International Journal of Engineering and Technical Research (IJETR)
ISSN: 2321-0869 (O) 2454-4698 (P) Volume-7, Issue-9, September 2017

Information Technology for Processing of Industrial Big Data with Distributed Infrastructure on the Basis of Smart Agents and Parallel Algorithms

Bogdan Poddubny, Andrey Kupin, Ivan Muzyka, Oleksandr Savytskyi, Sergii Ruban, Vadim Kharlamenko

***Abstract*— The paper presents a brief description of engineering and scientific problems aimed at implementation of IT for industrial Big Data processing with a distributed infrastructure on the basis of smart-agents and parallel algorithms. Emphasis is given to innovative methods based on smart agents and principles of Industry 4.0. Implementation and simulation of parallel algorithms for Big Data processing was carried out.**

***Index Terms*— Industry 4.0; Big Data; Smart-agent; parallel algorithms.**

I. INTRODUCTION

The question of implementing of intelligent management of repairs and maintenance services in large industrial enterprises on the basis of new approaches within the modern concept of Industry 4.0 is considered. Actuality of the problem and potential ways of its solving are presented in previous papers of authors [1-2].

Modern trends in the development of scientific and technical progress for the world industry quite often describe in such terms like «Smart Factory», «Smart Manufacturing», «Intelligent Factory» and «Factory of the Future». Now the development of these research areas is well enough formalized by the concept of the 4th Industrial Revolution (Industry 4.0) [3]. The implementation of the concept is related with the use of some key technological trends such as Big Data processing, cyber-physical systems, autonomous robots with different intelligent sensors, simulators for 2D- and/or 3D-modeling, 3D-printers, Internet of Things, augmented reality etc. [4]. Thus, according to estimates of leading world experts, these tendencies will determine the main vector of modern competitive industries [3-5].

II. PROBLEM STATEMENT

We have examined the technological complex of automated section for monitoring, repair, calibration, search and replace of unfit electronic equipment. Figure 1 shows the receiving and dispatch of equipment that needs of maintenance and repairing from/on vehicles used for transportation of equipment both between the workshops of one company and between the workshops of another companies.

Bogdan Poddubny, Kryvyi Rih National University (Post Graduated Student), Ukraine, 50027, V. Matusevicha str., 11

Andrey Kupin, Ivan Muzyka, Oleksandr Savytskyi, Sergii Ruban,Vadim Kharlamenko, Kryvyi Rih National University (Faculty of Information Technology), Ukraine, 50027, V. Matusevicha str., 11

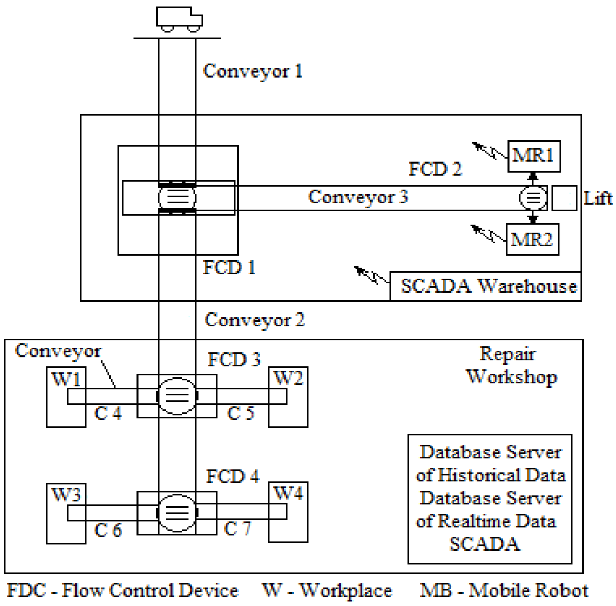

Figure 1. Scheme of the location of equipment and devices of warehouse reception system and repair workshop

Initially, the cargo gets on a roller conveyor R1 and reaches the flow control device (FCD1). It is a rolling rotary table which sorts the flow of cargo in the direction of the warehouse to the conveyor C3 or repair workshop to the conveyor C2. At the warehouse obtained electronic devices are put on shelves by mobile robots 1 and 2 (MR1, MR2). Loading of robots is carried by flow control device FCD2 at the end of Q3 of the conveyor, where also installed a lift L, which sends the devices with corresponding address to the 2nd floor, where another mobile robots work. At the repair workshop the sorting of processed devices to concrete workplaces Wn is carried out by FCD3 and FCD4. Delivery to workplaces is carried out by roller conveyors C4, C5, C6 and C7. Workplaces are equipped with displays (information boards). Each workplace is connected to the real-time database server of SCADA-system and the database server of repair history. Functioning of connections between objects is implemented in 2 ways: mobile robots are connected with SCADA system of warehouse by wireless network. SCADA of warehouse system, as well as SCADA of repair workshop, FCD1-FCD4, W1-W4 are connected by fixed network. Proposed modernization of the system will be based on the use of ideology with traditional multi-rank communications system or a progressive peer-to-peer (P2P) fast network of direct access to the data of the object and data processing points.

