

Программные системы: теория и приложения

Двуязычный электронный научный журнал

№2  2025

Bilingual Online Scientific Journal

Program Systems: Theory and Applications

Том 16 Выпуск 2(65) 2025 г.

СОДЕРЖАНИЕ

Научная статья

ТЕОРЕТИЧЕСКИЕ ОСНОВАНИЯ ПРОГРАММНЫХ СИСТЕМ

Джудакизаде М., Бельтюков А. П.[✉]. *Сложность вычислений с путешествиями во времени* (Англ., Рус.) 3–28, **29**–54

Научная статья

ПРИКЛАДНЫЕ ПРОГРАММНЫЕ СИСТЕМЫ

Абрамов Н. С.[✉], Фраленко В. П.. *Особенности организации единого научно-технологического пространства* **55**–76, 77–79

Научная статья

ПРОГРАММНОЕ И АППАРАТНОЕ ОБЕСПЕЧЕНИЕ РАСПРЕДЕЛЕННЫХ И СУПЕРКОМПЬЮТЕРНЫХ СИСТЕМ

Стецюра Г. Г.[✉]. *Скоростные ассоциативные операции и алгоритмы управления и обработки данных для распределенных цифровых систем, действующих в условиях неопределенности* ... **81**–106, 107–109

Научная статья

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И МАШИННОЕ ОБУЧЕНИЕ

Винокуров И. В.[✉]. *Повышение точности сегментирования объектов с использованием генеративно-состязательной сети* (Англ., Рус.) 111–131, **132**–152

Авторский указатель **153**

Click the flag at a top corner of any page to switch the language, please!

Author index 154

Contents 155



Complexity of Computations with Time Travel

Milad Joudakizadeh¹, Anatoly Petrovich Beltiukov²

^{1,2}Udmurt State University, Izhevsk, Russia.

²belt.udsu@mail.ru

Abstract. This paper explores a mathematical model of computation that can be interpreted as a computer capable of receiving data from future states of its own computational process. Under certain combinations of input data and programs, such computations may become infeasible due to emerging contradictions or lead to ambiguous outcomes. We investigate programs for which this process is always feasible and yields a unique result.

It is demonstrated that, in the absence of computational complexity constraints, such machines can output the value of any recursively decidable predicate within a fixed time after the computation begins—referred to as the response delay time—while the computation process must continue even after the result is produced. When the total runtime of these machines is bounded by a polynomial function of the input size, they precisely recognize languages belonging to the intersection of the NP and co-NP complexity classes, with the same constant response delay time in the aforementioned sense.

Possible practical implementations of such a computer are examined, including an analysis of operational protocols leveraging quantum annealing to select the appropriate computational process. It is shown that, with parallelization of the computational process, the class of problems solvable by these machines in polynomial time corresponds to the PSPACE complexity class.

Additionally, a mode of operation is studied in which these machines have direct access to input data. In this case, if the runtime is limited to a logarithmic function of the input size, the class of problems solvable by such a parallelized computer encompasses LOGSPACE

The findings of this study can be applied to develop new programming principles for nondeterministic computational machines, where data transmission from the future is employed instead of nondeterministic choices. (*Linked article texts in English and in Russian*).

Key words and phrases: Turing Machine, Computational Complexity, Recursive Sets, Nondeterministic Computation, Quantum Annealing, Time Machine, Artificial Intelligence

2020 *Mathematics Subject Classification:* 68Q09; 68Q10, 68Q15

For citation: Milad Joudakizadeh, Anatoly P. Beltiukov. *Complexity of Computations with Time Travel*. Program Systems: Theory and Applications, 2025, **16**:2(65), pp. 3–54. (*In English, in Russian*).

https://psta.psiras.ru/read/psta2025_2_3-54.pdf

Introduction

This article presents an unconventional perspective on computational processes. What happens if a machine is allowed to peek into its own future and use this information to make decisions in the present? This idea, which at first glance seems inherently paradoxical, lies at the heart of the computational model under investigation—Turing machines with a backward-in-time communication channel. This model not only redefines the understanding of time in computation but also raises fundamental questions about causality and the interaction between past and future in computational processes.

Our machine is defined in such a way that it can receive information from future states of its computational process. This model represents a theoretical framework that introduces new ways of organizing computations. It allows obtaining the result of a complex computation before its completion and using this information. As with any time travel scenario, complications arise: transmitting data from the future may lead to contradictions and ambiguities.

In this article, we focus on cases where such problems can be avoided and demonstrate how this model can be implemented and applied to solve complex problems. This approach offers an alternative computational paradigm that connects time, causality, and computation, enabling new problem-solving strategies.

One of the results is that such a machine can compute the value of any recursively enumerable predicate in constant time. This means that tasks requiring significant time in classical models can be solved almost instantaneously (although the computational process itself must continue, and its total duration may still be substantial).

We also prove that if the total runtime of such machines is bounded by polynomials in the input size, they can recognize all languages in the intersection of NP and co-NP. These results indicate that the proposed model is not only theoretically interesting but also possesses significant computational power.

We also consider a modification of the machine that enables parallel data processing. It is shown that in this case, the class of problems solvable by such machines in polynomial time corresponds to PSPACE (i.e., polynomial time on the new machines is equivalent to PSPACE on ordinary Turing machines).

Another modification involves adding a mechanism for direct access to input data by addressing cells of the input tape using a special address tape. In this case, with runtime limited logarithmically in the input size, the class of problems solvable by our parallel computer includes LOGSPACE (here, LOGTIME also translates into corresponding space complexity). It is worth noting that this mode of operation does not entail excessive complexity and is feasible on conventional computers for reasonable input sizes.

From a practical standpoint, our work introduces new programming mechanisms. Instead of complex nondeterministic choices, which can be difficult to comprehend, we propose using data from the future, making the computational process more intuitively understandable. We also explore possible implementations of this model, including deterministic backtracking and quantum annealing.

