с числовым программным управлением и/или от установленных на него датчиков, могут применяться для создания «цифрового двойника» этого оборудования. Также хорошим примером может послужить ситуация, когда для некоторых технологических операций в цеховом помещении необходимо поддерживать определенную температуру и влажность. В этом случае данные будут использоваться для «цифрового двойника» цехового помещения.

Беспроводные персональные сети (Wireless personal area network, WPAN) применяются на уровнях системы управления, оборудования, производственной ячейки, цеха и, реже, в случае малых производств, на уровне предприятия. Такой охват и возможность применения в производственной среде они получили по причине активной поддержки комитетом IEEE (Institute of Electrical and Electronics Engineers), что способствовало получению целого ряда технологий, охватываемых набором стандартов1. К ним относятся известные технологии — Bluetooth, Zigbee и Thread. Вместе с тем быстрыми темпами наращивается аппаратная база. Сейчас на рынке присутствует большой выбор чипов и готовых плат, поддерживающих сразу несколько технологий WPAN. Несмотря на изначально небольшую дальность прохождения сигнала, применение ячеистой топологии позволяет достигнуть практически неограниченного радиуса действия. Также стоит отметить возможность построения WPAN со стеком TCP/IP (Transmission Control Protocol/Internet Protocol), что является удобным при совместной работе с другими сетями.

К основным достоинствам WPAN можно отнести сокращение издержек на проводную коммутацию. Это связано с тем, что часто производственные помещения занимают очень большие площади (около 1 км^2), а диспетчерские и аппаратные помещения сильно удалены от рабочих.

Реализация проекта по организации проводной сети, пуско-наладочных работ и обслуживания требует высокой квалификации персонала. Ввод в эксплуатацию проводной сети занимает больше времени, чем беспроводной, так как появляется необходимость прокладывать как сетевые кабели, так и кабели питания для коммуникационного оборудования.

Беспроводные сети с ячеистой топологией создают избыточность путей прохождения данных по сети, повышая отказоустойчивость сети в целом. Появляется возможность простого масштабирования беспроводных сетей, которые являются энергоэффективными благодаря существующим протоколам обмена, например, стандарта[2].

К недостаткам WPAN можно отнести то, что диапазоны частот, в которых функционируют такие сети — 868 МГц, 915 МГц и 2,4 ГГц — в России являются нелицензируемыми и имеют ограничения на мощность передатчика[3]. Такие диапазоны носят название «ISM-частот» (Industrial Scientific Medical). Как следует из названия, в таких диапазонах частот работает достаточно большое количество оборудования, что может вносить помехи в работу сети [7, 8].

Оценка помехозащищенности беспроводного сигнала

На производстве существует множество факторов, которые могут ослаблять и искажать беспроводной сигнал. Определение этих факторов перед процедурой развертывания беспроводной сети является важным мероприятием. На рис. 2 представлена классификация таких факторов по основным признакам. При проектировании нового предприятия многие из указанных факторов можно определить еще на этапе разработки схемы размещения оборудования и компоновки подразделений. Например, тип среды можно определить по функциональному назначению помещения. А зная готовую компоновку подразделений, возможно выявить, где офисные Wi-Fi-сети 2,4 ГГц будут пересекаться с производственной беспроводной сетью. Учитывая это, есть возможность провести корректировки для того, чтобы повысить помехозащищенность сети.

Однако более актуальной в данный момент задачей является внедрение беспроводной сети на уже функционирующем производстве. Кроме того, для получения более точной картины необходимы измерения непосредственно на предприятии. Существуют различные характеристики, которые используются для определения качества сигнала. К ним относятся: профиль ослабления мощности, амплитудные и временные

[1] [Электронный ресурс]. Режим доступа: http://www.rovdo.com/ieee-802-15-4-networks, свободный. Яз. рус. (дата обращения: 31.05.2020).

[2] [Электронный ресурс]. Режим доступа: http://www.rovdo.com/ieee-802-15-4-networks, свободный. Яз. рус. (дата обращения: 31.05.2020).

[3] Заседание ГКРЧ от 28.04.2008 (Протокол № 08-24) [Электронный ресурс]. Режим доступа: https://digital.gov.ru/ru/documents/4035/, свободный. Яз. рус. (дата обращения: 31.05.2020).

Рис. 1. Уровни производственной системы

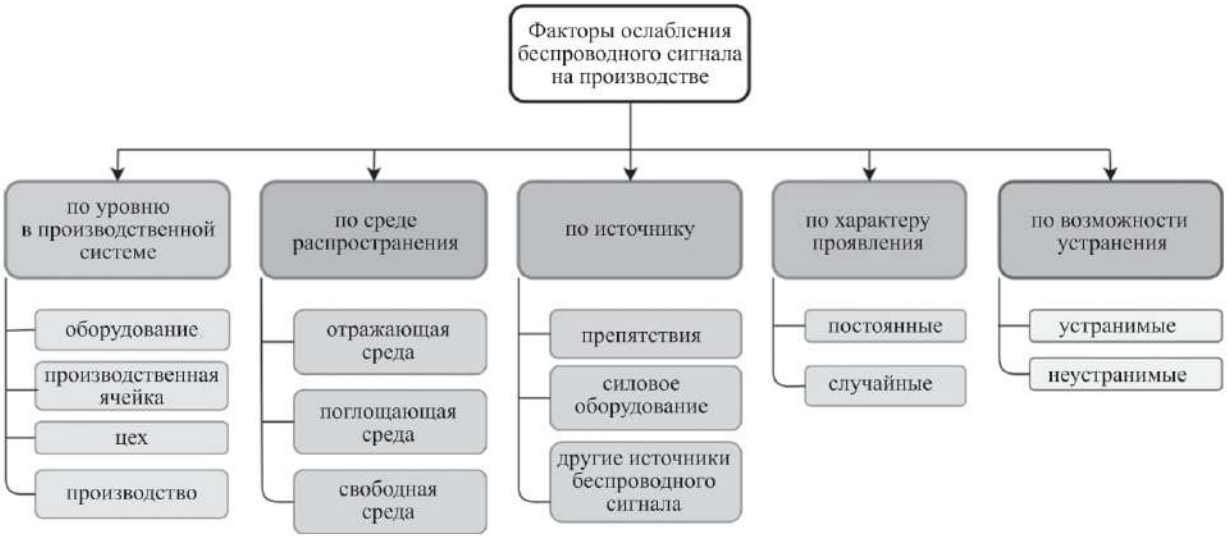

Рис. 2. Факторы ослабления беспроводного сигнала на производстве

характеристики, соотношение сигнал-шум (Signal to Noise Ratio, SNR) [9] и т. д. К сожалению, получение значений этих характеристик требует применения дорогостоящих приборов: анализатора спектра, векторного анализатора сети и специализированных антенн, а требования к квалификации человека, выполняющего такие измерения, оказываются очень высокими.

Методика исследования уровня помехозащищенности

В предлагаемой методике используется показатель уровня принимаемого сигнала (Received Signal Strength Indicator, RSSI). Его основным достоинством является нативная поддержка практически на любом сетевом оборудовании, в том числе на используемой в эксперименте аппаратной платформе. Это означает, что получить значение RSSI можно от приемника без использования дополнительного измерительного оборудования. Тем не менее использовать RSSI напрямую для измерения качества беспроводного сигнала нельзя, поскольку этот параметр может содержать значительную составляющую помех. Другими словами, сильные помехи могут приводить к потере пакетов, одновременно повышая значение RSSI [10].

Значение отклонения $\Delta RSSI$, полученного как разница измеренного значения $RSSI_p$ и теоретического $RSSI_T$:

$$\Delta RSSI = RSSI_T - RSSI_p. \quad (1)$$

При этом $RSSI_T$ рассчитывают на основе характеристик антенн приемника и передатчика, а также учитывая частоту беспроводного сигнала и справочный параметр затухания среды:

$$RSSI_T = A - 10\mu\lg(d), \quad (2)$$

где d — расстояние до источника, м; μ — экспонента затухания, $\mu = 2$;

$$A = P_{\text{out}} + G_{tx} + G_{rx} - FSPL,$$

где P_{out} — выходная мощность передатчика, дБм; G_{tx} и G_{rx} — коэффициенты усиления антенны источника и приемника соответственно, дБи; $FSPL$ — затухание в свободном пространстве, дБ;

$$FSPL = 10\lg(d) + 20\lg(f) + 20\lg\left(\frac{4\pi}{c}\right),$$

где f — частота, Гц; $c = 299792458$ м/с.

При проектировании сети, особенно ячеистой топологии, в первую очередь определяют то, как будут

располагаться узлы в помещениях. В зависимости от отклонения $\Delta RSSI$ можно перейти к более удобному для такой задачи показателю – максимальному расстоянию между узлами $D_{\max}$:

$$D_{\max} = 10^{\frac{A - RSSI_T}{\mu 10}}.$$

Сигнал со значением RSSI, меньше –80 дБм, считается слабым[1]. Исходя из этого утверждения и выражения (1), можно рассчитать $RSSI_T$:

$$RSSI_T = -80 - \overline{\Delta RSSI},$$

на определенном расстоянии от передатчика, с учетом среднего отклонения $\overline{\Delta RSSI}$, на величину которого влияют факторы конкретной производственной среды:

$$\overline{\Delta RSSI} = \frac{1}{n}\sum_{i=1}^{n}\Delta RSSI_i,$$

где n — количество проведенных измерений.

План эксперимента

Эксперимент проводился с целью определения значимых факторов, ослабляющих беспроводной сигнал в производственной среде. Для этого необходимо было собрать экспериментальные данные, а также оценить помехозащищенность WPAN по разработанной методике под влиянием заданных факторов.

Измерения проводились с помощью экспериментальной установки, представленной на рис. 3.

В ее состав входят две платы на базе чипа nRF52840[2], которые выступают в роли приемника и передатчика, на них установлены чип-антенны, характеристики которой указаны в табл. 1. Кроме того, используется персональный компьютер (ПК), на котором выполняется логирование данных. Для получения измеренных значений RSSI ПК соединяется с приемником по интерфейсу UART (Universal Asynchronous Receiver-Transmitter — универсальный асинхронный приемопередатчик). Внутри WPAN передача сообщения происходит по протоколу Thread[3].

Эксперимент выполнен на территории производственных помещений компании «Лар Технологии», специализирующейся на создании технологического оборудования, и проводился отдельно под влиянием каждого фактора ослабления сигнала: фактор отражающей среды (склад металлопроката); наличие силового оборудования (асинхронные, шаговые двигатели, аппарат TIG-сварки (tungsten inert gas)); наличие препятствий (стальная толстостенная труба с толщиной стенки 20 мм). Исследования проводились при наличии

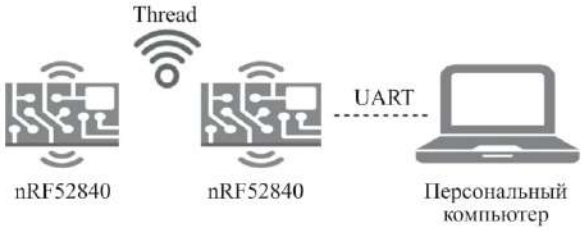

Рис. 3. Схема экспериментальной установки

Таблица 1. Спецификация чип-антенны ACA-2012-A1-CC-S

Характеристика	Значение
Диапазон частот, МГц	2400–2483
Коэффициент стоячей волны	менее 3,0
Поляризация	линейная
Входное сопротивление, Ом	50
Усиление, дБи	1,72
Габариты, мм	2,0 × 1,2 × 0,55

сетей 2,4 ГГц (сети Wi-Fi и Zigbee) и для сравнения измерения — на открытом пространстве без влияния перечисленных факторов.

Значения RSSI фиксировались на определенном расстоянии 0,5–24,5 м с шагом в 1 м. На каждом шаге проводилась серия замеров в течение одной минуты с периодом в 1 с. Для статистического анализа результатов эксперимента применен метод выборочных обследований.

Результаты эксперимента

По мере увеличения расстояния между приемником и передатчиком значение RSSI уменьшается логарифмически в зависимости от теоретического значения $RSSI_T$ (2).

Поскольку распространяемая волна имеет сферическую форму, то с увеличением расстояния энергия распределяется на возрастающую площадь поверхности сферы[4]. На рис. 4, *а* приведен график, отражающий теоретическую кривую значений RSSI, полученную в соответствии с (2), и зависимость RSSI от расстояния при измерении на открытом пространстве. Можно отметить, что кривые оказались схожими по форме. Однако имеются небольшие отклонения RSSI (рис. 4, *б*).

Одной из отличительных характеристик промышленного производства является тип среды, в которой должен распространяться беспроводной сигнал. Чаще всего производство характеризуется отражающей средой с определенными исключениями, например, целлюлозно-бумажные и деревообрабатывающие предприятия, где преобладают поглощающие среды. Распространение сигнала в отражающей среде подвер-

[1] Mobile Signal Strength Recommendations [Электронный ресурс]. Режим доступа: https://wiki.teltonika-networks.com/view/Mobile_Signal_Strength_Recommendationsl, свободный. Яз. англ. (дата обращения: 31.05.2020).

[2] nRF52840 MDK USB Dongle [Электронный ресурс]. Режим доступа: https://wiki.makerdiary.com/nrf52840-mdk-usb-dongle/, свободный. Яз. англ. (дата обращения: 31.05.2020).

[3] Openthread protocol description [Электронный ресурс]. Режим доступа: https://openthread.io, свободный. Яз. англ. (дата обращения: 31.05.2020).

[4] Линейная и круговая поляризация электромагнитных волн в системах RFID [Электронный ресурс]. Режим доступа: http://asupro.com/gps-gsm/means-identification/automatic/linear-circular-polarization-rfid-system.html, свободный. Яз. рус. (дата обращения: 31.05.2020).