Thus it is necessary to solve the following tasks:

1. Parallel processing:

1.1. Finding analogues of appliances, parts.

1.2. Evaluation of the degree of wear on the basis of regression analysis and decision making on necessity of repair or replacement.

1.3. Ordering into logistics.

2. Development of monitoring and control algorithms taking into account real delays in hardware, as well as the delays caused by the necessity of process information. For instance, data about processing of certain signals:

2.1. Reading of RFID tags – 2 ms.

2.2. Reading the instrument parameters – 70–100 ms.

2.3. Data transfer via Ethernet100 – 0.5–1 microseconds.

2.4. Reading from the server - 2 microseconds.

3. Determination of bandwidth peer-to-peer network considering a large number of types of devices and their accessories (more than 16000) and simultaneously processing of tasks of determination the need for purchase of the device or driver, software, terms of deliveries, etc.

4. Logistics of delivery to the warehouse or workshop for calibration.

5. Data acquisition and filling of measurement database for diagnostics of apparatuses using regression analysis.

6. Implementation of diagnostics without removing the device from the object.

The formulation of the problem of the implementation of intelligent management of repairs using cyber-physical systems (smart-agents) was made in [1] as applied to metallurgical enterprises (in the condition of PJSC "ArcelorMittal Kryviy Rih"). As a result, structural schemes have been presented and generic algorithms (Fig. 2-3) of implementation of similar approaches within the Industry 4.0 concept have been developed [5-7].

Taking into account the substantial computational complexity of the problem it will be discussed the implementation of these parallel algorithms and subsequent computer simulation of such technology in conditions of Big Data.

III. Methodology of Parallelization

The proposed algorithm uses modern approaches for handling data streams using the technology of parallelization (Fig. 2).

For this purpose, algorithm provides decomposition of task for parts processing. After forming data streams for parallel processing, mutex mechanism is used, to avoid conflicts between parallel processes that try to access to the shared data. After entering into the critical section (function mutex lock ()) data processing is occurred, and upon completion of work on a common resource is carried out with a critical section by calling mutex lock().

After processing of parts they are transferred to the warehouse (Warehouse Agent). Information that is obtained by this processing via PLC (PLC) is transferred for further processing and storage in a database that is built according to the standards ISA 95, ISA 88, ISO 22400.

For task management and forecasting that are difficult to formalize, to increase performance of a computer system situational control should be used. This approach is based on the fact that for each value of the vector that describes the current situation three is the known value of the vector that describes the solution that should be taken. If it is impossible to describe all situations, then you need to use interpolation.

The initial situation is characterized by the vector $X = \{x_1, x_2, \ldots, x_n\}$. Y decision is made by the X vector components. Y is also vector $Y = \{y_1, y_2, \ldots, y_m\}$. Associative memory is used by this approach. The main idea is based on confronting each X situation with Y decision.

It is advisable to carry out education of the system with precise model, which uses the most accurate calculation of component solutions. Applying the model to study not only used for specifically prescribed by step training system, but also beyond the real cycle management, along with functional control. Algorithm of its work shown in Fig. 3.

In this block diagram Y* – vector control actions that are precisely designed for model, ε – specified precision, r(Y, Y*) – distance between vectors Y and Y*.

The figure shows the following blocks:

- block 0 – simulation modeling of technological situation;
- block 1 – determining the decision from associative memory according to the simulated technological situation;
- block 2 – calculation of new solutions using certain model;
- block 3 – determining the optimal new solutions that meet specific criteria: $r(Y, Y^*) \leqslant \varepsilon$;
- block 4 – adding new solutions to knowledge base.