In the first approach, systematic backtracking ensures that all possible paths are explored to prevent paradoxes. In the second approach, quantum adiabatic machines simultaneously explore multiple states of the computational process, selecting a self-consistent sequence of actions.

Thus, we propose a model that expands our understanding of computational processes and provides new approaches to programming.

1. Related Work

Over the past few decades, the idea of utilizing temporal paradoxes and time-traveling data transmission has attracted attention in theoretical computer science and physics. Many researchers have explored the possibility of extending classical computational models by incorporating mechanisms that allow information to be retrieved from the future, which is hypothesized to redefine traditional views on computational resources and complexity.

In the seminal work by Cook [1], the concept of NP-completeness was introduced, laying the foundation for further research in computational complexity and the reducibility of various problems to a universal problem. Expanding on these ideas, R. M. Karp [2] analyzed the reducibility of combinatorial problems, demonstrating that many seemingly disparate problems are interconnected.

The classic work by Garey and Johnson [3] became a fundamental reference for researchers studying computational complexity, presenting

a wide range of NP-complete problems and serving as a starting point for analyzing the limitations of traditional computational models.

In recent years, significant attention has been devoted to models that leverage physical effects to solve complex computational problems. Aaronson [4] investigated the potential of quantum computing for solving traditionally hard problems, showing that quantum methods can provide substantial advantages over classical algorithms.

Deutsch [5] introduced the concept of a universal quantum computer, marking a major milestone in quantum computing and influencing subsequent research on hybrid computational models.

Regarding models that violate classical causality, substantial contributions have been made in studies related to closed timelike curves (CTCs). Brun [6] demonstrated that the use of CTCs can significantly accelerate problem-solving. Moreover, Aaronson and Watrous [7] established the equivalence of quantum and classical computations in the presence of CTCs, highlighting the potential of such models to expand computational capabilities.

Beyond classical approaches, the idea of integrating time-travel-based data transmission with complexity theory has sparked interest. The works of Papadimitriou [8] and Arora and Barak [9] provide a deep analysis of the structure of computational problems, including the relationship between NP and co-NP classes. These studies serve as a foundation for analyzing how nonstandard computational models that utilize future data transmission can solve problems that are difficult for traditional models.

Classical works on quantum computing, such as the book by Nielsen and Chuang [10], as well as the algorithms of Grover [11] and Shor [12], illustrate how quantum methods can significantly accelerate the solution of certain problem classes.

More recent studies, such as those by Harrow and Montanaro [13] and Iqbal et al. [14], confirm the practical significance of quantum algorithms for solving complex computational problems.

An important aspect of this field is the concept of time-travel-based data transmission, which allows machines to predict their future states and use this information to adjust current computations. In this context, the work by Orlicki [15] explores a model of time-traveling machines and examines the problem of self-consistent halting. This approach extends

traditional models, offering new perspectives on the interaction of temporal flows in computation.

Furthermore, Gödel's contribution [16] demonstrated the theoretical possibility of closed timelike curves in cosmological solutions, providing additional physical justification for computational models involving future data transmission.

Studies by Bacon [17] and Wolpert [18] indicate that physical constraints on future prediction and feedback between temporal states significantly influence computational processes.

Finally, experimental research, such as the work by King et al. [19], shows that quantum systems with thousands of qubits can already simulate complex physical phenomena, raising hopes for the practical feasibility of computational models that exploit time-travel effects.

Thus, these works illustrate that the idea of future data transmission and temporal loops in computation has strong theoretical and practical foundations. Our time-traveling computation model naturally fits into this context, offering a fresh perspective on computational complexity and enabling the solution of problems that are intractable for traditional models.

2. Structure and Features of the Turing Machine with Time Data Transmission Channel

This section describes the construction of a Turing machine equipped with a time data transmission channel. This model extends the classical Turing machine by including a mechanism that allows data to be transmitted to previously executed steps of the computational process and to receive data transmitted from future steps. Let us consider the structure and features of such a machine.

2.1. Main Components of the Machine

The Turing machine with a time data transmission channel is a multi-tape machine (with one head for each tape) and includes the following elements:

Input Tape: This tape contains the input word that needs to be processed.

At the beginning of the machine's operation, the input data is written to this tape.

Output Tape: The result of the machine's operation is recorded on this tape. For a recognition machine, the result can be represented as a binary value: 0 (no) or 1 (yes).

Multiple Working Tapes for performing necessary computations.

Communication Tape (Connection Tape): A special tape designed for transmitting data in time. The machine can write data to this tape for sending to the past and receive data from the future. To send and receive data, the machine must transition to special states for data transmission and reception, respectively.

Address Tape for Transmitted Data: This tape records labels that allow data to be retrieved by referencing these labels in requests. Special machine states are used for data transmission and retrieval.

In the transmission state, the contents of the communication tape along with the label are sent to the time data transmission channel. If data with such a label has already been sent earlier, it is considered that an error has occurred during the machine's operation.

In the reception state, the communication tape receives data from the transmission channel associated with the specified label. If such data was not found in the future, it is also considered that an error occurred during the machine's operation. For solving recognition tasks where errors may occur during operation, usage is prohibited.

Tape Alphabets: Each tape has its own alphabet of symbols that can be written on it.

Control Unit: At each step of operation, the control unit can be in one of the states of the state alphabet. The following special states are distinguished:

initial state (in which the machine begins the computational process),

final state (in which the machine's operation is completed),

data transmission state (to which transitioning sends data with the specified label to the data transmission channel),

data reception state (to which transitioning receives data from the data transmission channel with the specified label),

output state (to which transitioning sends the contents of the output tape to the outside world while the machine continues its operation), considering that the machine operates with a single write to the output tape, meaning that a repeated transition to the output state is an error in the machine's operation.

Program: A function (table) that maps the machine's state and the symbols visible to the heads on each tape to a new state, new symbols to be written on the tapes, and the directions of head shifts.