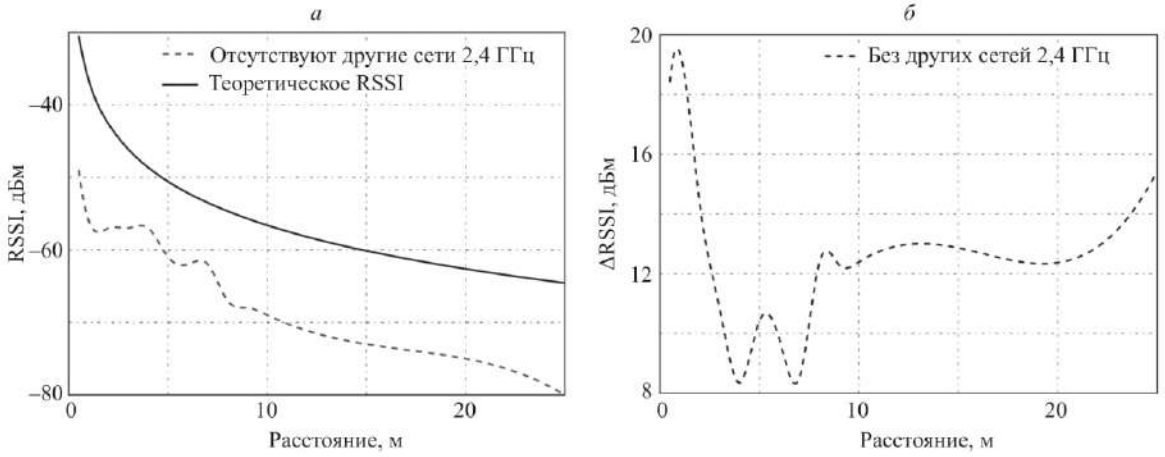

Рис. 4. Влияние открытого пространства и аппаратных особенностей приемника и передатчика на значение RSSI (*а*) и отклонение ΔRSSI от теоретического (*б*)

жено эффекту многолучевого распространения [11]. В результате осуществляется прием не только прямых лучей, но и отраженных, что приводит к флуктуации амплитуды, фазы и угла приема сигнала. На рис. 5 показаны графики зависимости RSSI и ΔRSSI от расстояния соответственно. Очевидно, что на графиках присутствуют значительные флуктуации, порождаемые эффектом многолучевого распространения.

Препятствия, возникающие на пути распространения сигнала, оказывают аналогичное влияние. Кроме того, могут также проявляться явления дифракции (на поверхностях с резкими неровностями) и рассеяния (на шероховатых поверхностях). Стены между цехами и производственными ячейками, с одной стороны, могут отражать сигнал благодаря арматуре, а с другой, поглощать электромагнитные излучения из-за наличия шумоизоляции. Более того, на производстве встречаются ситуации, когда приемник или передатчик располагается только за определенным препятствием или внутри короба. На рис. 6 приведены данные для подобного случая. Полученный результат подобен представленному на рис. 5, однако отклонение RSSI больше, поскольку преграда имеет распределенный вид.

Так как производственная сетевая инфраструктура характеризуется гетерогенностью сетей, появляется вероятность одновременной работы нескольких беспроводных сетей в одном частотном диапазоне. В таком случае может наблюдаться явление интерференции. Происходит наложение когерентных волн друг на друга, что приводит к увеличению или снижению результирующей амплитуды. Такой эффект может особенно сильно проявляться, когда устройства работают на одном и том же канале, что характерно для диапазона 2,4 ГГц, где пересекаются каналы Wi-Fi, Bluetooth, Zigbee и Thread сетей [12]. В соответствии с графиками (рис. 7) становится очевидным значительное влияние этого фактора.

Множество силового оборудования, которое находится в цеховых помещениях, может искажать проходящий беспроводной сигнал и создавать помехи. Чаще всего это наблюдается, когда по силовым линиям оборудования распространяется импульсный сигнал. Это происходит по той причине, что при формировании фронта амплитуда сигнала изменяется с высокой скоростью, а это приводит к появлению в спектре сигнала большого количества высокочастотных гармоник.

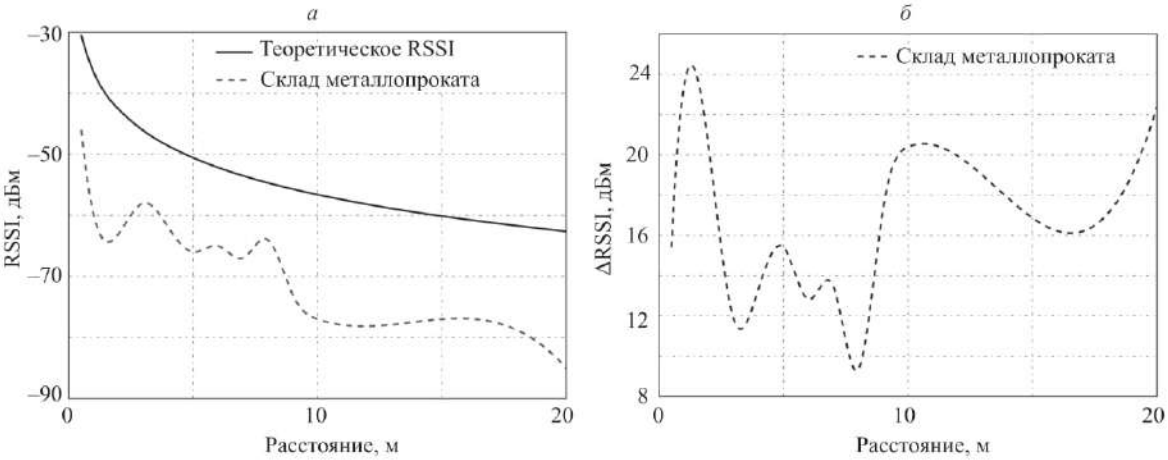

Рис. 5. Влияние отражающего характера пространства среды на значение RSSI (*а*) и отклонение ΔRSSI от теоретического (*б*)

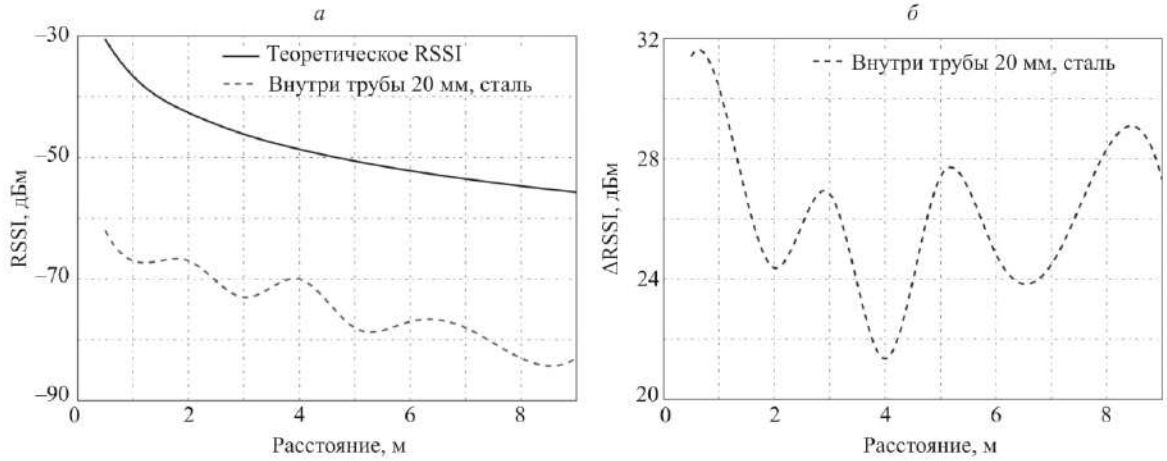

Рис. 6. Влияние препятствий в производственном помещении на значение RSSI (*а*) и отклонение ΔRSSI от теоретического (*б*)

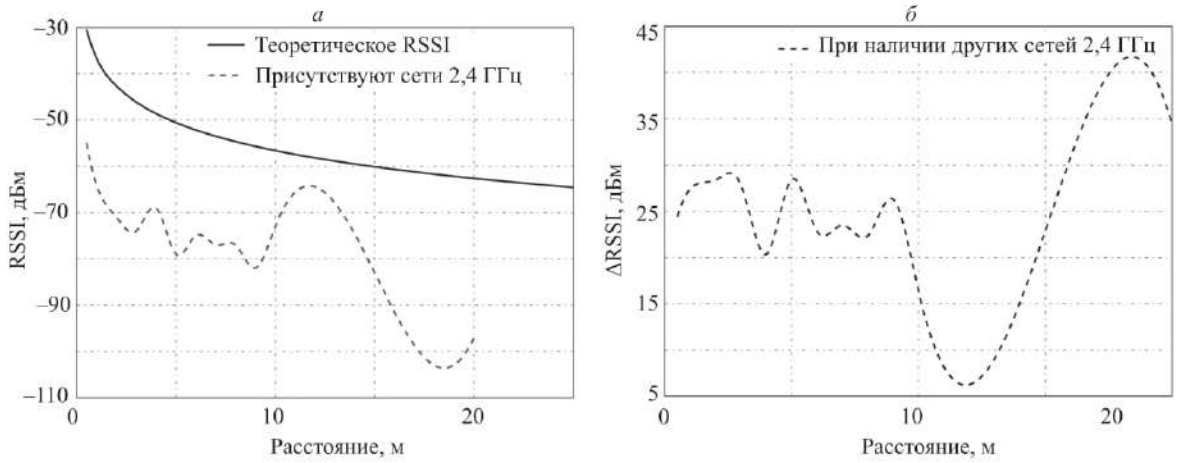

Рис. 7. Влияние соседних сетей 2,4 ГГц на значение RSSI (*а*) и отклонение ΔRSSI от теоретического (*б*)

Применение шагового двигателя, рассматриваемого в эксперименте, создает помехи, связанные с управлением двигателя посредством широтно-импульсной модуляции. Это приводит к искажению беспроводного сигнала (рис. 8). В случае асинхронного двигателя (рис. 9) сильных флуктуаций не наблюдается, поскольку управление таким двигателем осуществляется без применения частотных преобразователей.

В случае применения аппарата TIG-сварки наблюдались наибольшие искажения сигнала (рис. 10), поскольку процесс сварки связан с переключением мощных транзисторных ключей инвертора и появлением затухающих высокочастотных колебаний[1].

В табл. 2 приведены сравнительные результаты экспериментов. Видно, что для производственных помещений, где располагается оборудование для металлообработки, необходимо обеспечить расположение узлов беспроводной сети на расстоянии не более, чем 7 м друг от друга. Отклонение экспериментальных значений от теоретических скорее связаны с некоторыми особенностями антенны, неучтенными при расчете, а также усреднением значения экспоненты затухания. Стоит отметить, что организация ячеистой топологии беспроводных персональных сетей позволяет нивелировать отрицательные эффекты за счет плотного расположения узлов сети согласно рассчитанному D_{max}.

В соответствии с ОНТП 14-93[2] оптимальная норма расстояния между технологическим оборудованием в механообрабатывающих цехах единичного и мелкосерийного производства составляет величину не более 2,6 м. Наиболее рациональным местом расположения узла беспроводной сети является тыльная сторона оборудования, которая согласно вышеуказанным нормам должна располагаться на расстоянии 0,5 м от цехового

[1] Переходные процессы в сварочных инверторах [Электронный ресурс]. Режим доступа: https://303421.selcdn.ru/soel-upload/clouds/1/iblock/e03/e036d5c6f8c271c7ebbf44799619f068/201002052.pdf, свободный. Яз. рус. (дата обращения: 31.05.2020).

[2] Нормы технологического проектирования предприятий машиностроения, приборостроения и металлообработки (доработка ОНТП-14-93). Механообрабатывающие сборочные цехи: утверждены Роскоммаш взамен ОНТП-14-86/Минстанкопром СССР.

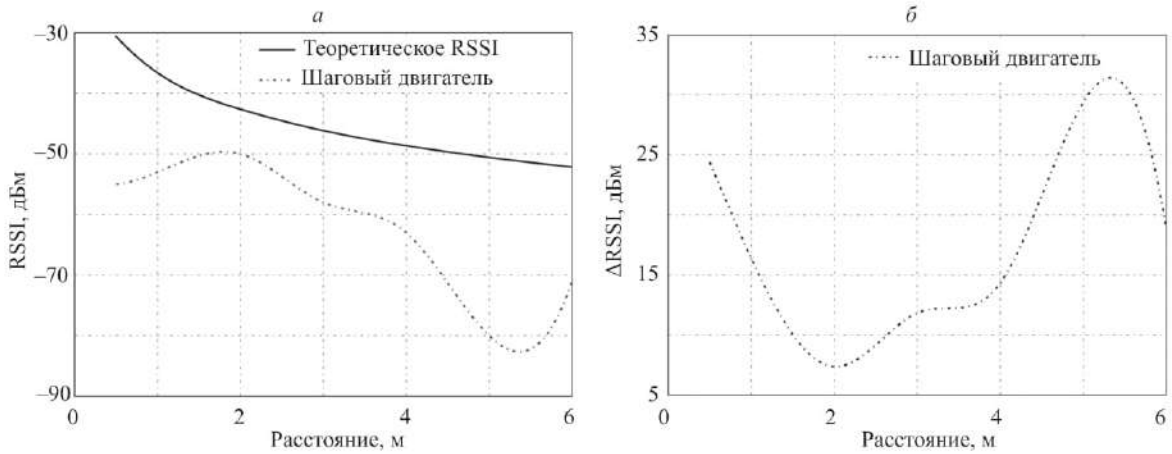

Рис. 8. Влияние работы шагового двигателя на значение RSSI (*а*) и отклонение ΔRSSI от теоретического (*б*)

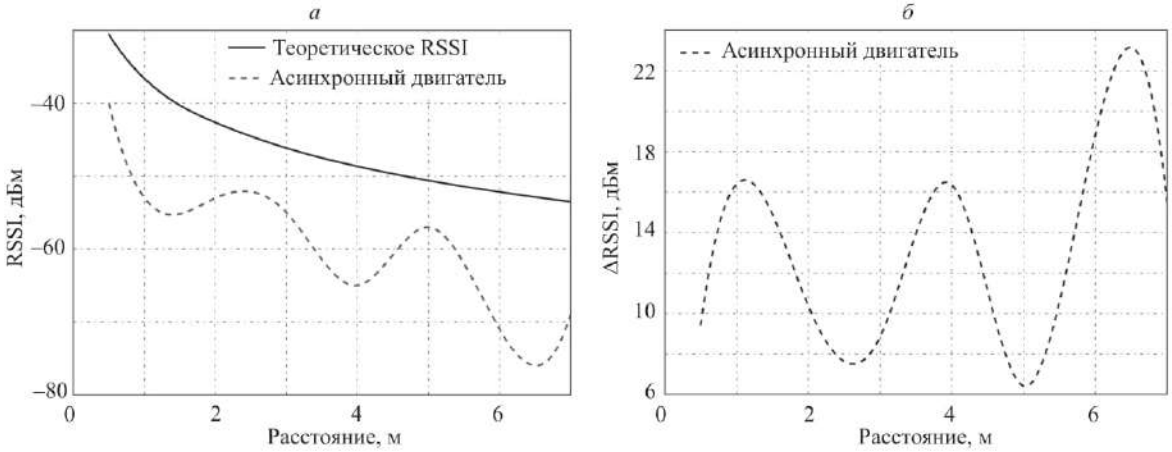

Рис. 9. Влияние работы асинхронного двигателя на значение RSSI (*а*) и отклонение ΔRSSI от теоретического (*б*)