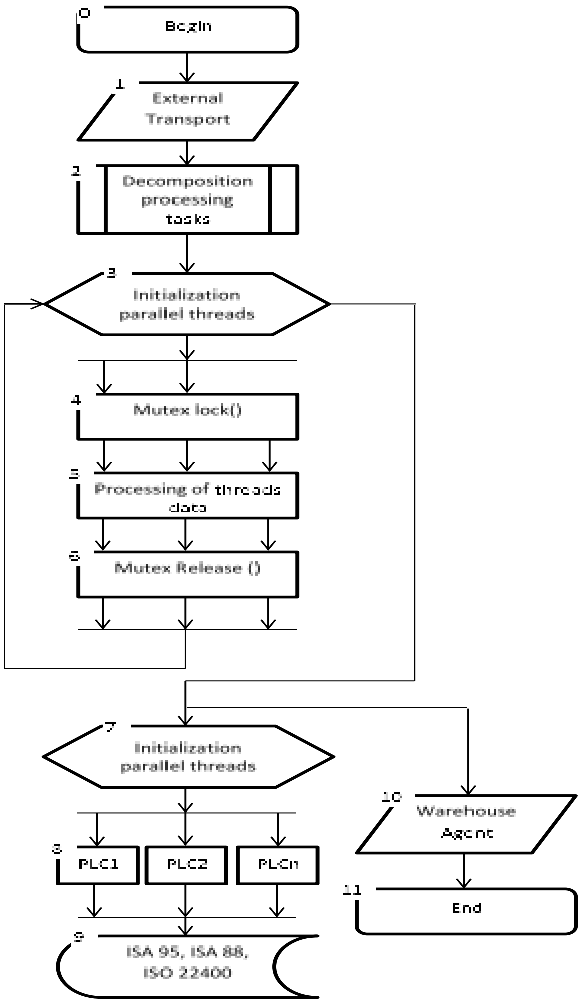

Figure 2. Parallel algorithm of smart information system functioning

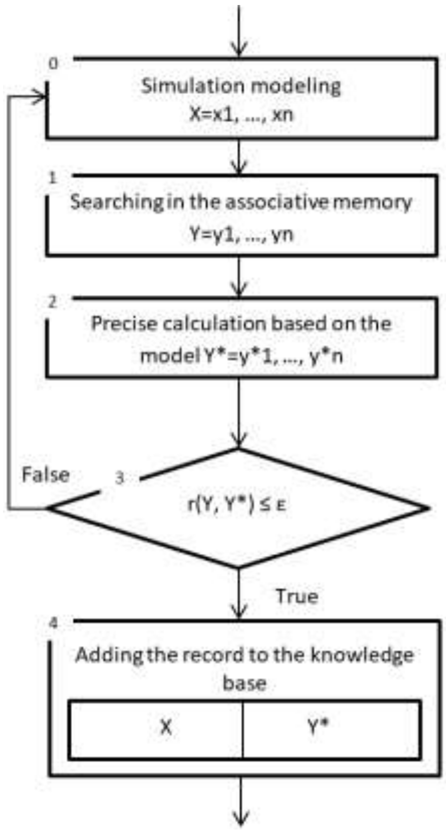

Figure 3. Fragment of block diagram using for control and predictive tasks

Search counterparts implemented by using combinatorial methods. Combinatorial search problem refers to the final of combinatorial sets whose elements are combinatorial objects. These objects are combinations of elements of other final (perhaps too combinatorial) sets – a combination, permutation, partition, cover etc. [8].

Almost all of combinatorial algorithms have exponential complexity [8]. However, in practice it is important to be able to solve combinatorial problems rather large dimensions. One possibility of reducing the time of combinatorial problems solution is to break combinatorial set of classes and explore these classes in parallel, i.e. simultaneously on multiple processors of a computer system. The problem is how to break combinatorial set of classes and handing out the latest processors so as to ensure the highest degree of acceleration computing. The latter depends on many factors, including uneven load on the processor system of overhead communication (especially important for systems with distributed memory) of redundancy viewed search area and others. Paper [8] provides an overview of existing methods for the development of parallel algorithms for solving combinatorial problems.

Suppose objects are defined with vectors of finite length (a_1, a_2, ..., a_n) with components of finite ordered sets A_1, A_2, …, A_n ($a_i \in A_i$) [8]. This method of presentation is for all of combinatorial objects. We assume that the objects are ordered according to the lexicographical order corresponding vectors, and they start with the prefix 1. The object (which is its vector) (a_1, ..., A_n) is the vector (a_1, ..., a_t), $0 \le t \le n$. If t = 0, then we have an empty prefix (). Suppose p = (a_1, ..., A_{t-1}) is a prefix. Denote N(p) the number of objects that have the prefix p. Many applicants for the following position on p, that is, a set of values that can take element, let Z(p). The following expressions give algorithms for calculating the number of object and construct the object by its number [8].