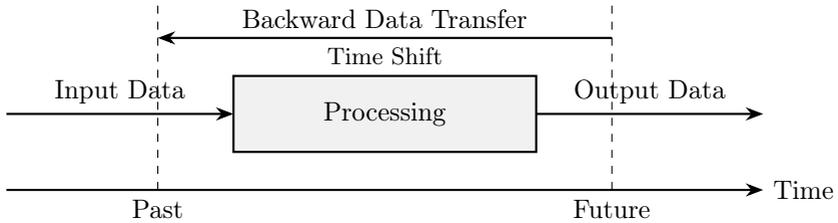


FIGURE 1. Schematic of reverse-time data transfer in a Turing machine with closed timelike curves

Computation Time: A sequence of steps. Each step performs one machine command or transmits/receives one symbol through the time data transmission channel (i.e., transmitting a long word to the data transmission channel takes as many steps as the length of that word). For a recognition machine, computation time must be finite.

In diagram Figure 1, the process of data processing with the possibility of transmitting information from the future to the past through a time shift is shown, implemented by closed time-like curves in Kerr metric.

2.2. Features of the Machine's Operation

The operation of the Turing machine with a time data transmission channel has several distinctive features:

Output of the result before the completion of computations: The machine can output the result to the output tape before the computation process is completed. This leads to a separation of two concepts of time:

response delay time,
total computation time.

Uniqueness and definiteness of computations: Due to the ability to obtain data from the future, the operation of the machine may become ambiguous or indeterminate. For example, ambiguity arises if the machine, having received some data from the future, writes the same data with the same label, which may lead to non-determinism in the process and, consequently, to different final results, depending on which data ends up being read and written. If the machine, having read data with one label, then necessarily writes different

data with the same label, a contradiction arises, making the process impossible, leading to indeterminacy. A machine is considered to recognize a certain property of the input data (i.e., computing the characteristic function of some set) if its operation is defined for all input data and the result is unique.

Programmability: At the end of the article, we will describe tools of algorithmic languages that will allow programming such actions not only for Turing machines but also for other computers.

2.3. Temporal Aspects of the Machine's Operation

A key feature of the machine is the ability to obtain data from the future, which leads to the following temporal effects:

Constant response delay time: Thanks to the mechanism of time data transmission, the machine can produce a simple result (YES or NO) in constant time, regardless of computational complexity.

Total computation time: Despite the constant response delay time, the total computation time may be significantly longer, as the computation process must continue even after the result is output.

2.4. Explanation of Ambiguity and Indeterminacy

To illustrate the ambiguity and indeterminacy in the operation of the machine, consider the following example. Suppose the machine queries one bit from the future and chooses a branch of computations based on its value. If the bit value of 1 is received, the machine follows the first branch; if a value of 0 is received, it follows the second branch. Ambiguity in the computation process arises if, in each branch, the machine writes the same bit with the same label to the communication tape. This creates two possible paths of computation. It is important to distinguish between the ambiguity of computation and the ambiguity of the result. In an ambiguous computation process, the result may turn out to be the same.

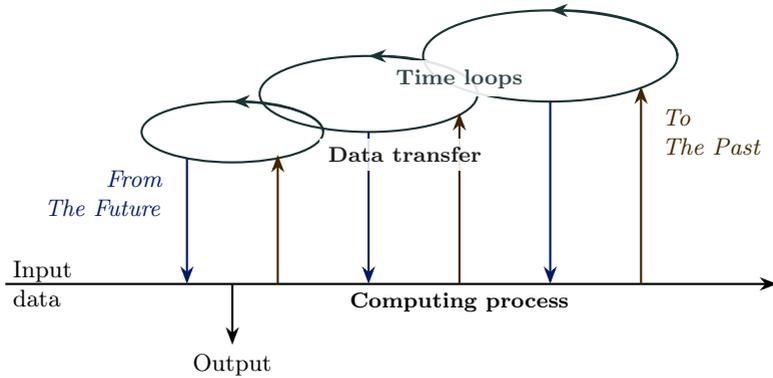


FIGURE 2. Schematic of temporal data flow in a Turing machine with closed timelike curves

If the machine reads one bit from the future and tries to write a contradictory bit to the communication tape with the same label, a contradiction arises, making such computation impossible.

Combinations of these situations may lead to some branches of the computation process, arising from internal ambiguity, subsequently becoming contradictory and thus being discarded. Only non-contradictory branches will remain. It may happen that all remaining non-contradictory branches lead to one final result. If this always occurs, we will consider such a machine correct for solving the recognition problem.

Thus, the described situations may lead to the following:

ambiguity, if several branches of computations end with different results,
indeterminacy, if no branch can complete without contradictions.

This demonstrates how the use of data from the future can affect the computational process.

3. Capabilities of Machines with Time Data Transmission

The scheme of data transmission in time for a Turing machine with closed time-like curves is presented in Figure 2.

It shows the main stages of information processing, the feedback mechanism through time loops, and the direction of data transmission.

In this section, we examine the computational capabilities of the Turing machine with a time data transmission channel. We present two theorems that demonstrate how the use of data from the future affects the computational complexity and classes of problems solvable by such a machine. These theorems show that the proposed model possesses properties that allow solving problems inaccessible to classical computational systems.

3.1. Recognition of Decidable Sets

THEOREM 1. *A Turing machine with a time data transmission channel recognizes all recursively decidable sets with constant response delay time.*

PROOF. The proof is based on the machine's ability to use data from the future to predict the results of computations. Let us recognize the membership of input strings in a given recursively decidable set. There exists a standard recognizing Turing machine for this set. From this machine, we will construct the required Turing machine with a time data transmission channel that will operate as follows.

At the start of its operation, the machine requests the contents of the communication tape labeled 0. This value is output as the result of its operation. Then, the process of the standard Turing machine recognizing membership in the given set is initiated. When the recognition is complete, its result is transmitted to the past with the label 0. This ensures that in the past, the machine provided the correct result.