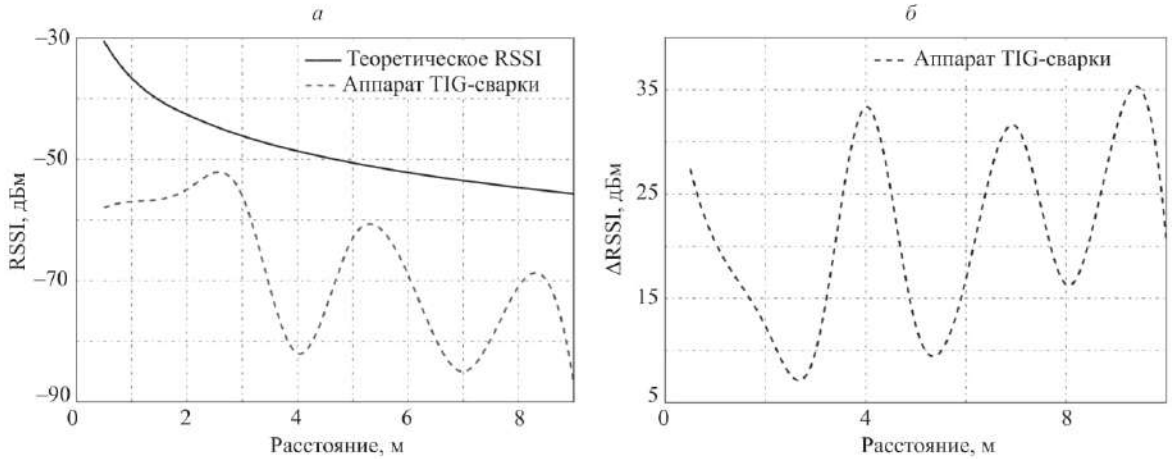

Рис. 10. Влияние работы аппарата TIG-сварки на значение RSSI (*а*) и отклонение ΔRSSI от теоретического (*б*)

проезда. При этом для любых возможных конфигураций расположения оборудования в цеху максимальная ширина проезда составляет 4 м. Можно сделать вывод о том, что при расположении узлов, как в проезде, так и непосредственно на оборудовании, будет возможно получить устойчивое соединение между узлами беспроводной сети при использовании ячеистой топологии. Последнее обусловлено тем, что при потере связи между узлами, расположенными в проезде, останется возможность прохождения сигнала непо-

Таблица 2. Сравнение факторов производственной сети на помехозащищенность беспроводных персональных сетей (WPAN)

Факторы	Средняя ошибка ΔRSSI, дБм	Максимальная ошибка $\Delta RSSI_{max}$, дБм	Максимальное расстояние между передатчиками D_{max}, м
Свободная среда без других сетей 2,4 ГГц	**12,37**	**18,82**	**34,22**
Асинхронный двигатель	12,69	19,37	33,99
Отражающая среда (склад металлопроката)	16,33	23,39	22,54
Шаговый двигатель	17,51	29,41	19,67
Аппарат TIG-сварки	21,10	34,35	13,01
Зашумленная среда другими сетями 2,4 ГГц	25,03	32,37	8,27
Экранирование толстостенной стальной трубой	**26,67**	**34,41**	**6,85**

В таблице жирным шрифтом выделены наилучшие и наихудшие значения показателей.

средственно между узлами, установленными на самом оборудовании.

Заключение

Выполнена серия экспериментов по определению помехозащищенности беспроводных персональных сетей в производственной среде. Предложен способ оценки помехозащищенности таких сетей на основе показателя уровня принимаемого сигнала.

Показана принципиальная возможность развертывания беспроводных персональных сетей в условиях реального производства при наличии силового оборудования. Выявлены факторы ослабления беспроводного сигнала, и вычислены значения помехозащищенности на реальном производстве.

Учитывая полученные данные, можно сделать вывод о том, что многие производственные факторы оказывают ощутимое влияние на качество сигнала беспроводных персональных сетей, воздействие которых не настолько значительное, чтобы отказываться от их применения. Более того, влияние многих факторов может быть нивелировано применением ячеистой топологии и плотным расположением приемников и передатчиков в соответствии со стандартом IEEE 802.15.4. Стоит отметить важность проведения соответствующих измерений для каждого конкретного производства, поскольку это обеспечит эффективное расположение приемников и передатчиков в цеховых помещениях.

Также полученные результаты могут быть полезными в процессе цифровой трансформации производства в соответствии с концепцией Индустрия 4.0, подразумевающей большую гибкость и мобильность производственного процесса. Очевидно, что при реализации данной концепции, беспроводной метод передачи данных с полевых датчиков становится предпочтительнее проводного. Разработанная методика оценки помехозащищенности беспроводных сетей может быть применена на практике при создании схем размещения оборудования в производственных цехах с учетом данных требований. При этом проведение анализа помехозащищенности беспроводной сети повышает ее надежность и отказоустойчивость в условиях цифрового производства.

Литература

1. Aheleroff S., Philip R., Zhong R.Y., Xu X. The degree of mass personalisation under Industry 4.0 // Procedia CIRP. 2019. V. 81. P. 1394–1399. doi: 10.1016/j.procir.2019.04.050
2. Mehta B.R., Reddy Y.J. Industrial Process Automation Systems: Design and Implementation. Chapter 14. Wireless communication. Butterworth-Heinemann, 2015. P. 417–457.
3. Stenumgaard P., Chilo J., Ferrer-Coll J., Angskog P. Challenges and conditions for wireless machine-to-machine communications in industrial environments // IEEE Communications Magazine. 2013. V. 51. N 6. P. 187–192. doi: 10.1109/MCOM.2013.6525614
4. Ängskog P., Karlsson C., Coll J.F., Chilo J., Stenumgaard P. Sources of disturbances on wireless communication in industrial and factory environments // Proc. of the Asia-Pacific Symposium on Electromagnetic Compatibility (APEMC 2010). Beijing, China. 2010. P. 281–284. doi: 10.1109/APEMC.2010.5475862
5. Li H., Liu L., Li Y., Yuan Z., Zhang K. Measurement and characterization of electromagnetic noise in edge computing networks for the industrial Internet of Things // Sensors (Switzerland). 2019. V. 19. N 14. P. 3104. doi: 10.3390/s19143104
6. Saaifan K.A., Henkel W. Measurements and modeling of impulse noise at the 2.4 GHz wireless LAN band // Proc. 5th IEEE Global Conference

References

1. Aheleroff S., Philip R., Zhong R.Y., Xu X. The degree of mass personalisation under Industry 4.0. *Procedia CIRP*, 2019, vol. 81, pp. 1394–1399. doi: 10.1016/j.procir.2019.04.050
2. Mehta B.R., Reddy Y.J. *Industrial Process Automation Systems: Design and Implementation. Chapter 14. Wireless communication.* Butterworth-Heinemann, 2015, pp. 417–457.
3. Stenumgaard P., Chilo J., Ferrer-Coll J., Angskog P. Challenges and conditions for wireless machine-to-machine communications in industrial environments. *IEEE Communications Magazine*, 2013, vol. 51, no. 6, pp. 187–192. doi: 10.1109/MCOM.2013.6525614
4. Ängskog P., Karlsson C., Coll J.F., Chilo J., Stenumgaard P. Sources of disturbances on wireless communication in industrial and factory environments. *Proc. of the Asia-Pacific Symposium on Electromagnetic Compatibility (APEMC 2010)*, Beijing, China, 2010, pp. 281–284. doi: 10.1109/APEMC.2010.5475862
5. Li H., Liu L., Li Y., Yuan Z., Zhang K. Measurement and characterization of electromagnetic noise in edge computing networks for the industrial Internet of Things. *Sensors (Switzerland)*, 2019, vol. 19, no. 14, pp. 3104. doi: 10.3390/s19143104
6. Saaifan K.A., Henkel W. Measurements and modeling of impulse noise at the 2.4 GHz wireless LAN band. *Proc. 5th IEEE Global Conference*

on Signal and Information Processing (GlobalSIP 2017). Montreal, Canada. 2017. P. 86–90. doi: 10.1109/GlobalSIP.2017.8308609
7. Takaya K., Maeda Y., Kuwabara N. Experimental and theoretical evaluation of interference characteristics between 2.4-GHz ISM-band wireless LANs // Proc. of the 1998 IEEE EMC Symposium. International Symposium on Electromagnetic Compatibility. Symposium Record. Denver, CO, USA. 1998. V. 1. P. 80–85. doi: 10.1109/ISEMC.1998.750064
8. Guo W., Healy W.M., Zhou M. Impacts of 2.4-GHz ISM band interference on IEEE 802.15.4 wireless sensor network reliability in buildings // IEEE Transactions on Instrumentation and Measurement. 2012. V. 61. N 9. P. 2533–2544. doi: 10.1109/TIM.2012.2188349
9. Zheng G., Han D., Zheng R., Schmitz C., Yuan X. A link quality inference model for IEEE 802.15.4 Low-Rate WPANs // Proc. 54th IEEE Global Telecommunications Conference (GLOBECOM 2011). Houston, TX, USA. 2011. P. 6133782. doi: 10.1109/GLOCOM.2011.6133782
10. Konings D., Faulkner N., Alam F., Noble F., Lai E.M. The effects of interference on the RSSI values of a ZigBee based indoor localization system // Proc. 24th International Conference on Mechatronics and Machine Vision in Practice (M2VIP 2017). Auckland, New Zealand. 2017. P. 1–5. doi: 10.1109/M2VIP.2017.8211460
11. Cheffena M. Propagation channel characteristics of industrial wireless sensor networks [wireless corner] // IEEE Antennas and Propagation Magazine. 2016. V. 58. N 1. P. 66–73. doi: 10.1109/MAP.2015.2501227
12. Musăloiu-E.R., Terzis A. Minimising the effect of WiFi interference in 802.15.4 wireless sensor networks // International Journal of Sensor Networks. 2008. V. 3. N 1. P. 43–54. doi: 10.1504/IJSNET.2008.016461

on Signal and Information Processing (GlobalSIP 2017), Montreal, Canada, 2017, pp. 86–90. doi: 10.1109/GlobalSIP.2017.8308609
7. Takaya K., Maeda Y., Kuwabara N. Experimental and theoretical evaluation of interference characteristics between 2.4-GHz ISM-band wireless LANs. *Proc. of the 1998 IEEE EMC Symposium. International Symposium on Electromagnetic Compatibility. Symposium Record*, Denver, CO, USA, 1998, vol. 1, pp. 80–85. doi: 10.1109/ISEMC.1998.750064
8. Guo W., Healy W.M., Zhou M. Impacts of 2.4-GHz ISM band interference on IEEE 802.15.4 wireless sensor network reliability in buildings. *IEEE Transactions on Instrumentation and Measurement*, 2012, vol. 61, no. 9, pp. 2533–2544. doi: 10.1109/TIM.2012.2188349
9. Zheng G., Han D., Zheng R., Schmitz C., Yuan X. A link quality inference model for IEEE 802.15.4 Low-Rate WPANs. *Proc. 54th IEEE Global Telecommunications Conference (GLOBECOM 2011)*, Houston, TX, USA, 2011, pp. 6133782. doi: 10.1109/GLOCOM.2011.6133782
10. Konings D., Faulkner N., Alam F., Noble F., Lai E.M. The effects of interference on the RSSI values of a ZigBee based indoor localization system. *Proc. 24th International Conference on Mechatronics and Machine Vision in Practice (M2VIP 2017)*, Auckland, New Zealand, 2017, pp. 1–5. doi: 10.1109/M2VIP.2017.8211460
11. Cheffena M. Propagation channel characteristics of industrial wireless sensor networks [wireless corner]. *IEEE Antennas and Propagation Magazine*, 2016, vol. 58, no. 1, pp. 66–73. doi: 10.1109/MAP.2015.2501227
12. Musăloiu-E.R., Terzis A. Minimising the effect of WiFi interference in 802.15.4 wireless sensor networks. *International Journal of Sensor Networks*, 2008, vol. 3, no. 1, pp. 43–54. doi: 10.1504/IJSNET.2008.016461

Авторы

Афанасьев Максим Яковлевич — кандидат технических наук, доцент, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, Scopus ID: 57194081345, ORCID ID: 0000-0003-4061-1407, amax@niuitmo.ru

Федосов Юрий Валерьевич — кандидат технических наук, доцент, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация; инженер, ОАО «Российский институт мощного радиостроения», Санкт-Петербург, 199048, Российская Федерация, ORCID ID: 0000-0003-1869-0081, yf01@yandex.ru

Крылова Анастасия Андреевна — аспирант, инженер, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация; программист, ООО «ЛАР Технологии», Санкт-Петербург, 197342, Российская Федерация, ORCID ID: 0000-0002-5822-6702, ananasn94@gmail.com

Шорохов Сергей Александрович — аспирант, инженер, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, ORCID ID: 0000-0002-5412-7723, stratumxspb@gmail.com

Зименко Ксения Владимировна — инженер, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, ORCID ID: 0000-0002-3792-136X, zksenia@yahoo.com

Authors

Maxim Ya. Afanasiev — PhD, Associate Professor, ITMO University, Saint Petersburg, 197101, Russian Federation, Scopus ID: 57194081345, ORCID ID: 0000-0003-4061-1407, amax@niuitmo.ru

Yuri V. Fedosov — PhD, Associate Professor, ITMO University, Saint Petersburg, 197101, Russian Federation; Engineer, Russian Institute for Power Radioengineering (JSC RIPR), Saint Petersburg, 199048, Russian Federation, ORCID ID: 0000-0003-1869-0081, yf01@yandex.ru

Anastasiia A. Krylova — Postgraduate, Engineer, ITMO University, Saint Petersburg, 197101, Russian Federation; Software Developer, LLC Lar Technologies, Saint Petersburg, 197342, Russian Federation, ORCID ID: 0000-0002-5822-6702, ananasn94@gmail.com

Sergei A. Shorokhov — Postgraduate, Engineer, ITMO University, Saint Petersburg, 197101, Russian Federation, ORCID ID: 0000-0002-5412-7723, stratumxspb@gmail.com

Ksenia V. Zimenko — Engineer, ITMO University, Saint Petersburg, 197101, Russian Federation, ORCID ID: 0000-0002-3792-136X, zksenia@yahoo.com

Machine Vision for Auto Positioning in the Modular Industrial Equipment: A Case Study

Maxim Ya. Afanasev, Yuri V. Fedosov, Yuri S. Andreev, Anastasiya A. Krylova, Sergey A. Shorokhov, Kseniia V. Zimenko, Mikhail V. Kolesnikov
Faculty of Control Systems and Robotics, ITMO University
St. Petersburg, Russia
amax@niuitmo.ru

***Abstract*—The unification principle of industrial equipment involves splitting the installation into interchangeable components, from which the required type of equipment is assembled for specific production tasks according to the order. For the common industrial equipment, zero position setting usually causes no issues. However, for the modular machines with different sizes of modules, the mounting error can be appeared and it has to be corrected automatically. In addition, due to a possible frequent change of equipment during the production process, clamps can wear out, increasing the error. It is necessary to introduce an automatic adjustment of tool position relative to the carriage suspension, implemented using machine vision systems. The cost of video capture equipment and computing tools decreases, and their characteristics, on the contrary, grow. Moreover, algorithms for recognizing markers' position and angle together with augmented reality development tools are also constantly being improved with their application complexity being significantly reduced, which emphasizes the practicability of their use. The paper describes the problem of eliminating the installation error of instrumental modules on a three-axes platform. The proposed method involves the use of a number of reference markers on the platform, its suspension, and on the plug-in modules.**

***Index Terms*—Machine vision, Modular industrial equipment, Marker-based method, CNC, ArUco markers.**

I. Introduction

The field of industrial production is an extremely dynamic and rapidly developing area of human activity. Current trends require an increasingly frequent change of a product range, as well as a gradual transition to personalized production, when even consumer goods are customized for a specific user.