This task assumes that combinatorial objects sorted according to the lexicographical order corresponding vectors are numbered and natural numbers. Then number I of object (a_1, a_2, ..., A_n) is calculated by formula (1)

$$I = \sum_{t=1}^{n-1} \sum_{x \in Z(a_1, a_2, \ldots a_{t-1}), x < a_t} N(a_1, a_2, \ldots, a_{t-1}, x) + \\ + \sum_{x \in Z(a_1, a_2, \ldots a_{n-1}), x \le a_n} N(a_1, a_2, \ldots, a_{n-1}, x) \quad (1)$$

Vector p = (a_1, a_2, ..., a_{t-1}) is the prefix of object with the number I in lexicographical order, Z(p) = {z_1, z_2, ..., z_m} i N – the number of the first object with the prefix p. Then $a_t = z_k$, where k is such that the inequality is true (2)

$$N + \sum_{j=1}^{k-1} N(a_1, \ldots, a_{t-1}, z_j) \le I < \\ < N + \sum_{j=1}^{k} N(a_1, \ldots, a_{t-1}, z_j) \quad (2)$$

To apply the method to the specific combinatorial objects it is need to calculate the number N(p) and Z(p). The paper [9] shows the numbering of the method for the parallel transfer of combinations and permutations of set partitioning. Using the numbering when paralleling for high performance with an apportionment calculations and no exchange of data between processors, as evidenced by the results of computational experiments. For combinations of parallel transfer efficiency averages 0.87, permutations – 0.94, the efficiency of the parallel algorithm partitioning the transfer set is in the range of 0.8-0.85 [8].

Acceleration (speed up factor) of parallel algorithm in the N-processor system is determined by the expression

$$S(N) = T_1 / T_N, \quad (3)$$

where T_1 – execution time of the algorithm on one processor or single-processor system; T_N – execution time of the algorithm on multiprocessor system with N processors.

Execution duration of the program on a parallel system with N processors was estimated by the formula:

$$T_N = f * T_1 + (1 - f) * \frac{T_1}{N}, \quad (4)$$

It is defined acceleration rate (Speedup) of the system with N processors on the basis of the expression

$$S_N = \frac{T_1}{T_N} = \frac{N}{1 + f * (1 - N)}. \quad (5)$$

Since $0 < f < 1$, following dependency is fair:

$$1 \le S_N \le N, \quad (6)$$

There is a measure of the acceleration achieved in respect of a maximum efficiency of the system with N processors

$$E_N = \frac{S_N}{N}. \quad (7)$$

CPU usage as well as the rate of acceleration decrease with increasing of sequential part of the program.

Experimental research suggests the presence of a minimum in the dependence of time on the number of search threads. (Fig.

4). It is used test database with 10000 records of various devices, each of it has 5 parameters. Simulations carried out on a computer with a 4-core processor Core i5. As can be seen from the graph, the minimum processing time information coincides with the condition when the number of threads equals the number of processor cores. At excessive increase in the number of threads and also increased the overhead of context switching threads and due to the limited number of physical calculators (nuclei) begins to increase processing time.

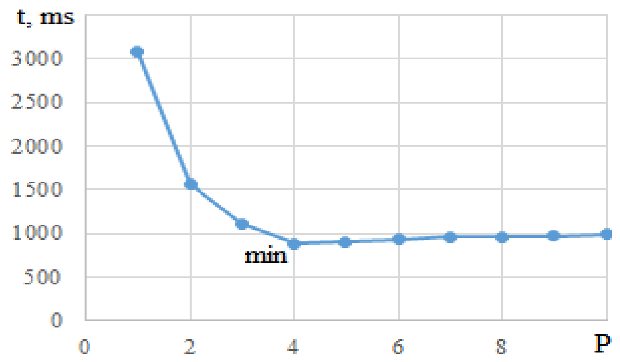

Figure 4. Dependence of search time t on the number of threads P, which are used for parallel processing

Fig. 5 shows dependence of time processing on the number of records in the database. This treatment was carried out in single-threaded mode. The dependence is linear – the more entries in the database require more time for search.