Now, let us assume we have the original machine with a time data transmission channel that recognizes the membership of input strings in some set. By definition, the total running time of such a machine is always finite. We will construct a new standard Turing machine that recognizes membership in this set, which enumerates all possible strings that could serve as a partial protocol of the original machine's operation (chains of instantaneous descriptions), in order of their lengthening.

The machine checks for each string whether it is a correct completed protocol of the original machine's operation. As soon as such a protocol is found, the result of the original machine's operation is output as the result of the new machine. Since the operation of the original machine, by definition, is definite and unique, the new machine always provides the correct answer regarding the membership of the input string in the given set. \square

3.2. Recognition of Sets in $NP \cap co-NP$

THEOREM 2. *A Turing machine with constant response delay time and polynomial total computation time recognizes exactly all sets in $NP \cap co-NP$*

IDEA OF THE PROOF. By interpreting the reception of data from the future as branching into several paths, we construct a machine that operates in polynomial time and ensures the consistency of results across all branches of computations.

4. Proof of Theorem 2

In this section, we will present the proof of Theorem 2, formulated in the previous section. First, we will reformulate the operation of our machine in terms of nondeterministic computations. This will allow us to transform the operation of our machine into the operation of standard nondeterministic Turing machines and vice versa. Since the mechanism for transitioning to constant response delay has been described in the proof of the previous theorem, we will initially prove Theorem 2 without mentioning this, which we will do.

4.1. Transition to Nondeterministic Computations

The idea is to interpret the reading of each data symbol on the communication tape, received from the future by some label, as branching in the computation process. Each branch corresponds to one of the possible values of that symbol. When the machine reaches the moment in time denoted by this label, it checks whether the assumed read symbol matches the transmitted one. If the value of the symbol recorded in the communication tape cell matches the expected value in the current branch, the computations continue. Otherwise, the computation branch ends due to a contradiction.

A correct recognizing machine must provide the same answers (either 0 — NO or 1 — YES on all consistently terminating branches for a given input word. Next, we will transform our original machine with time data transmission into a nondeterministic one, as shown above, and then split it into two nondeterministic Turing machines, handling the aforementioned contradictions as follows:

- (1) The first machine accepts the recognized set, returning 1 when the original machine outputs 1, and returns 0 in all other cases (including contradictions).
- (2) The second machine accepts the complement of this set, returning 1 when the original machine outputs 0, and returns 0 in all other cases (including contradictions).

Thus, if the original machine with time data transmission outputs YES then the first of the resulting machines outputs YES on some branches of the computation, while the second always outputs NO. Conversely, if the original machine outputs NO then the first always outputs NO while the second outputs YES on some branches of the computation. Thus, the first of the constructed deterministic machines accepts the recognizable set itself, while the second accepts its complement. It is easy to see that the polynomial time bound is preserved in the specified transformation, as all performed restructurings give at most polynomial slowdown. So, if the original machine recognized the given set in polynomial time, the newly constructed machines operate in the same time. This means that the originally recognized set belongs to $NP \cap \text{co-NP}$.

4.2. Constructing a Machine with Time Data Transmission from Nondeterministic Machines

The reverse statement is also true: if there exist two nondeterministic machines, bounded by polynomial time, one of which accepts some set and the other accepts its complement, we can construct a single machine with time data transmission from the future that recognizes this set and operates in polynomial time. This machine operates as follows:

- (1) At the beginning of the process, it guesses which of the two nondeterministic machines to run.
- (2) It then executes the chosen machine.
- (3) If the first selected machine outputs 1, the computation branch ends with the result 1.
- (4) If the first selected machine outputs 0, the computation branch leads to a contradiction.
- (5) If the second selected machine outputs 1, the computation branch ends with the result 0.

- (6) If the second selected machine outputs 0, the computation branch leads to a contradiction.

By assumption, exactly one of these machines must yield a positive result. Thus, our machine always provides a definite answer: either 1 or 0.

4.3. Constant Response Delay Time

Constant response delay time is achieved because the decision about the result of the computations is made at the very beginning of the process. The machine requests data from the future and uses it to immediately output the answer. Meanwhile, the computation process must continue. \square

5. Implementation Questions

This section discusses possible approaches to the practical implementation of a Turing machine with a time data transmission channel. Although such a machine is a theoretical construct, it can be emulated using existing computational methods and technologies. The main approaches to implementation include deterministic backtracking, annealing methods, particularly quantum annealing. Each of these methods has its own features and limitations, which are discussed below.

5.1. Deterministic Backtracking

One of the simplest ways to implement a machine with data transmission from the future is to use deterministic backtracking on a conventional computer. In this approach, the computation process is modeled as a tree of possible states, where each branch corresponds to one of the possible values of the symbol received from the future.

Working Principle:

- (1) At each step of the computation, when reading a data symbol from the communication tape, the machine checks all possible values of that symbol.
- (2) If the chosen value leads to a contradiction, the system returns to the decision point and tries the next value.
- (3) The process continues until a consistent solution is found or all possible values are exhausted.

Advantages:

- Simplicity of implementation on existing computational platforms.
- Ability to use standard search algorithms with backtracking.

Disadvantages:

- High computational complexity, especially for problems with a large number of branches.
- Exponential growth of computation time as the depth of the decision tree increases.

5.2. Annealing Methods

Annealing methods represent a more complex approach to implementing a machine with time data transmission. In this case, the computation process is viewed as a whole structure (time unfolds in space). Then, a search is performed for the state of this structure with the lowest energy, where energy corresponds to the number of contradictions in the system.

Working Principle:

- (1) Initially, the system is in an excited state with a high energy level, corresponding to the presence of many contradictions.
- (2) Gradually, the system cools down, transitioning to states with lower energy.
- (3) If a state with the lowest energy (without contradictions) is reached during the annealing process, the process is considered successful.
- (4) If the lowest energy is achieved for only one answer (0 or 1), the result is considered final.