All of the above leads to Research & Design works coming to the fore. There are more and more private design bureaus and organizations that are involved in producing high-tech devices and technologies to order. Such small innovative enterprises are usually called Startups. The goal of any startup is to design and test new products and bring them to the market as quickly as possible. Evidently, this requires an organization of a continuous cycle of prototypes development, and therefore, the availability of its own production enterprise.

The development of such experimental production is highly difficult for startups. Modern products are becoming more complex, combining both mechanical, electrical, and electronic components. The latter is especially relevant in connection with the rapid development of the Internet of Things and the Industrial Internet of Things, as well as with the widespread introduction of "Smart Technologies". The prototypes design of such devices requires a large number of sophisticated equipment units, so most startups in early stages of their development are forced to use the services of large organizations that have their own production facilities. This need leads to the following negative consequences:

- The cost of a prototype increases, since large enterprises are primarily aimed at the production of large batches and are ready to take a single order only if its cost is comparable to the cost of the batch. A typical example of the described situation is the production field of printed circuit boards. Here most of the order cost falls on the preproduction engineering, which is practically independent of the serial production.
- The prototype production time increases, since contacting a third-party manufacturer requires preparation of a full set of technical documentation, and involves coordination process, cost negotiations, etc.

These problems can be solved in different ways, for example, in the printed circuit boards production, the most common solution is the combination of orders from different manufacturers (which definitely has a positive effect on the cost of a prototype, but not on the production time). So-called core facilities and industrial co-working centers are also being developed. However, the concept of developing a modular equipment of the "office" class, which is characterized by lower accuracy and performance parameters, seems more promising.

Unquestionably, the main objective of a prototype is to clearly demonstrate the operability of a developed product. Prototypes are made in a single copy or in small batches, so the technology for their production may not be optimal. For example, the prototype of a case can be made on a 3D printer, and instead of a complex multilayer printed circuit board, several single-layer ones can be used. At the same time, modular equipment occupies much less production space and can easily be moved if the organization itself moves to a different place.

The paper describes the current development stage of a multipurpose modular platform for adjustable manufacturing equipment. The developed platform is a combination of chassis positioning the operating carriage in three-dimensional

978-1-7281-6389-5/20/$31.00 ©2020 IEEE

Authorized licensed use limited to: St Petersburg Natl Uni of Info Tech Mech & Optics. Downloaded on January 09,2021 at 17:20:13 UTC from IEEE Xplore. Restrictions apply.

Cartesian space. On the carriage suspension various modules, that ensure the operation of the equipment, are mounted. By replacing the modules, various types of industrial equipment are created such as: equipment for the selective curing of photopolymers, drilling machines, milling machines, engravers, laser cutters, 3D printers, coordinate-measuring machines, machines for installing components on a printed circuit board, marking machine, sorting equipment, dispensers of chemical reagents, cartesian robots.

The developed platform is based on the principles of unification and hybridization. Unification refers to an open software and hardware architecture that allows creating new types of equipment and software on the basis of the "intelligent constructor set". Unification is achieved by dividing a single product into bigger interchangeable modules with each having a clear description of input and output parameters. To put it another way, a line of chassis of different overall dimensions and rigidity is being created for various types of equipment. At the same time, the principles of building electronic, electrical, mechanical, and software parts of each line representative remain invariable, and the transition to a new type of equipment is achieved by changing modules.

Undoubtedly, modules may vary in size depending on the problem being solved. Moreover, in accordance with the principle of unification, the control system must determine the position of the connected module independently and without the need for manual adjustment or calibration. It is known that processing accuracy depends on the accuracy of mutual geometric position of all components in industrial equipment. The latter is extremely difficult to achieve using the modular approach, since the position of the module with respect to the chassis coordinate system is not known in advance.

Higher positioning accuracy can be obtained by changing the way modules are mounted on a carriage suspension and, for example, using dovetail mounts with a wedge clamp. The described scheme is widely used in quick change toolposts of metalworking machines. This approach provides excellent accuracy and repeatability when reinstalling modules manually. However, it greatly complicates the overall suspension design, due to the need for high precision machining of mating parts (grinding is required for all mating surfaces in a joint), and the overall weight of the moving part in a system also increases. Furthermore, all modules must have the same dovetail mating part, depending not on the module size itself, but on the chassis dimensions. Consequently, the principle of universality is violated because in this case it is impossible to install a small-sized module on a large chassis.

It is also worth noting that one of the options for using modular equipment is to include it in flexible production lines, where automatic module replacement with the use of an industrial robotic arm will be required. The analysis of modern equipment has shown that sufficient pressing force for a wedge clamp can only be achieved using pneumatic clamps. This leads to the need of an additional air line, which also complicates the initial proposed approach. Instead, it is suggested to use a magnetic mounting system with an additional guide groove positioning on a plane parallel to the direction of carriage movement.

The described fastening system allows using modules of any size; however, the problem of combining the module and main chassis coordinate systems arises. In this paper, a combined automatic positioning system for a module on a carriage suspension using machine vision is suggested. This system determines the location of specialized optical markers on the chassis and automatically changes geometric parameters used in the control system.

The article is organized as follows. Section II is dedicated to an overview of related work. In Section III the description of the operational setup is provided. Section IV describes the methodology of research. Section V focuses on the results at this stage of development. The main limitations and drawbacks of the proposed approach are discussed in Section VI. Finally, Sections VII contains conclusions.

II. Related Work

The application of machine vision is quite ubiquitous for today, especially within the industrial context. It is an essential part of industrial robotics that allows manufacturers to accomplish various tasks such as pick and place operations over different pieces [1], quality inspection [2], [3], human safety [4], etc. Besides, machine vision is getting relevant within CNC machine tools that are an integral part of manufacturing. A sufficient number of surveys are carried out regarding marker-based methods.

Researchers from Aarhus University [5] combine augmented reality and markers to create a CNC system for the laser cutter. The system is based on the WYSIWYG approach, where a projector is used to depict current contours, and markers are used to set its position on the working area. Along with that, Keio University specialists [6] expand the fiducial markers functionality for a laser cutter. To set cutting parameters, they put a set of fiducial markers near the workpiece, including material, operation order, and command markers.

In [7] a contouring error detection method based on machine vision is described. A special measurement fixture covered with markers is designed that allows researchers to measure contouring error without a cross-grid encoder.

Moreover, verification of the machining process is one of the tasks that can be addressed using machine vision. It includes collision avoidance of moving parts of the machine, tool cutting into the workpiece or fixture, and tool breakage. Commonly, a special software simulation that needs a detailed kinematics model is usually used for these purposes. However, such a method is not suitable for modular equipment.

Work [8] provide a method for a direct on machine tool processing simulation. Special markers are glued to the machine to determine tool and workpiece position, and augmented reality is applied to simulate toolpath and workpiece processing. Nevertheless, modular equipment requires the determination of the geometric parameters of the modules and working area on the fly. Thus, an additional study is needed.

Authorized licensed use limited to: St Petersburg Natl Uni of Info Tech Mech & Optics. Downloaded on January 09,2021 at 17:20:13 UTC from IEEE Xplore. Restrictions apply.

III. Operational setup

The proposed system for the automatic module positioning is a further development of the architecture of the modular production system described earlier in [9], [10], [11]. In accordance with this concept, a machine vision module is mounted on a chassis along with other modules used for processing (Figure 1). Nevertheless, the presence of this module is not mandatory, because it is needed only for those operations that require machine vision capabilities. In particular, it helped to solve the problem of determining the position of a workpiece by tracking reference points located on it [12].

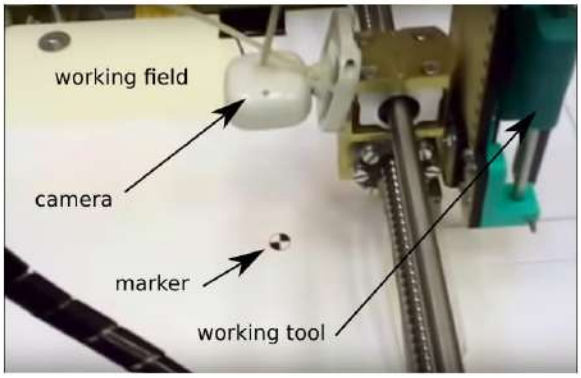

Fig. 1. Marker detection system prototype.

However, this system has a number of significant drawbacks. In particular, the method of quick module reinstallation on a carriage suspension is not implemented. The modules are mounted on screw connections, and after installing each module, it is required to manually input the optical axis offset value of the machine vision module relative to the instrument axis. In other words, the task of determining the exact position of a module on the suspension of a 3-axis platform is solved only mechanically and requires additional actions from an operator.

Currently, a new construction has been developed to simplify and automate the reinstallation of modules using magnetic mounting on a plane with an additional groove. Standard machine mounts with switchable (on/off) permanent magnets that have a separation force from 30 to 160 kg are used for experiments (Figure 2).

With the described configuration, the initial scheme with the camera being on a moving suspension does not allow precise module positioning with respect to the base coordinate system of a 3-axis platform. Modules can have different sizes, so the processing area can also change. In this case, the exact location of an actuator (a spindle axis, a laser head lens, a probe of the measuring module, a nozzle of a three-dimensional printing module, etc.) remains unknown at the stage of module connection. Moreover, the location of the module with a camera on a carriage suspension makes it impossible to see the entire work area fully. The carriage must be moved to find markers and determine their exact position, which takes extra time during module initialization.

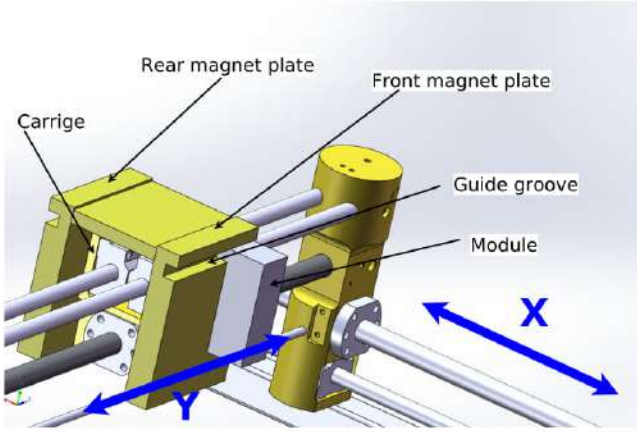

Fig. 2. Carriage design.

Therefore, a second camera is included in the new construction, located above the work area. This camera has a larger viewing angle, which allows determining the modules position with respect to the carriage and corners of the work area. Consequently, an additional upper camera is used to automatically position the modules with respect to the coordinate system of the 3-axis platform. On the contrary, the camera, located directly on the carriage suspension, is responsible for automatic instrumental equipment (machine vise, angle plate, jig, etc) positioning relative to the coordinate system of the 3-axis platform and part positioning relative to the machine tool (Figure 3).

Markers are made of polished stainless steel plates and mechanically attached to parts of a three-axis platform and instrumental equipment. Currently, images are applied by using ultraviolet printing. Experiments on the use of electrochemical blackening of stainless steel are also being conducted. Mechanical attachment of markers is preferred over gluing. The possibility of applying markers using laser engraving directly during the manufacturing process is also considered. This will increase the positioning accuracy and eliminate the need for manual tagging.

IV. Methodology

A. The selection of software development tools

The task of positioning by using optical markers relates to the field of augmented reality. There already exist a sufficient number of fully developed software libraries that implement the solution to this problem. The most well known one is OpenCV [13], [14], an open source library for computer vision. It has a convenient set of tools that simplify such routine procedures of machine vision systems as receiving a video stream from various external sources, camera calibration, image filtering, etc. In particular, this library has already been used to determine the position of a workpiece on a table and receive its coordinates using the Canny edge detection filter [12].

Authorized licensed use limited to: St Petersburg Natl Uni of Info Tech Mech & Optics. Downloaded on January 09,2021 at 17:20:13 UTC from IEEE Xplore. Restrictions apply.