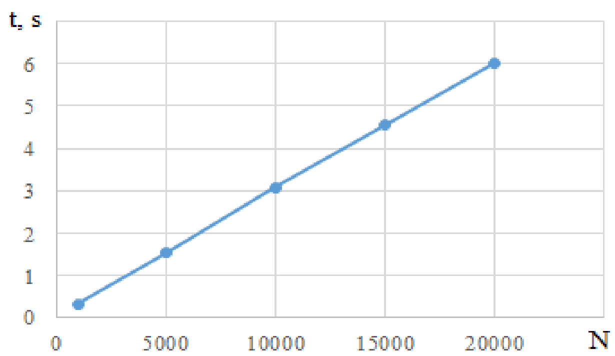

Figure 5. Dependency of search time t on the number of records in database N

Fig. 6 shows the dependency of the processing time on the count of parameters describing a device. Database size is 1000 devices.

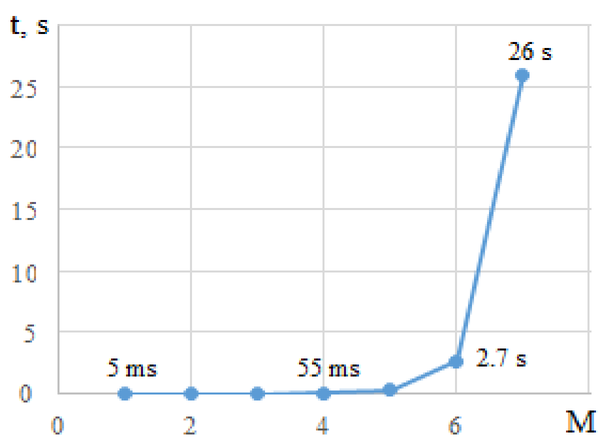

Figure 6. Dependency of search time t on the number of device parameters M

Conclusions

Thus, conducted research shows that search time of analogs is the most sensitive on M indicator – the number of parameters that describe a device in the database. With the value of $M \geq 10$ processing request is unacceptable for use in industrial environments. Therefore, further research of authors will be directed to search for more optimized algorithms under conditions of incomplete information about devices in the database.

References

[1] A. Kupin, I. Muzyka, M. Romanov, S. Ruban, O. Savitskyi, and V. Kharlamenko, “Intelligent management of repairs industrial facilities with geographically distributed infrastructure on the basis of CPS”,. Cyber-Physical System: Achievements and Challenges, II scientific seminar, Lviv Polytechnic National University, 2016, pp. 23-29 (in Ukrainian).

[2] A. Kupin, I. Muzyka, “Analysis of potential opportunities and rationale of cyber-physical systems for mining and metallurgical complex”, Computer science, information technology, automation, no. 4, 2016, pp. 25-29.

[3] A.Kupin Neural identification of techno-logical process of iron ore beneficiation // Proceedings of 4th IEEE Workshop on Intelligent Data Acquisition and Advanced Computing Systems Technology and Applications (IDAACS’2007).– *Dortmund, Germany*.–2007.– pp.225–227. J. Clerk Maxwell, *A Treatise on Electricity and Magnetism*, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp. 68–73.

[4] A.Kupin, Y.Kumchenko Usage of training methods to parameterization of multilayer neural computing structures for technological processes // Radioelectronic and Computer systems. - *Vol. 5(69).- Kharkiv «KhAI»*, 2014.- pp. 100-104.

[5] M. Hermann, T. Pentek, B. Otto (2015). Design Principles for Industrie 4.0 Scenarios: A Literature Review. Technische Universitat Dortmund.

[6] J. Lee, Industry 4.0 in Big Data Environment, Harting Tech News 26, 2013, http://www.harting.com/fileadmin/harting/documents/lg/hartingtechnologygroup/news/tec-news/tec-news26/EN_tecNews26.pdfA.

[7] E. A. Lee and S. A. Seshia, Introduction to Embedded Systems — A Cyber-Physical Systems Approach, http://LeeSeshia.org, 2011.

[8] N. E. Timoshevskaya, “Development and research of parallel combinatorial algorithms”, Applied Discrete Mathematics, vol. 2(4), 2009, pp. 96-103 (in Russian).

[9] N. E. Timoshevskaya, “Parallel enumeration of partition sets with the numeration method”, Journal of the Tomsk State University, no. 17, 2006, pp. 260-264 (in Russian).