Advantages:

- Ability to find the global energy minimum even in complex systems with many local minima.
- Efficiency for problems where contradictions can be mitigated or eliminated during optimization.

Disadvantages:

- Requires significant resources to reach a state with minimal energy.
- Does not guarantee finding the optimal solution in polynomial time.

5.3. Quantum Annealing

Quantum annealing is a promising method for implementing a machine with data transmission from the future, based on the use of quantum devices. This method utilizes the effect of quantum tunneling to find states with the lowest energy.

Advantages:

- Potentially high speed of finding solutions due to quantum effects.
- Ability to solve problems that are difficult for classical computers.

Disadvantages:

- The speed and computational efficiency of quantum annealing are not yet fully understood.
- Existing quantum devices are limited in memory capacity and noise levels, complicating their use for complex tasks.

Prospects: Despite current limitations, quantum annealing remains a promising direction for experimentation. Existing quantum annealing machines, such as those from D-Wave (see, for example, [19]), are already suitable for research and can be used to explore the possibilities of implementing machines with data transmission from the future.

5.4. Summary

The implementation of a Turing machine with a time data transmission channel is possible using various approaches, including deterministic backtracking, annealing methods, and particularly quantum annealing. Each of these methods has its advantages and limitations, and the choice of a specific approach depends on the characteristics of the problem being solved and the available computational resources. Quantum annealing, in particular, remains a promising direction, although its practical efficiency requires further study.

6. Constructive Recognizers: An Example of Solving a Problem with a Time Data Transmission Machine

6.1. Non-constructiveness of Classical Recognizers

Classical recognition algorithms are usually non-constructive: they provide only a binary answer (YES or NO without explaining the result.

For example, when asked whether there exists a y that satisfies the condition $P(x, y)$, such algorithms merely confirm or deny the existence of such a y , without providing the actual object y upon a positive answer or proof of its absence upon a negative one. This limits their use in tasks where justification of the solution is required.

6.2. Constructive Problems in the Class $\text{NP} \cap \text{co-NP}$

A Turing machine with a time data transmission channel allows the implementation of constructive recognizers that provide not only an answer but also a certificate—proof or witness of the result. We consider symmetric problems in the class $\text{NP} \cap \text{co-NP}$, where the property $Q(x)$ is true if there exists a certificate y such that $Q_1(x, y)$, and false if there exists a certificate z such that $Q_0(x, z)$, with the sizes of y and z polynomially bounded relative to $|x|$.

6.3. Example: Checking for a Non-trivial Divisor

Consider the problem of determining whether the number x has a non-trivial divisor y such that $1 < y \leq d$ (the threshold d is given). This property $Q(d, x)$:

- If $Q(d, x) = 1$, then the certificate is the divisor y , verified by division in polynomial time.
- If $Q(d, x) = 0$, then the certificate is the factorization of x into prime factors, all $> d$, verified in polynomial time (for example, using the Miller test [20] under the assumption of the generalized Riemann hypothesis).

6.4. Implementation on a Time Data Transmission Machine

The constructive recognizer on a time data transmission machine works as follows:

(1) *Receiving Data from the Future:*

- In the first step, the machine requests a pair (b, C) with label 0, where $b \in \{0, 1\}$ is the answer, and C is the certificate (divisor or factorization).
- The answer b is immediately written to the output tape (if no fixed delay response is required, the certificate can also be issued, making the recognizer constructive).

(2) *Certificate Verification:*

- If $b = 1$, it checks that $C = y$ divides x and $1 < y \leq d$.
- If $b = 0$, it checks that C is a factorization of x into prime factors $> d$, and their product equals x .

(3) *Self-consistency:*

- After verification, if the certificate is correct, then (b, C) is written to the communication tape with label 0.
- If incorrect, a contradiction is created, and the computation branch is discarded.

It is easy to see that this machine operates in polynomial time.

Execution Example:

Input: $x = 15, d = 4$.

Step 1: Import $(1, 3)$ from the future, output 1.

Step 2: Check: $15 \div 3 = 5, 1 < 3 \leq 4$, correct.

Step 3: Transmit $(1, 3)$ with label 0, self-consistency achieved.

Input: $x = 17, d = 4$.

Step 1: Import $(0, \{17\})$, output 0.

Step 2: Check: 17 is prime (Miller test), $17 > 4$, correct.

Step 3: Transmit $(0, \{17\})$, self-consistency achieved.

6.5. Features and Advantages

Linearity: The process is deterministic.

Instant Response: The answer is available in the early steps due to data transmission from the future.

Transparency: If a certificate is issued, it makes the result interpretable, enhancing the clarity of the computations.

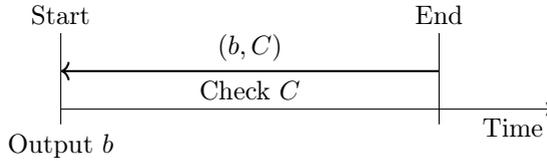


FIGURE 3. Scheme of the Constructive Recognizer's Operation

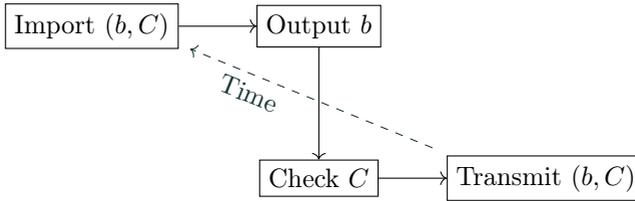


FIGURE 4. Flowchart of the Process

Figure 3 presents a scheme of the constructive recognizer's operation: data (b, C) is transmitted from the end of the computations to the beginning for immediate output and subsequent verification. The flowchart of the process from importing the result to verification and self-consistent data transmission is shown in Figure 4.

7. Actions for Time Data Transfer

To transform a conventional programming language into one that supports time data transfer, it is necessary to add two new operations: transferring data to the past and receiving data from the future. These actions allow the machine to interact with both future and past states of the computational process, providing new programming capabilities. In this section, we will examine these actions and their role in organizing computations with time data transfer.