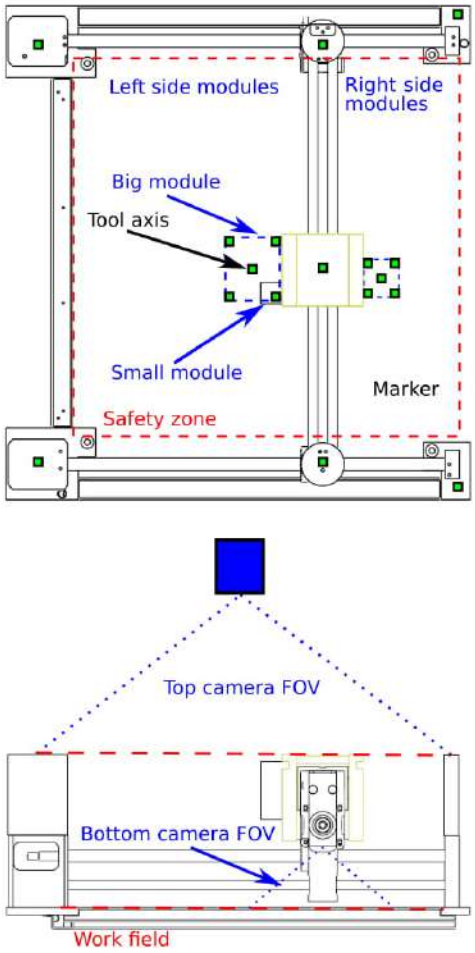

Fig. 3. Layout of cameras and modules.

To solve the current problem, the following requirements for the optical marker recognition software library were defined:

1) The library should be based on the OpenCV functions.
2) The library should be able to recognize markers by using a special dictionary.
3) The library must consider optical parameters of a camera (distortion, optical axis tilt, etc).

Markers should be simple enough for manufacturing (black-and-white markers with simple geometry, which can be made by UV printing, laser engraving or electrochemical blackening, are preferred).

Several software libraries that met the specified criteria were analyzed. Some of them are used to recognize markers that store data in binary form (Data Matrix, Maxicode, Quick Response Code). The described markers are convenient for creating an arbitrary dictionary when each marker describes itself. However, for the particular task under consideration, this possibility is redundant, because for the complete description of the 3-axis chassis geometry only 50 markers are required.

Therefore, application of dictionary encoded markers is appropriate. Firstly, in the case of its usage, an array containing the same type of markers is generated. It is important that the condition of maximum dissimilarity of markers is satisfied in order to simplify the recognition task.

All dictionary encoded markers can be divided into the following conditional categories [15]:

- markers consisting of square and rectangular components (ARTag, Cybercode, Matrix, ArUco, AprilTag, VisualCode);
- markers consisting of circle and arc elements (TRIP, RUNE-Tag, CCTag, Intersense, BinARyID);
- markers with mixed elements, including complex shapes (ARTollkit, STag, ReacTIVision, SCR, SVMS).

Each of the listed marker types has its own advantages. In particular, circular markers show better performance when determining points (for example, in three-dimensional scanning systems), the rectangular type is optimal for determining planes in three-dimensional space, and mixed markers are best for high accuracy recognition in difficult conditions (poor lighting, overlapping markers, etc.). Unquestionably, rectangular markers are the most suitable for the considered system of automatic positioning of modules. Therefore, after the literature research and a series of experiments, the ArUco [16] was chosen as the main software library for the recognition process.

The ArUco library allows implementing a system for pose estimation, based on binary square fiducial markers. The main advantage of these markers is that the presence of only one marker in the camera's field of view allows an accurate determination of the camera's position relative to the plane where marker is located. Moreover, the guaranteed asymmetry of each marker makes it possible to determine not only the coordinates of the marker's center but also its angle of rotation. To sum up, with ArUco markers it is possible to obtain the exact location and direction of coordinate axes in a Cartesian space, which is extremely important for the problem under consideration.

B. Camera Positioning and Calibration

To implement an automatic positioning system, it is necessary to solve two main tasks:

- *Proper camera placement*. The camera must capture the entire work area and at the same time not interfere with any working equipment.
- *Camera calibration*. The process of camera calibration lies in obtaining its intrinsic parameters and distortion coefficients. These parameters remain constant until the camera optics are changed, so the camera must be calibrated only once.

To obtain the most accurate result during the optical marker recognition, it is advisable to place the camera directly above the work area. In this case, the optical axis should coincide with the normal vector to the surface of the latter. Such configuration minimizes the marker overlap by components

Authorized licensed use limited to: St Petersburg Natl Uni of Info Tech Mech & Optics. Downloaded on January 09,2021 at 17:20:13 UTC from IEEE Xplore. Restrictions apply.

of the 3-axis chassis and simplifies the camera calibration procedure. Based on the geometric dimensions of the 3-axis chassis, it is necessary to select the parameters of camera lens, allowing seeing the entire work area in one frame. The main parameter in this case is the height of camera placement. Knowing it, it is possible to determine the lens' Angle of View (AOV) by using (1):

$$\alpha_v = 2\arctan\frac{S}{2h}, \tag{1}$$

where α_v is AOV of the lens, S is the size of the work area, h is the distance between the camera and the work area.

Once the location and parameters of the optical system are determined, it is possible to proceed to camera calibration. The ArUco library includes all the necessary tools for carrying out this process. The camera calibration procedure consists of the following steps. Firstly, a marker size is selected in bits as well as the size of a dictionary. The minimum marker size is 4x4 bits; a dictionary can contain a minimum of 50 different markers. Then, in accordance with the selected marker type and dictionary, the corresponding calibration table is generated, which is a field of a specific size, where markers are placed. The number of markers is determined experimentally; their size can be arbitrary and differ from the final size, which will be used directly for recognition. The latter is caused by the fact that the calibration table is needed only for obtaining parameters of the optical system. The ArUco library implements two main types of calibration tables: a table consisting only of ArUco markers and a table representing a checkerboard pattern where, in its white cells, ArUco markers are placed (this calibration method is called ChArUco). After creating the calibration table, an array of its images from different angles is obtained. All images should vary as much as possible, and the more images there are, the more accurate the calibration will be. It should be noted that the advantage of the calibration method with the ChArUco table is that, while obtaining a set of images, occlusions and partial views, when the table may not completely fall into the frame, are also acceptable.

C. Accuracy Evaluation

Undoubtedly, the positioning accuracy depends on the camera resolution. To calculate the theoretical resolution of a camera, which directly affects the accuracy of determining the optical marker position, the following equation is used (2):

$$C_r = N\frac{S}{C_s}, \tag{2}$$

where C_r is a camera resolution in mm, S is the size of a work area in mm, C_s is a camera resilution in px, N is the Nyquist factor, which value must be 2 or more according to the Nyquist-Shannon Signal Sampling Theorem.

However, this statement is true only if machine vision is the only way to determine the size of objects in a frame, while in the system under consideration there is an additional method of measurement. Optical markers are located on a moving carriage of a 3-axis chassis, controlled by servo-step drives. Each drive is equipped with an incremental encoder capable of determining the relative displacement value with a given accuracy. Thus, knowing the physical accuracy of the camera and the resolution of the incremental encoder, the maximum interpolation coefficient can be obtained as (3):

$$k = \frac{C_r}{E_r}, \tag{3}$$

where C_r is a camera resolution in mm, E_r is an incremental encoder resolution in mm.

The proposed approach will allow determining the size of objects in a frame with subpixel accuracy. In this case, it will be possible to assume that 1 pixel of an enlarged image corresponds to the resolution of the encoder, and errors in finding the exact position will be determined by the accuracy of the interpolation algorithm.

V. Results

In the experimental installation, camera height is set to 400 mm, work field size is $500\,\text{mm} \times 500\,\text{mm}$. Therefore, based on the (1), the angle of view equals 65°. A Withrobot oCam-5CRO-UM camera (OmniVision OV5640 CMOS Image Sensor, max resolution 2592×1944) and a Withrobot 3018PL002 lens (focal length (3.00 ± 0.15) mm, relative aperture $1.8 \pm 5\%$, Angle of View $65° \times 111° \times 126°$ $(V \times H \times D)$ with $1/2.9"$ sensor size) were used during experiments.

During tests, a computer with the following configuration was used: 6-core AMD Ryzen 5 2600 3.3 GHz, 16 GB RAM DDR4 3000 MHz, Gigabyte GeForce GTX 1070 Ti, Windows 10 Pro x64. The Python 3.8.1 programming language and the OpenCV 4.1.2 library were used for the software implementation. According to developers' recommendations, the calibration process was performed using a ChArUco board (Figure 4) with 6×6 markers. The size of a calibration table is $200\,\text{mm} \times 300\,\text{mm}$, with 7 columns and 5 rows. The table was made using UV printing on a white foam-core. In the experimental installation, Leadshine CS-D1008 servo-step drives were used, which made it possible to obtain a resolution of 12.5 µm for all axes.

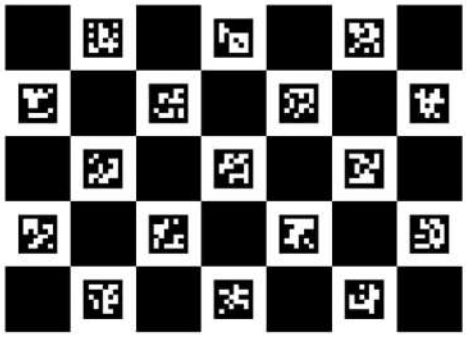

Fig. 4. ChArUco board.

Authorized licensed use limited to: St Petersburg Natl Uni of Info Tech Mech & Optics. Downloaded on January 09,2021 at 17:20:13 UTC from IEEE Xplore. Restrictions apply.

After performing the calibration using (2) and (3), the camera resolution ($C_r = 0.51\,\mathrm{mm}$) and the maximum interpolation coefficient ($k = 40.8$) were calculated. However, it should be noted that the system that determines the position of a module on a carriage suspension may have lower positioning accuracy than the system located on a suspension itself (the accuracy of which must be a priori equal to or greater than the resolution of servo encoders). This happens due to the fact that the procedure of determining the maximum size of the processing area and starting point coordinates of a tool was previously carried out manually with accuracy being within 0.1 mm. It is evident that such precision is sufficient enough to solve the given task. Based on the foregoing description, the interpolation coefficient was taken to be equal to 4, which gave the resolution after bicubic interpolation as 10368×7776, and $C_r = 0.1275\,\mathrm{mm}$. Using a test computer with libraries, mentioned above, the speed of processing one image frame from the camera (together with interpolation) takes app. 130 ms, which can be considered a sufficient result for the recognition task being solved statically. An example of recognized markers is shown in Figure 5.

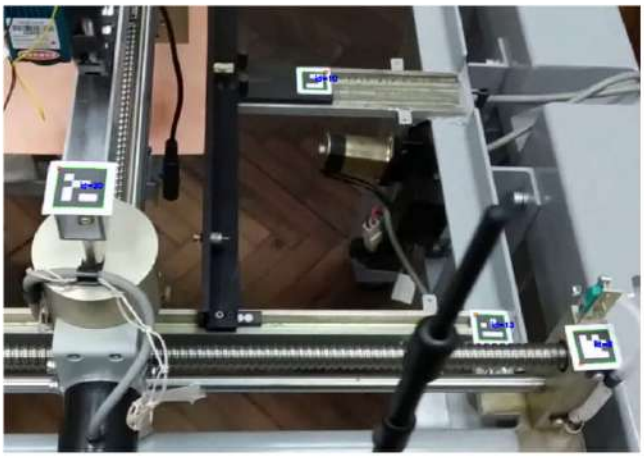

Fig. 5. An example of recognized markers.

To sum up, the following initialization algorithm for a new module was implemented based on the conducted experiments:

1) An operator installs a module on the carriage suspension, turns on its magnetic mount, turns on the power supply and the safety line [9], and then turns on the general power supply of the installation.
2) The control system accesses the machine vision module to obtain a preliminary configuration of the workspace.
3) The machine vision module takes an image from the camera in a normal (without interpolation) resolution, this image determines the dimensions of the work area, security zone, suspension coordinates, dimensions and location of the module (on the left or right mount), as well as the position of the point that will be taken as a preliminary machine zero. All the obtained data is transmitted to the control system. The accuracy at this stage is $\pm 1\,\mathrm{mm}$ (similar accuracy is inherent in mechanical limit switches).
4) The control system begins to move the carriage to the zero position, considering its location and positioning accuracy at this stage ($\pm 1\,\mathrm{mm}$).
5) Having reached the zero position, the control system resets the coordinates obtained from incremental encoders and starts moving the carriage to the maximum coordinate.
6) After reaching the maximum coordinate of the axes, the control system memorizes the values obtained from encoders. These values will be the maximum size of the work area with accuracy being equal to accuracy of the encoder (12.5 µm in this case).
7) At this point, the homing procedure is considered complete. The control system returns the carriage to the center of the work area and relinquishes control to the machine vision module.
8) The machine vision module takes a series of images, shifting the carriage along the axes and remembering the exact coordinate values received from encoders.
9) Based on the obtained series of images, the interpolated image with markers is obtained. It determines the exact coordinate of the tool axis with respect to machine coordinate system. Machine zero is shifted to this coordinate.
10) Geometric parameters of the chassis and the installed module are considered fully defined and the control system proceeds to the main program cycle.

VI. Discussion and Future Work

Methods of machine vision make it possible to accurately determine distances on a plane. However, for the full-fledged operation of modules in three-dimensional space, it is additionally required to determine the distance from a tool base point to a processing plane. At the moment, the problem is successfully solved by using the point triangulation method. This approach is most widely used when building 3D scanners, which accuracy can reach several micrometers. However, the described approach requires at least two cameras with a predetermined location. Contrast lighting is also often required. Therefore, at the current stage of development, a decision was made to additionally include a probe in a module, which needs precise height positioning, to accurately set it relative to the processing plane. Currently, a contact sensor mounted directly in the work area is used for experiments. To set the Z-axis value to zero, the carriage drives up to the point where the sensor is located and goes down to touch a probe or tool.

The optimal size of ArUco markers has been experimentally determined. For the considered chassis with a working area of $500\,\mathrm{mm} \times 500\,\mathrm{mm}$, $25\,\mathrm{mm} \times 25\,\mathrm{mm}$ markers are used. However, as noted in Section 1, the concept of a modular platform suggests using a chassis line of different sizes. It is undeniable that for a chassis with a large work area, the camera height will be different, and therefore the optics too. As a result, it is necessary to identify a relation that, when knowing the size of a work area and a camera height, can

Authorized licensed use limited to: St Petersburg Natl Uni of Info Tech Mech & Optics. Downloaded on January 09,2021 at 17:20:13 UTC from IEEE Xplore. Restrictions apply.

determine the optimal marker size, which allows achieving a positioning accuracy not lower than in the experimental setup.

It should also be noted that optical markers can become dirty when equipment is in operation. The maintenance schedule should describe the procedure of marker cleaning. This operation is not technically complicated and is usually performed by staff with a low-skilled labor.