7.1. Action of Transferring Data to the Past

The action of transferring data to the past allows the machine to record data at a specific moment in time, which can be used in previous computations. This action is formalized as follows:

Formal Description:

- $transfer(l, d)$: transfers object d to the moment in time l .
 - l — the label of the moment in time for which the data is being transferred.
 - d — the object being transferred (e.g., bit, number, or string).

Working Principle:

- (1) The machine records object d in the time data transfer channel with the timestamp l .
- (2) The data becomes available for use in previous steps of computations related to the moment in time l .
- (3) If a record with label l already exists, an error occurs.

Example: If the machine executes the action $transfer(1, 1)$, the value 1 is transferred with label 1. In previous computations related to 1, this value can be used for decision-making.

7.2. Action of Receiving Data from the Future

The action of receiving data from the future allows the machine to obtain data from a specific moment in the future, making it possible to use future information in current computations. This action is formalized as follows:

Formal Description:

- $accept(l)$: accepts data from the moment in time with label l .
 - l — the name of the moment in time from which data is requested.

Working Principle:

- (1) The machine requests data from the time data transfer channel with timestamp l .
- (2) If the data exists, it is returned and can be used in current computations.
- (3) If the data is absent, an error occurs.

Example: If the machine executes the action $accept(1)$, it receives the value d , which was transferred with label 1 using the action $transfer(1, d)$. This value can then be used for decision-making in current computations.

7.3. Interaction of Data Transfers

The actions of transferring and receiving data work together, ensuring a connection between current and future states of the computational process. This interaction allows the machine to obtain data from the future for decision-making in current computations. It is important to note that the volume of data transferred through time is a significant indicator of the complexity of the computation. For example, when implementing a backtracking process, the runtime slows down exponentially with a binary indicator of the volume of data in bits, as for each bit of data, it effectively requires entering two branches of computation. Similar complexities arise in annealing methods.

7.4. Summary

The actions of transferring and receiving data through time are key elements of our computational model. In principle, we can incorporate these capabilities into any programming language, transforming it into a language for controlling a machine with data transfer from the future to the past. These tools provide new opportunities for creating intuitive computational systems.

8. Parallelization of the Computational Process

This section describes a parallel machine with time data transfer—an extension of our machine with an additional capability aimed at increasing its computational power: executing another instance of the machine with specific input data. The input string for the new process is placed on a special argument tape. The parallel process is initiated when the machine transitions to a special state for starting a parallel process. The new process begins with a special initial state of the subprocess. Initially, the new process runs until it reaches a state where it outputs a result (while the old process waits). This result is then transferred from the output tape of the new process to the argument tape of the old process. Both processes then continue to execute in parallel, with the old process no longer interacting with the new process. The parallel execution of the new process is necessary for transferring data to the past within it, ensuring the correctness of the result produced by this process.

The new process can be started as many times as necessary, either by the main process or by spawned subprocesses. Since subprocesses can launch their own subprocesses in parallel and independently, the number of subprocesses can grow exponentially as the machine operates in parallel mode.

The total parallel runtime of the machine is considered to be the time until all parallel processes complete, as if they were executed synchronously, step by step.

THEOREM 3. *On a parallel machine with time data transfer, the operation of a conventional Turing machine with polynomial memory can be modeled in polynomial time (which corresponds to the complexity class PSPACE).*

PROOF. Consider a Turing machine with a known estimate of its capacity complexity in the class PSPACE. We will perform the following steps:

- (1) First, estimate the runtime of this machine (exponential in the volume of memory).
- (2) Launch a subprocess to model the operation of the machine for half of this time (the modeling time is passed to the subprocess along with the initial configuration of the process).
- (3) After obtaining the result and allowing the subprocess to continue forming it in the future, immediately launch a new subprocess to model the operation of the original machine for the remaining half of the allocated time, passing this subprocess the configuration obtained from the first subprocess.
- (4) Output the result.
- (5) The subprocesses continue working in a similar manner until the allocated time converges to a predetermined constant.

It is easy to see that the depth of subprocess calls here is polynomial, and the total parallel runtime is also polynomial. This completes the proof of the theorem. \square

THEOREM 4. *The operation of a parallel machine (with time data transfer) in polynomial time can be modeled by a conventional Turing machine with polynomial memory.*

PROOF. The reverse statement follows from the fact that the operation of such a parallel machine can be modeled on a conventional Turing machine using polynomial memory. This is achieved by sequentially executing all launched processes, as they do not interact with each other. Since the total parallel time of the modeled machine is polynomially bounded, the depth of subprocess calls is also polynomially bounded. Thus, the total memory volume for modeling also remains polynomial. This proves the theorem. \square

9. Direct Access to Input Data

In this section, we modify the model of our machine by adding the capability to work with large data through direct access to the cells of the input tape. By large data, we mean data that cannot be sequentially examined within the current time constraints. The direct access mechanism allows for quick access to remote cells, overcoming the specified limitation. This is achieved as follows: a special address tape is introduced, on which the binary representation of the cell number on the input tape is recorded. The input tape head instantly moves to the specified cell. No other mechanisms for moving this head are provided. We assume that all launched subprocesses work with the same input data. Thus, we have defined a parallel machine with time data transfer and direct access to input data.

If we replace *PSPACE* with *LOGSPACE* in the previous section, we obtain the following theorem.

THEOREM 5. *On a parallel machine with time data transfer and direct access to input data, the operation of a conventional Turing machine with logarithmic memory (*LOGSPACE*) can be modeled in logarithmic time (*LOGTIME*).*

PROOF. It is similar to the proof of Theorem 3, but with the replacement of polynomial constraints with logarithmic ones. Since access to input data occurs in constant time, the total execution time remains logarithmic. \square

The reverse statement is problematic since direct modeling of logarithmic time on our machine using a conventional Turing machine requires a memory volume polynomial in the logarithm of the size of the original data. However, a similar approach can be used to prove the following theorem.