In machine vision systems, where high-precision object location is required (3D scanning, automated high-speed robotic lines, etc.), special shadowless light sources are used, which, depending on their configuration, allow better recognition of the object contours or its volume. The proposed approach is based on the use of augmented reality tools that were originally designed to work in adverse lighting conditions. Therefore, at the current stage of development, diffuse LED lamps are used to illuminate the work area, and the selection of specialized light sources is considered inadvisable.

Changing the method of mounting modules on a carriage suspension made it possible to use two modules simultaneously. However, it is not currently implemented in the software. In the future, it is planned to add this feature to obtain hybrid devices that work with two modules simultaneously. However, it will also be necessary to further divide modules into classes according to the principle of processing (measuring) compatibility.

Since all components in a modular 3-axis platform, as well as instrumental equipment, are provided with coordinate markers, known by the system, it is possible to use external augmented reality tools. In particular, it is planned to create an application for mobile devices and augmented reality glasses. Such system will allow creating interactive manuals for modules installation and configuration, visualizing the machining process (for example, to create an animation of a cutting tool movement, or as a visual progress indicator for three-dimensional printing), as well as displaying additional information about the machining or measuring process (for example, three-dimensional heat map for a coordinate measuring machine or sensors readings).

VII. Conclusion

In conditions of increasing the flexibility of production equipment, a carefully thought-out technology for securing interchangeable modules plays an important role, and helps minimizing the operator's efforts to install, fix and calibrate interchangeable devices on a 3-axis platform. The current stage of an experimental modular platform development is focused on methods of equipment reconfiguration with the possibility of its automatic preparation for operation.

In the paper, a novel method for plug-in module installation on a 3-axis platform carriage, which uses a machine vision system for a subsequent tool registration and calibration, is proposed. As an input to the system, a set of labels is used, located directly on the plug-in module and processing head, and on the desktop, which allows calculating the relative positions of the module, carriage, and work area relative to each other. The calculation accuracy, undoubtedly, depends on both the accuracy of camera mount and on the resolution of a frame. As a mounting method, it is suggested to use electromagnets located on a platform suspension and providing a clamping force sufficient for the module to perform its direct functions without changing its location. In the future, it is planned to develop this approach and conduct a series of experiments with various types of equipment.

Acknowledgment

This work was financially supported by Government of Russian Federation, Grant 08-08.

References

[1] A. Dömel, S. Kriegel, M. Brucker, and M. Suppa, "Autonomous pick and place operations in industrial production," in *2015 12th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI)*, October 2015, pp. 356–356.

[2] Y. Zhong, X. Fengyu, and W. Yue, "Analysis and experiment of workpiece quality detection based on industrial robot," in *2016 23rd International Conference on Mechatronics and Machine Vision in Practice (M2VIP)*, November 2016, pp. 1–6.

[3] W. Zuxiang, Z. Lei, and F. Junpeng, "Design of safety capacitors quality inspection robot based on machine vision," in *2017 First International Conference on Electronics Instrumentation Information Systems (EIIS)*, June 2017, pp. 1–4.

[4] S. Stankov, S. Ivanov, and T. Todorov, "An application of deep neural networks in industrial robotics for detection of humans," in *2019 IEEE XXVIII International Scientific Conference Electronics (ET)*, September 2019, pp. 1–3.

[5] K. Winge, R. Haugaard, and T. Merritt, "Val: Visually augmented laser cutting to enhance and support creativity," in *2014 IEEE International Symposium on Mixed and Augmented Reality - Media, Art, Social Science, Humanities and Design (ISMAR-MASH'D)*, September 2014, pp. 31–34.

[6] T. Kikuchi, Y. Hiroi, R. Smith, B. Thomas, and M. Sugimoto, "Marcut: Marker-based laser cutting for personal fabrication on existing objects," in *TEI 2016 - Proceedings of the 10th Anniversary Conference on Tangible Embedded and Embodied Interaction*. Association for Computing Machinery, Inc, 2 2016, pp. 468–474.

[7] X. Li, W. Liu, Y. Pan, H. Li, X. Ma, and Z. Jia, "A monocular-vision-based contouring error detection method for cnc machine tools," in *2018 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, May 2018, pp. 1–6.

[8] G. Kiswanto and D. Ariansyah, "Development of augmented reality (ar) for machining simulation of 3-axis cnc milling," in *2013 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, September 2013, pp. 143–148.

[9] M. Y. Afanasiev, Y. V. Fedosov, A. A. Krylova, and S. A. Shorokhov, "Modular industrial equipment in cyber-physical production system: Architecture and integration," in *Proceedings of the 21th Conference of Open Innovations Association FRUCT*, November 2017, pp. 3–9.

[10] ——, "An application of microservices architecture pattern to create a modular computer numerical control system," in *Proceedings of the 20th Conference of Open Innovations Association FRUCT*, April 2017, pp. 10–19.

[11] ——, "Problems of trajectory building during laser CNC processing," in *Proceedings of the 20th Conference of Open Innovations Association FRUCT*, April 2017, pp. 585–591.

[12] ——, "Machine vision for selective polymer curing devices: Challenges and solutions," in *Proceedings of the 21th Conference of Open Innovations Association FRUCT*, November 2017, pp. 391–397.

[13] Opencv official site. [Online]. Available: https://opencv.org/

[14] G. Bradski, "The OpenCV Library," *Dr. Dobb's Journal of Software Tools*, 2000.

[15] G. Yu, Y. Hu, and J. Dai, "Topotag: A robust and scalable topological fiducial marker system," *arXiv preprint arXiv:1908.01450*, 2019.

[16] S. Garrido-Jurado, R. Muñoz-Salinas, F. J. Madrid-Cuevas, and M. J. Marín-Jiménez, "Automatic generation and detection of highly reliable fiducial markers under occlusion," *Pattern Recognition*, vol. 47, no. 6, pp. 2280–2292, 2014.

Authorized licensed use limited to: St Petersburg Natl Uni of Info Tech Mech & Optics. Downloaded on January 09,2021 at 17:20:13 UTC from IEEE Xplore. Restrictions apply.

Modular Approach in CNC Kernel Development

Ksenia Zimenko, Maxim Afanasev, Yuri Andreev,
Anastasiya Krylova, Sergey Shorokhov, Yuri Fedosov, Mikhail Kolesnikov
ITMO University, St. Petersburg, Russia
zksenia@yahoo.com, amax@niuitmo.ru, ysandreev@itmo.ru, {ananasn94, stratumxspb}@gmail.com,
yf01@yandex.ru, km@hexaxis.ru

***Abstract*—Most small design bureaus have to order prototypes from third-party organizations which elongates development time. A possible solution to the problem may be using modular equipment which reduces the cost of obtaining the required machines and development time. However, such equipment needs a computer numerical control system that could quickly adapt to hardware requirements. An approach to control system design is proposed that implies the development of numerical control kernel from independent modules interacting via a unified programming interface with a high level of granularity which will allow rapid development of a required configuration. The possibility to use existing open control systems as a basis is also considered in the paper, which can minimize the design time. The solution is based on a multiprotocol control system and provides the possibility to combine software and hardware components from different manufacturers. A motion planning module was developed based on the proposed approach which can be embedded in an open system Smoothieware and expand the its capabilities in terms of trajectory complexity and quality of obtained surface. Simulations of the motion planner performance were carried out for linear and complex trajectories and showed minimal contouring and linear errors. The work is aimed at increasing the economic independence and competitiveness of small design organizations and enterprises. The proposed modular approach allows obtaining the required equipment and its control system with minimal design time which can significantly expand the capabilities of rapid prototyping and ensure the prompt production of pilot batches.**

I. Introduction

Among the main problems faced by small organizations and design bureaus, also called start-ups, that do not have their own manufacturing premises is the development of prototypes and pilot batches.

These organizations are most often forced to use services of third-party manufacturers [1]. However, this decision leads to an increase in time and cost of production. Since the distinguishing feature of these design bureaus is the development of innovative products with the minimal design and launch time, the success of the project relies greatly on the speed of prototype development and tests.

Another possible solution to this problem for an organization is to purchase the required equipment and create its own production site. This decision also has a number of disadvantages. Firstly, due to current trends in the field of instrumentation, as well as the wide introduction of the Industrial Internet of Things, most of products have complex geometry, as well as electronic and electric components. Development of this type of products requires the use of a wide number of expensive specialized equipment, purchasing which may be unprofitable or even financially unavailable for a small organization. This decision is also considered impractical due to rapidly changing nomenclature.

The optimal solution may be in the use of modular equipment, which allows obtaining the required type of a machine tool by replacing individual physical components i.e. modules. It makes it possible to get the equipment needed for a specific project. This way, a small enterprise can achieve the required level of economic independence and produce prototypes in a short period of time without the involvement of third-party organizations and with minimum costs.

It should be noted that, since modular equipment provides the possibility of various types of processing, a numerical control system (CNC) suitable for each type of machine tool is required for its successful operation. Therefore, the problem of obtaining CNC systems for each type of required processing arises. Possible solutions include the following:

- The use of specialized commercial CNC systems.
- Development of a CNC system from scratch.
- The use of CNC systems with open source code.

Purchasing a commercial CNC system, despite its obvious advantages like high processing speed and precision, can be considered inexpedient. Most of CNC systems are designed either for one type of processing or for a small variety most often consisting of milling, laser cutting and turning [2]–[4]. And since the modular equipment capabilities include a much wider selection of possible machine tools, it requires the purchase of multiple systems to control each of these units. This, together with the high cost of specialized CNC software, makes this solution impractical.

Another solution may include the development of control systems for each type of processing individually. However, this will significantly increase the development time. Therefore, the main approach in this case is universality, which is applied in the proposed solution.

Finally, it is possible to introduce a CNC system with open source code. It allows obtaining a machine control system with minimal time and monetary loss. Open architecture allows future modifications and upgrades to meet the needs of users and accepts machine programs from a wide variety of software packages. However this leads to a few drawbacks that need to be addressed. Firstly, the capabilities of open source systems (GRBL, Smoothieware), are most often limited to milling, laser processing and three-dimensional printing [5], [6]. And

ISSN 2305-7254

Authorized licensed use limited to: St Petersburg Natl Uni of Info Tech Mech & Optics. Downloaded on May 06,2021 at 12:56:41 UTC from IEEE Xplore. Restrictions apply.

it leaves the same problem of obtaining the CNC software for other types of equipment. Apart from that algorithms used in most open source systems result in low machining accuracy and surface quality and need be improved. In the present paper a way to overcome the listed disadvantages while using open source CNC system is suggested i.e. to both achieve the solution that is universal for every type of machining and increases the processing efficiency.

The paper proposes the use of a modular approach in the CNC software development, when all the main functions of the CNC system are designed as separate program modules interacting via a unified program interface (Application Programming Interface, API). There are several developments in this area [7]–[9]. The distinctive feature of this particular approach lies in a high level of granularity, i.e. division of a CNC kernel into modules, including the trajectory planning stage and embedding it into existing CNC systems with open source code as a basis. Thus, the following advantages can be achieved:

- Ensuring the operation of various types of modular equipment on demand by integrating the CNC system from the required modules.
- Possibility of continuous improvement and modification of the system by replacing modules that affect processing efficiency.
- Minimization of development cycle.

The paper aims to obtain the part of a numeric control kernel (NCK) responsible for the trajectory planning, based on a modular approach. It is intended for application on universal modular equipment. The paper also considers the possibility of using an open CNC system as a basis for the modular kernel. The task at this stage of development is to obtain a motion planner for milling and laser processing in such a way as to ensure the possibility of importing the developed modules into the existing open source CNC system.

The paper is organized as follows. The first Section is dedicated to analysis of modern trends in CNC systems and universal equipment development. In Section II the main aspects of the proposed modular approach are disclosed and its advantages are listed and explained. In Section III the obtained trajectory planning library and its structure is described. In Section IV the possibility to import additional and replace existing modules into open CNC systems to ensure the effective operation of different types of processing is considered. Section V shows the universal modular equipment in development and simulation results of the obtained trajectory planner. Finally in Section VI further clarifications to the work of the developed motion planner are given, as well as directions for the ongoing development of algorithms.

II. Related Work

An active development of the methodology for creating universal equipment is currently taking place. The aim of most researches is to combine the capabilities of additive and subtractive machining, as well as other types of technological equipment.

Among the existing developments of universal equipment there can be mentioned a combination of Rapid Prototyping (RP) equipment (three-dimensional printer) and a milling machine [10]–[12]. Apart from that, there are works that aim at combining laser and milling processing in one machine tool [13]. For example, in [14] a scheme for combining a CNC system to produce hybrid equipment for Fused Deposition Modeling (FDM) printing and three-axis milling was proposed. However, since, in this case, neither hardware nor software was obtained using the modular approach, its application in other types of processing is significantly limited.

On the contrary, the capabilities of the modular equipment considered in the paper make it possible to create not only subtractive and additive machine tools, but also robots, sorting and marking machines, and other technological units. Obviously, due to the fact that the machine organs are reinstallable and are not fixed rigidly, the processing accuracy is lower than when using specialized equipment. However, since a prototype is usually being developed to show the general performance of a product, the capabilities of modular machines are sufficient enough for this purpose.

The requirements for the CNC system that can be used in modular equipment are directly related to the main areas of development in this field and include minimizing the development cycle by ensuring a cross-platform approach, using an open architecture and the possibility of code reuse [7], [15].

Grigoriev and Martinov [7], proposed an approach for scaling channels control during data transfer, which allows reduction of interpolation and cycle time of a programmable logic controller (PLC), aimed at multi-axis processing. It leads to the acceleration of data processing in a CNC system. However, this approach does not support scalability in terms of using the resulting CNC system for other types of machining other than multi-axis milling.

The closest to the proposed solution is a cross-platform CNC kernel obtained by Grigoriev and Martinov [16], which can adapt to multi-axis machining. Also in [17], a model of the control system of an assembly robotic system is proposed with the ability to customize the CNC functions for the required physical configuration. In this case, the core of the control system represents a single module, but the composition of the core itself is constructed as a black box, and modification of the trajectory planning and interpolation methods is not considered.

A. Methodology

An analysis of existing methods for the development of control systems and the possibility of their application in modular equipment was carried out.