THEOREM 6. *The class POLYLOGTIME on a parallel machine with time data transfer and direct access to input data coincides with the class POLYLOGSPACE for conventional Turing machines.*

PROOF. The proof is based on the fact that the operation of our machine with direct access can be modeled on a conventional Turing machine using polylogarithmic memory. Thus, the total modeling time remains polylogarithmic. \square

Conclusion

This work presents a fundamentally new computational model that expands classical representations of computational processes through the mechanism of transferring data across different moments in time. The main results of the research are as follows:

- (1) *Complexity Classes and Time Constraints:* It has been proven that a machine with data transfer from the future, without time constraints, recognizes exactly the decidable sets with a constant response delay. Under polynomial time constraints, the model corresponds to the class of problems in the intersection of *NP* and *co-NP*. These results demonstrate how controlled violations of causal order can assist computations.
- (2) *Parallelization and Space Complexity:* The extension of the model to allow the launching of subprocesses that inherit input data enables the coverage of the *PSPACE* class in polynomial time, demonstrating how time complexity transforms into space complexity.
- (3) *Direct Access and Data Processing Efficiency:* The modification with addressable access to the input tape and logarithmic time complexity ensures compliance with the *LOGSPACE* class, which is particularly relevant for large data processing tasks. This result is supported by the mechanism of instant movement of the input tape head, eliminating linear delays.
- (4) *Practical Implementation:* Two approaches to the practical implementation of the model are proposed:
 - *Deterministic Backtracking* — a systematic search for self-consistent computational trajectories;

- *Annealing Methods* (including quantum systems) — searching for states with minimal energy of contradictions.

Experiments with quantum annealers, such as D-Wave, could be the next step in assessing the physical realizability of the model.

The prospects for further research encompass both theoretical and practical aspects:

- Optimization of the volume of transmitted data to reduce computational load;
- Development of specialized programming languages with operators for time data transfer;
- Analysis of the connection with quantum computing.

This work is relevant not only for theoretical computer science but also for interdisciplinary research. It demonstrates that the connection of states in time may be the key to solving problems inaccessible to traditional systems. The proposed model could serve as a bridge between complexity theory and physical realizations, opening new possibilities for algorithms.

References

- [1] S. A. Cook. “The complexity of theorem-proving procedures”, *Logic, automata, and computational complexity: The works of Stephen A. Cook*, ed. Kapron B.C., ACM, New York, 2023, ISBN 979-8-4007-0779-7, pp. 143–152.  ^{↑5}
- [2] R. M. Karp. “Reducibility among combinatorial problems”, *50 Years of Integer Programming 1958–2008. From the Early Years to the State-of-the-Art*, eds. Jünger M., et al., Springer, Berlin–Heidelberg, 2010, ISBN 978-3-540-68274-5, pp. 219–241.  ^{↑5}
- [3] M. R. Garey, D. S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*, Series of Books in the Mathematical Sciences, Freeman, San Francisco, 1979, ISBN 978-0716710455, 340 pp. ^{↑5}
- [4] S. Aaronson. “Guest column: NP-complete problems and physical reality”, *ACM Sigact News*, **36**:1 (2005), pp. 30–52.  ^{↑6}
- [5] D. Deutsch. “Quantum theory, the Church–Turing principle and the universal quantum computer”, *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **400**:1818 (1985), pp. 97–117.  ^{↑6}
- [6] T. A. Brun. “Computers with closed timelike curves can solve hard problems efficiently”, *Foundations of Physics Letters*, **16**:3 (2003), pp. 245–253.  ^{↑6}

- [7] S. Aaronson, J. Watrous. “Closed timelike curves make quantum and classical computing equivalent”, *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **465**:2102 (2009), pp. 631–647. ⁶
- [8] Ch. Papadimitriou. *Computational Complexity*, Addison-Wesley, Reading, Massachusetts, 1994, ISBN 978-0201530827, 523 pp. ⁶
- [9] S. Arora, B. Barak. *Computational Complexity: A Modern Approach*, Cambridge University Press, New York, 2009, ISBN 978-0521424264, 594 pp. ⁶
- [10] M. A. Nielsen, I. L. Chuang. *Quantum computation and quantum information*, Cambridge University Press, New York, 2010, ISBN 9781107002173, 702 pp. ⁶
- [11] L. K. Grover. “A fast quantum mechanical algorithm for database search”, *STOC '96: Proceedings of the twenty-eighth annual ACM symposium on Theory of computing* (22–24 May 1996, Pennsylvania, USA), ACM, New York, 1996, ISBN 978-0-89791-785-8, pp. 212–219. ⁶
- [12] P. W. Shor. “Algorithms for quantum computation: discrete logarithms and factoring”, *Proceedings 35th Annual Symposium on Foundations of Computer Science* (20–22 November 1994, Santa Fe, NM, USA), IEEE, 1994, ISBN 0-8186-6580-7, pp. 124–134. ⁶
- [13] A. W. Harrow, A. Montanaro. “Quantum computational supremacy”, *Nature*, **549**:7671 (2017), pp. 203–209. ⁶
- [14] K. Iqbal, O. Ahmad, A. Y. Vandika. “Quantum computing and its implications for complex system analysis”, *Research of Scientia Naturalis*, **1**:5 (2024), pp. 238–247. ⁶
- [15] J. I. Orlicki. *Time-travelling Turing Machines and the self-consistent Halting Problem*, 2024. ⁶
- [16] K. Gödel. “An example of a new type of cosmological solutions of Einstein’s field equations of gravitation”, *Rev. Mod. Phys.*, **21**:3 (1949), pp. 447–450. ⁷
- [17] D. Bacon. “Quantum computational complexity in the presence of closed timelike curves”, *Phys. Rev. A*, **70**:3 (2004), id. 032309. ⁷
- [18] D. H. Wolpert. “Physical limits of inference”, *Physica D: Nonlinear Phenomena*, **237**:9 (2008), pp. 1257–1281. ⁷
- [19] A. D. King, J. Raymond, T. Lanting, R. Harris, A. Zucca, F. Altomare, A. J. Berkley, K. Boothby, S. Ejtemaee, C. Enderud, E. Hoskinson, Sh. Huang, E. Ladizinsky, A. J. R. MacDonald, G. Marsden, R. Molavi, T. Oh, G. Poulin-Lamarre, M. Reis, Ch. Rich, Y. Sato, N. Tsai, M. Volkman, J. D. Whittaker, J. Yao, A. W. Sandvik, M. H. Amin. “Quantum critical dynamics in a 5,000-qubit programmable spin glass”, *Nature*, **617**:7959 (2023), pp. 61–66. ^{7, 17}