The solution proposed in this paper includes the possibility to integrate a CNC kernel from separate modules-blocks with unified input data. When solving this problem, it is necessary to determine the required level of system granularity [18]. Determination of the optimal level of detailization affects the entire system architecture and should be defined at the design

Authorized licensed use limited to: St Petersburg Natl Uni of Info Tech Mech & Optics. Downloaded on May 06,2021 at 12:56:41 UTC from IEEE Xplore. Restrictions apply.

stage [19], [20]. At its minimum level, the system consists of a single unit, this approach does not imply modularity. At the maximum level of granularity, each small component of the system is handled individually, resulting in an extremely fine-grained system and non-integrated design [20]. In this work, at the highest level of detail, the program code is presented as a single motion planning module, and at the lowest level modules responsible for geometry analysis, speed control and interpolation can be distinguished.

The proposed modular solution will allow not only the possibility of fast configuration of a CNC system for the required equipment, but wide space for continuous improvement of processing efficiency by importing and modifying the NCK modules responsible for processing accuracy.

The method is based on the synthesis of the developed motion planning modules and the existing CNC system. Considering the architecture and the possibility of expanding the source code, the Smoothieware system was chosen for this study [5].

A motion planner was developed for its use in laser processing. The simulation of its performance was carried out using the Python programming language version 3.6.5, charts are made with the Matplotlib plotting library version 2.2.2 and calculations are made using Numpy library version 1.14.2. CPU's clock speed is 2.20 GHz, Random Access Memory volume is 4 Gb, operation system is 64-bit Windows 10.

III. Modular approach in CNC development

The possibilities of modular equipment are wide and allow obtaining different installations and machines by replacing physical modules. The potential is not limited to milling, turning, drilling and laser processing: it is possible to obtain a 3D printer, marker, sorter, industrial robot, etc. The operation of all types of machines is controlled by a CNC system.

From a functional perspective, a CNC consists of a Human-Machine Interface (HMI), a Numerical Control Kernel (NCK), and a Programmable Logic Control (PLC) [21]. The HMI is an interface between the CNC and the user that executes machine control commands, displays its status and offers functions for editing part programs for processing. The PLC block sequentially controls the spindle speed of the machine, change of workpiece, tool and processing of I/O signals. It controls the machine behavior, with the exception of servo control.

Finally, the NCK block interprets part programs and performs interpolation, position control, and error compensation. This unit drives and controls the machine servo drives. The paper's main focus is on the development of NCK.

Evidently, each type of equipment requires its own specific functions to be performed by the control system. Fig. 1 shows an example of needed trajectory planning modules for laser, milling and 3D printing. For each type of processing, certain sets of modules are preferred. Since the use of complex paths is more typical for laser processing and three-dimensional printing, in this case it is preferable to work with curves directly, without preliminary segmentation. Usually Non-Uniform Rational B-Splines (NURBS) are used. On the other hand, to reduce the computational complexity in milling, the approximation of curved paths by linear segments is often used. The use of segmentation or NURBS planning modules depends on specific requirements.

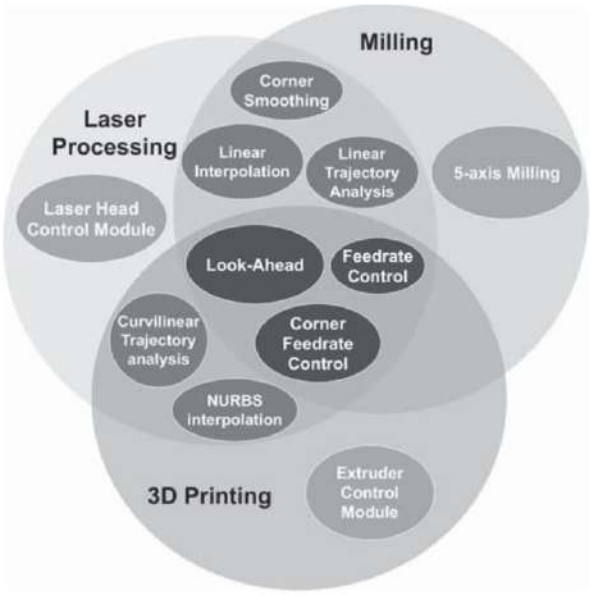

Fig. 1. Sets of modules for milling, laser processing and 3D printing

Apart from that, there are modules that are necessary to ensure equipment efficient operation. For example, the extruder control module allows providing the required material supply for 3D printing using FDM technology.

An approach to the development of CNC with a high level of granularity is proposed in this paper, when not only the main blocks of the system (Fig. 2 right), but also the NCK itself (Fig. 2 left) consist of separate modules-blocks, and the interaction between them is carried out through a unified Application Programming Interface (API).

Technological modules are optional and are used for certain types of processing or operations. For example, if a technology requires additional movement during the molding process (extruder control) or if the movement control is synchronized with devices such as laser head or electron beam gun.

The kinematic transformation module converts the coordinates according to the kinematic model of a machine. The use of the specific kinematic transformation is determined based on the kinematics of the machine.

Commands for controlling outputs and the status of inputs are integrated in the cyclic drive control data and are transmitted via the same interface. The trajectory planning phase is also divided into the following modules:

- Preliminary analysis of geometry.
- Acceleration/Deceleration control, during which the tool speed values are calculated for each interpolation period.
- Interpolation, where the control signals for drives that set the actuator in motion are generated.

Authorized licensed use limited to: St Petersburg Natl Uni of Info Tech Mech & Optics. Downloaded on May 06,2021 at 12:56:41 UTC from IEEE Xplore. Restrictions apply.

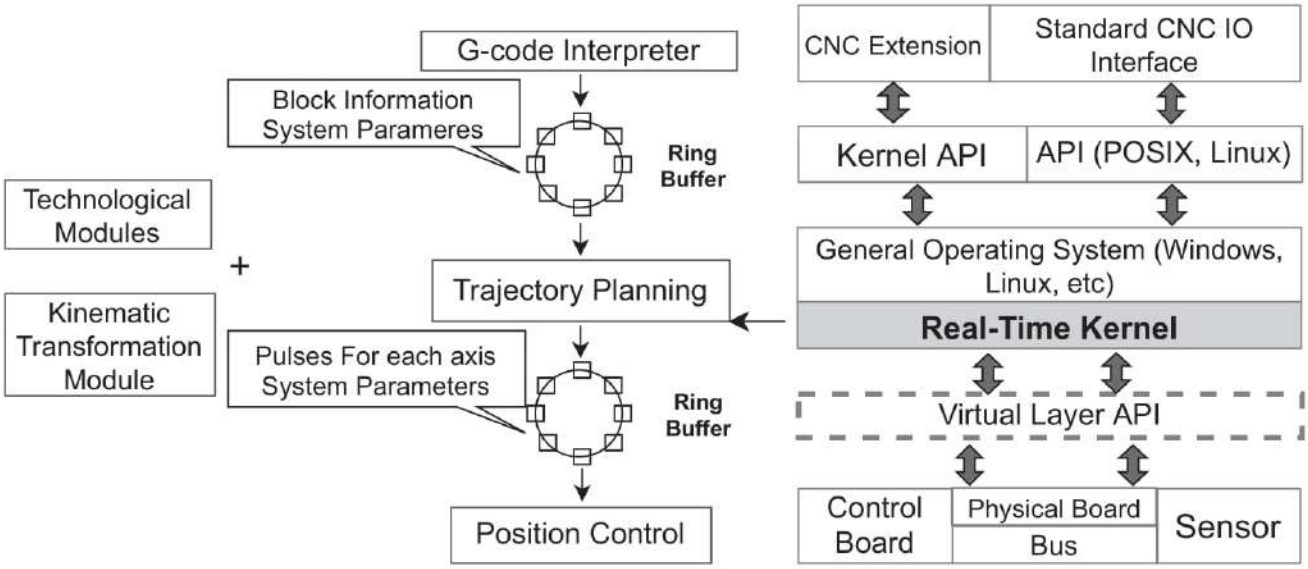

Fig. 2. The structure of CNC system (right) and NCK (left)

It is suggested to reduce the level of granularity and design each of the above stages of trajectory planning as a separate independent module. The main advantage of this modular approach is the ability to easily modify the existing algorithms by replacing modules that have the greatest impact on processing accuracy. Since the precision and speed of processing depends not only on the hardware of a machine, but also on the software component, the constant improvement of NCK algorithms will increase the equipment efficiency and the quality of the obtained part surface.

IV. Modular structure of NCK

As a result of practical aspect of the research, a motion planner software library for laser processing has been obtained, including modules for geometry analysis, speed control and interpolation. The structure of the library is shown in Fig. 3.

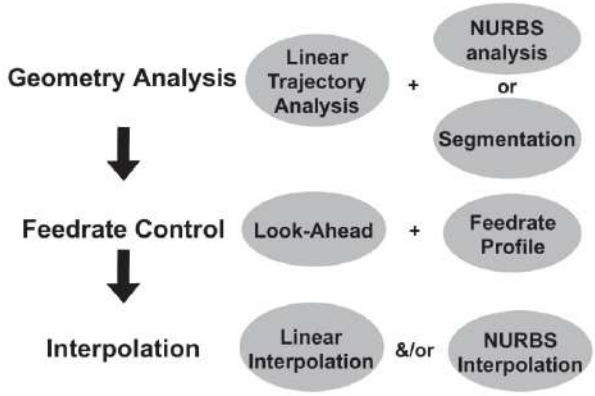

Fig. 3. The structure of developed trajectory planner

The modules can process both curvilinear trajectories and linear ones. The motion planner is designed in such a way that by adding appropriate modules (for example, NURBS interpolation), work with curves will be provided, which is necessary, for example, in 3D printers or in laser processing, where complex paths are often used. On the other hand, if it is replaced with a segmentation module, then all trajectories will be converted to linear segments, even if the original data in the control program is presented as curves, which decreases the time cycle and allows fast real-time processing.

The generation of a tool path begins after receiving data from the Interpreter module, where instructions of a part program are decoded and translated into the parameters of the tool trajectory.

Since the obtained data must be further analyzed before proceeding to interpolation, a set of modules for geometry preprocessing was developed. It includes the calculation of length of trajectory segments, corners smoothing, NURBS curves analysis or its segmentation with linear segments.

It should be noted that the trajectory contains tangential discontinuities at the corners between the linear segments. The module implements rounding with Bezier curves with six control points or exact traversal of the corner with a complete stop of the tool, depending on requirements. The allowable speed on the rounding is also calculated, taking into account the geometry of a parametric curve (curvature extremum) and allowable speed and acceleration. The detailed rationale for the choice of algorithms and methods for corner smoothing is given in one of the previous papers [22].

Also, when analyzing curvilinear trajectories at this stage, the trajectory can be divided into linear segments, if the segmentation module is selected, or the curvature of a NURBS curve can be analyzed to select areas where restrictions on the tool speed should be imposed so as not to exceed the allowable acceleration.

After geometry preanalysis, the speed or accelera-

Authorized licensed use limited to: St Petersburg Natl Uni of Info Tech Mech & Optics. Downloaded on May 06,2021 at 12:56:41 UTC from IEEE Xplore. Restrictions apply.

tion/deceleration control is executed, which consists in calculating the velocity values for each interpolation period T. To ensure accurate movement of the actuator during high-speed machining, the equipment must operate at speeds up to 600 mm/s with acceleration up to 2g. Acceleration/deceleration control consists of two main modules/stages: a speed profile generation of the current segment and Look Ahead algorithm.

First, the Look-Ahead module is executed, which calculates the start and end velocities for the current segment by analyzing the parameters of the next N segments. The N value is defined in the system parameters.

Further, the module for speed profile calculates a gradual change in speed along a segment. Based on the set feedrate, starting and ending velocity, and the allowable acceleration and jerk, the speed values for acceleration, deceleration and constant speed stages are calculated for each interpolation period T.

Finally, the interpolation is executed, which acts as a generator of the axial movement of the actuator based on data obtained during previous planning stages. Linear segments are processed according to the linear Reference Word interpolation algorithm. Curved segments and arcs, if not segmented at the stage of geometry analysis, are processed according to the second-order Taylor's expansion of the curve parameter with respect to the arc-length. Detailed rationale and description of interpolation algorithms used are presented in one of the previous papers [23].

The software is assembled from blocks that interact through a unified API. The use of a cross-platform kernel ensures the software independence from a specific platform and provides ample opportunities for CNC system configuration.

V. Possibilities for open CNC systems modification

One of the advantages of modular CNC development is the ability to integrate the developed modules into existing open source systems. The study considers the possibility to use the Smoothieware CNC system, developed for the 32-bit SmoothieBoard by a team of volunteers [5]. Its main benefit is the ability to extend the existing source code.

The structure of Smoothiewave NCK, input and output at each stage of trajectory planning are shown in Fig. 4. Since the interchangeability of modules in this system is carried out through a unified API, additional modules can be imported or replaced to improve processing efficiency. Fig. 4 also shows possibilities for replacement of existing and import of additional modules of path planning to ensure increased accuracy and processing speed.

This approach is beneficial for the following reasons:

- It reduces the time for software development, since the basis of the CNC system already exists.
- It helps to adjust the system to the required hardware, i.e., when replacing physical blocks in universal modular equipment, by adding necessary programming modules.

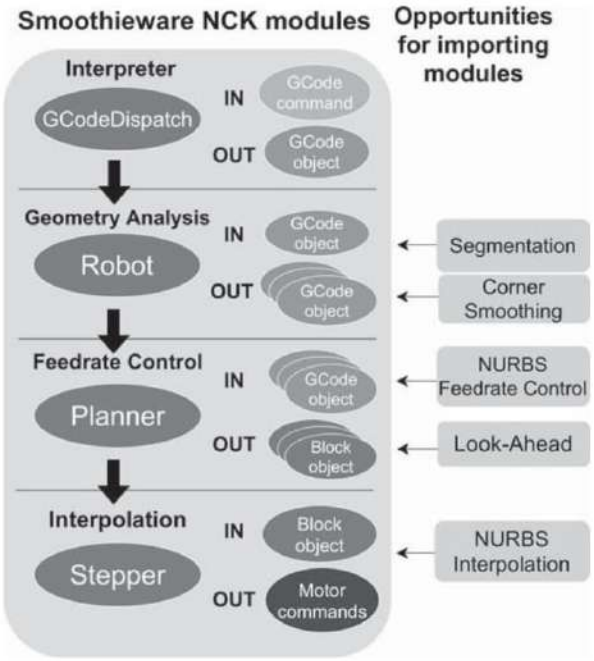

Fig. 4. Possibilities for Smoothieware CNC system modification

- It gives the opportunity to import modified algorithms of part program analysis and trajectory generation to improve the accuracy and speed of machining.

The paper suggests embedding previously developed modules that have the greatest impact on processing accuracy into the existing CNC. Among these modules the tool trajectory generation, feedrate control and interpolation can be mentioned. The interchangeability of modules is carried out through a unified API, which allows fast redesign of a system depending on the hardware requirements. This approach will significantly simplify and improve processing efficiency when using modular equipment.

Since most open systems offer the possibility of upgrading and extending the source code, a similar approach can be applied to systems other than Smoothieware. However, the architecture of this system turned out to be the most suitable for applying the proposed approach.

VI. Simulation results

The suggested approach and modules, described in Section III are used for designing a control system for a prototype of the modular technological equipment platform, which is shown in Fig. 5. At this stage of development, it is possible to implement laser and measuring equipment, three-dimensional printing using FDM technology, as well as milling and drilling processing of aluminum alloys [24]. The modular platform consists of a two-axis table with the possibility of setting an additional third axes. Work space size is 500×500 mm, a servo-step drive is used. The tool moves by means of a ball screw and cylindrical guides with linear rolling bearings.

Authorized licensed use limited to: St Petersburg Natl Uni of Info Tech Mech & Optics. Downloaded on May 06,2021 at 12:56:41 UTC from IEEE Xplore. Restrictions apply.

One ball screw is installed on the X axis, two on the Y axis, synchronized by a belt drive.

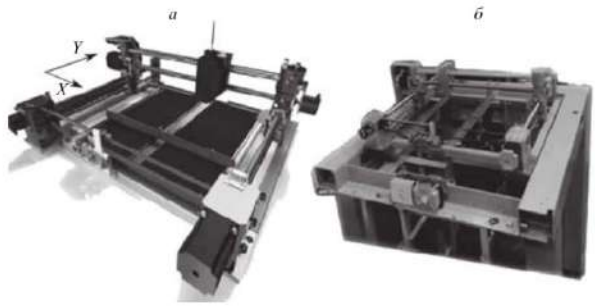

Fig. 5. Platform in Solidworks (a) and assembled (b) [24]

A Smoothieboard circuit board with a Smoothieware control system is used for controlling the platform operation. The modules responsible for the trajectory generation in the system have been replaced with the described above in Section III.

Simulations of the obtained software library were carried out. The following sample tool path has been used as an input: curvilinear trajectory with unit weights and a nodal vector K = [0, 0, 0, 0, 0.25, 0.5, 0.75, 1, 1, 1, 1] (Fig. 6, a). Required feed was set as $F = 50\ mm/s$, maximum acceleration – A_{max} = 2,000 mm/s^2, maximum jerk – J_{max} = 20,000 mm/s^3.

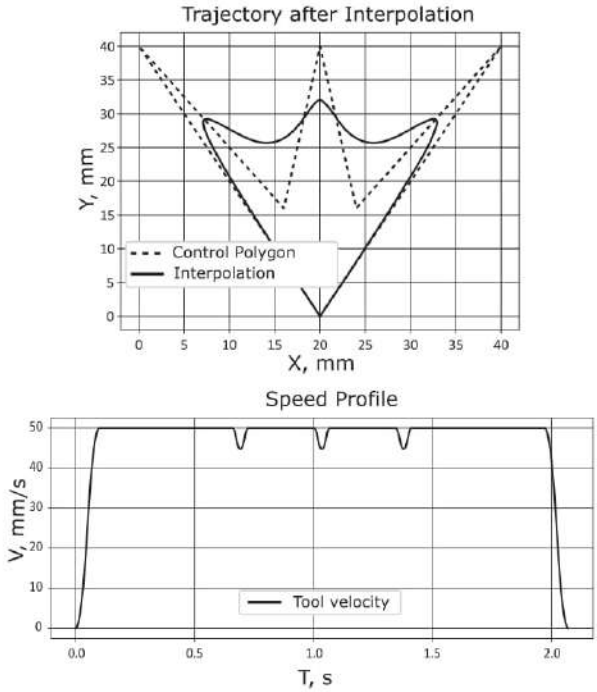

Fig. 6. Simulation example

The curve is interpolated directly, without preliminary approximation. The resulting velocity profile, where V is tool speed, mm/s; T - time, s, is shown in Fig. 5, b. The resulting linear processing error is within 17 μm, and the contour error does not exceed 7 μm, which is significantly less than the error that occurs when planning a similar trajectory in the Smoothieware system, where, among other things, preliminary segmentation of the original curve is required.

The resulting modules can receive data from the Smoothieware G-code interpreter as input. The generated output represents a class instance that contains the speeds and displacements required to generate control signals for drives. In this form, they can be used by the corresponding modules of Smoothieware system to control the tool movement.

VII. Discussion

The proposed approach can significantly increase the competitiveness of small organizations. It should be addressed that, in contrast to the proposed solution, it is possible to develop a universal CNC system that is not based on a modular approach, which will already contain all the required functions for any type of processing. In this case, there is no need to import additional modules. However, without modular approach, the possibility of modifying the algorithms is significantly complicated. And the large number of modules in such system can create difficulties in management and interaction.

The developed planner described in Section III successfully generates trajectories consisting of linear segments, circular arcs and curvilinear NURBS paths. Due to the versatility of the trajectory representation, this module can be applied both for segmented trajectories, and for part programs with direct curves representation.

All received deviations between desired and resultant trajectories are of small order and are comparable with the results of other developments and the performance characteristics of commercial machine tools of the middle price category. A smooth velocity profile with jerk considered, as well as parametric interpolation of curvilinear trajectories with division into equal segments, made it possible to minimize tool speed inconsistencies and vibrations.

Despite the fact that the developed software library was focused on laser processing, it is also possible to use it for milling, for this it is necessary to change some system parameters, such as permissible acceleration and jerk: for milling, their value must be reduced by an order of magnitude.

Finally, to further expand the capabilities of a modular installation, it is necessary to develop modules for other types of processing, including, for example, technological and kinematic transformation modules. It is required to form a list of required modules for the successful operation of each type of modular equipment.

As for practical application, it is assumed that employees of a small design organization could acquire the required modules for a modular platform of technological equipment and the Smoothieware system with a built-in motion planner developed within the framework of this work. In the future, it would be possible to develop other modules and improve the existing ones on the basis of the proposed approach to ensure the operation of various types of equipment and to improve the quality and accuracy of processing. Interchangeability

Authorized licensed use limited to: St Petersburg Natl Uni of Info Tech Mech & Optics. Downloaded on May 06,2021 at 12:56:41 UTC from IEEE Xplore. Restrictions apply.

and a high level of granularity, together with documentation, makes it easier to work with the code in the case of possible development team changes.

Plans for further research include the development of other modules to expand the modular equipment capabilities. Among them is an algorithm for generating drive pulses for motors, as well as an interpreter of G-codes and a feedback system (position control) to improve the existing open source software and possibly use the modules independently, and not as part of the existing CNC system. Also plans for further work include the development of an error compensation system for curved paths to improve processing accuracy.

Further areas of development may also include the modification of a CNC system to a Cyber-Physical Machine Tool (CPMT), which was proposed in [25] to enable continuous monitoring and control of the machining process. It is especially important in the case of modular equipment since the physical modules are not rigidly fixed ant there is higher possibility of change of their relative position during processing, which requires additional monitoring.

VIII. Conclusion

The paper proposes an approach to the development of an NCK of a CNC system, which allows integrating the required software from programming blocks for the use on modular equipment. The possibility of using existing open source CNC system Smoothieware as a basis was also considered and the modules required for each type of machining can be imported there. In this case, there is no need to develop a system from scratch, and it also becomes possible to modify existing algorithms that affect the processing accuracy to improve the efficiency of the processing.

A motion planner, including modules for geometry analysis, acceleration/deceleration control and interpolation, was obtained. The simulations show low contour and linear error while maintaining high processing speed. The resulting modules can be imported into the CNC system Smoothieware to improve the efficiency of existing algorithms, e.g. allow direct processing of complex curvilinear trajectories without negative consequences of segmentation. Also, the application field of the algorithm is not limited to laser processing, but can also be used for milling.

It is proposed to use the obtained planner for modular equipment, when the required installation is formed from separate physical blocks. In this case, it is possible to ensure the operation of the CNC system, depending on the hardware requirements. All the listed advantages of the development distinguish it favorably among the available open source projects.

Acknowledgment

This work was carried out under the project no. 619296 "Technologies of cyber-physical systems: management, computing, security" conducted at the Faculty of Control Systems and Robotics, ITMO University.

References

[1] M. Y. Afanasiev, Y. V. Fedosov, A. A. Krylova, S. A. Shorokhov "Modular industrial equipment in cyber-physical production system: Architecture and integration", *Proc. of the 21th Conf. of Open Innovations Association FRUCT*, 2017, pp. 3—9.

[2] Flexible CNC systems and solutions, Web: https://www.fanuc.eu/hu/en/cnc.

[3] SINUMERIK CNC automation system, Siemens. Ingenuity for life, Web: https://new.siemens.com/global/en/products/automation/systems/cnc-sinumerik/automation-systems.html.

[4] High-Speed Machining, Haas factory, Web: https://www.haascnc.com/productivity/control/hsm.html.

[5] Smoothieware official website, Smoothieboards, Web: http://smoothieware.org/smoothieboard.

[6] Grbl, GitHub Grbl repository, Web: https://github.com/grbl/grbl.

[7] S.N. Grigoriev, G.M. Martinov, "Scalable open cross-platform kernel of PCNC system for multi-axis machine tool", *Procedia CIRP*, vol. 1, 2012, pp. 238–243.

[8] L. I. Martinova, G. M. Martinov, "Organization of intermodule interactions in distributed CNC systems. Models and algorithms for implementation", *Mekhatronika, avtomatizatsiia, upravlenie*, vol. 11, 2010, pp. 50-55.

[9] L. I. Martinova, G. M. Martinov, "Organization of intermodule interactions in distributed CNC systems. Models and algorithms for implementation", *Mekhatronika, avtomatizatsiia, upravlenie*, vol. 11, 2010, pp. 50-55.

[10] W. Lee, C. Wei, S.-C. Chung, "Development of a hybrid rapid prototyping system using low-cost fused deposition modeling and fiveaxis machining", *J. Mater. Process. Technol.*, vol. 214, no. 11, 2014, pp. 2366–2374.

[11] K.P. Karunakaran, S. Suryakumar, V. Pushpa, S. Akula, "Low cost integration of additive and subtractive processes for hybrid layered manufacturing", *Robotics and Computer-Integrated Manufacturing*, vol. 26, no. 5, July 2010, pp. 490–499.

[12] J. Hur, K. Lee, Zhu-Hu, J. Kim, "Hybrid rapid prototyping system using machining and deposition", *CAD Computer Aided Design*, vol. 34, no. 10, July 2002, pp. 741–754.

[13] J.Y. Jeng, M.C. Lin, "Mold fabrication and modification using hybrid processes of selective laser cladding and milling", *J. Mater. Process. Technol.*, vol. 110, July 2001, pp. 98–103.

[14] A.N.M. Amanullah, Murshiduzzaman, T. Saleh, R. Khan, "Design and development of a hybrid machine combining rapid prototyping and CNC milling operation", *Procedia Engineering*, vol. 184, 2017, pp. 163–170.

[15] X.-B. Ma, Z.-Y. Han, Y.-Z. Wang, H.-Y. Fu, "Development of a PC-based open architecture software-CNC system", *Chinese Journal of Aeronautics*, vol. 20, no. 3, 2007, pp. 272–281. doi: 10.1016/S1000-9361(07)60044-2.

[16] S.N. Grigoriev, G.M. Martinov, "Research and development of a cross-platform CNC kernel for multi-axis machine tool", *Procedia CIRP*, vol. 14, 2014, pp. 517–522.

[17] T. AlGeddawy, "A new model of modular automation programming in changeable manufacturing systems", *Procedia Manufacturing*, vol. 11, 2017, pp. 198–206.

[18] H.-P. Wiendahl, H.A. ElMaraghy, P. Nyhuis, M.F. Zäh, H.-H. Wiendahl, N. Duffie, M. Brieke, "Changeable manufacturing – classification, design and operation", *CIRP Annals–Manufacturing Technology*, vol. 56, no. 2, 2007, pp. 783–809.

[19] N. Chiriac, K. Hölttä-Otto, D. Lysy, S. Eun Suk, "Level of modularity and different levels of system granularity", *Journal of Mechanical Design*, vol. 133, no. 10, 2011, pp. 329–339.

[20] T. AlGeddawy, H. ElMaraghy, "Optimum granularity level of modular product design architecture", *CIRP Annals–Manufacturing Technology*, vol. 62, no. 1, 2013, pp. 151–154.

[21] S.-H. Suh, S.-K. Kang, D.-H. Chung and I. Stroud, *Theory and Design of CNC Systems, 1st ed.* Springer Publishing Company, Incorporated, 2008.

[22] K.V. Zimenko, M.Ya. Afanasev, A.A. Krylova, S.A. Shorokhov, Y.V. Fedosov "Motion Profile Control Algorithm and Corner Smoothing Technique for Trajectory Optimization of High-Precision Processing", *Proceedings of the 23rd Conference of Open Innovations Association FRUCT*, 2018, pp. 425-431.

[23] K.V. Zimenko, M.Ya. Afanasev, Y.S. Andreev, A.A. Krylova, S.A. Shorokhov, Y.V. Fedosov, M.V. Kolesnikov, "Interpolation Algorithm for

Authorized licensed use limited to: St Petersburg Natl Uni of Info Tech Mech & Optics. Downloaded on May 06,2021 at 12:56:41 UTC from IEEE Xplore. Restrictions apply.

High-Speed Processing of Complex Curvilinear Trajectories", *Proceedings of the 25th Conference of Open Innovations Association FRUCT*, 2019, pp. 398-404.

[24] M.Ya. Afanasiev, Yu.V. Fedosov, A.A. Krylova, S.A. Shorokhov, "Microservice architecture application in the design of industrial equipment with computer numerical control", *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, vol. 18, no. 1(113), 2018, pp. 87–97. (in Russian).

[25] C. Deng, R. Guo, P. Zheng, C. Liu, X. Xu, R.Y. Zhong, "From open CNC systems to cyber-physical machine tools: A case study", *Procedia CIRP*, vol. 72, 2018, pp. 1270–1276.

Authorized licensed use limited to: St Petersburg Natl Uni of Info Tech Mech & Optics. Downloaded on May 06,2021 at 12:56:41 UTC from IEEE Xplore. Restrictions apply.