- [20] G. L. Miller. “Riemann’s hypothesis and tests for primality”, *STOC ’75: Proceedings of the seventh annual ACM symposium on Theory of computing* (5–7 May 1975, Albuquerque, New Mexico, USA), ACM, New York, ISBN 978-1-4503-7419-4, pp. 234–239.  [↑18](#)

Received	15.03.2025;
approved after reviewing	02.04.2025;
accepted for publication	02.04.2025;
published online	21.04.2025.

Information about the authors:



Milad Joudakizadeh

PhD student in Artificial Intelligence and Machine Learning, Research focuses on computational complexity, machine learning, and logic in computer science.

 0000-0002-6167-6237

e-mail:



Anatoly Petrovich Beltiukov

Doctor of Physics and Mathematics, Professor. Research focuses on computational complexity, complexity classes, subrecursive hierarchies, intuitionistic mathematics, constructive systems, proof mechanization, proof complexity, computation models, substructural logics, weak arithmetics, and logic in computer science.

 0000-0002-3433-9067

e-mail:

The authors contributed equally to this article.

The authors declare no conflicts of interests.

УДК 004.415.2:519.681

 10.25209/2079-3316-2025-16-2-3-54

Сложность вычислений с путешествиями во времени

Милад Джудакизде¹, Анатолий Петрович Бельтюков^{2✉}

^{1,2}Удмуртский государственный университет, Ижевск, Россия

[✉]belt.udsu@mail.ru

Аннотация. Эта работа рассматривает математическую модель вычислений, которая может интерпретироваться как компьютер, способный получать данные из будущих состояний своего вычислительного процесса. При определённых сочетаниях входных данных и программ такие вычисления могут становиться невыполнимыми из-за возникающих противоречий или приводить к неоднозначным результатам. Мы исследуем программы, для которых этот процесс всегда выполним и даёт однозначный результат.

Показано, что при отсутствии ограничений на вычислительную сложность такие машины могут выдавать значение любого рекурсивно разрешимого предиката через фиксированное время после начала вычисления – время задержки ответа (при этом процесс вычисления должен продолжаться и после получения ответа). Если общее время работы таких машин ограничить полиномами от размера входных данных, то эти машины будут распознавать в точности языки, принадлежащие пересечению классов NP и co-NP, за постоянное время задержки ответа в таком же смысле.

Рассматриваются возможные реализации такого компьютера на практике, включая анализ возможных протоколов работы с использованием квантового отжига для выбора нужного процесса. Показано, что при распараллеливании вычислительного процесса класс задач, решаемых рассматриваемыми машинами за полиномиальное время, соответствует классу PSPACE.

Кроме того, исследуется режим работы, при котором эти машины имеют прямой доступ ко входным данным. В этом случае, если время работы ограничено логарифмом от размера входных данных, то класс задач, решаемых таким параллельно работающим компьютером, содержит LOGSPACE.

Результаты данного исследования могут быть использованы для разработки новых принципов программирования недетерминированных вычислительных машин, в которых вместо недетерминированных выборов используется передача данных из будущего.

(Связанные тексты статьи на английском и на русском языках)

Ключевые слова и фразы: Машина Тьюринга, вычислительная сложность, рекурсивные множества, недетерминированные вычисления, квантовый отжиг, машина времени, искусственный интеллект

Для цитирования: Джудакизде М., Бельтюков А. П. *Сложность вычислений с путешествиями во времени* // Программные системы: теория и приложения. 2025. Т. 16. № 2(65). С. 3–54. (Англ.+русс.) https://psta.psisras.ru/read/psta2025_2_3-54.pdf

Введение

В данной статье мы рассматриваем необычный взгляд на вычислительные процессы. Что получается, если машине разрешать заглядывать в своё собственное будущее и использовать эту информацию для принятия решений в настоящем? Эта идея, которая на первый взгляд кажется внутренне противоречивой, лежит в основе исследуемой вычислительной модели — машины Тьюринга с каналом передачи данных против течения времени. Эта модель не только переосмысливает понимание времени в вычислениях, но и поднимает фундаментальные вопросы о причинности и взаимодействии между прошлым и будущим в вычислительных процессах.

Наша машина определена таким образом, что она может получать информацию из будущих состояний своего вычислительного процесса. Эта модель представляет собой теоретическую конструкцию, вводящую новые способы организации вычислений. Она позволяет получить результат сложного вычисления до его завершения и использовать эту информацию. Как и в любом путешествии во времени, здесь возникают сложности: передача данных из будущего может привести к противоречиям и неоднозначностям.

В данной статье мы сосредотачиваемся на случаях, когда таких проблем можно избежать, и демонстрируем, как эту модель можно реализовать и применить для решения сложных задач. Этот подход представляет собой альтернативную вычислительную парадигму, которая связывает время, причинность и вычисления, позволяя применять новые стратегии решения задач.

Одним из результатов является то, что такая машина может вычислить значение любого рекурсивно разрешимого предиката за постоянное время. Это означает, что задачи, требующие значительного времени в классических моделях, могут быть решены практически мгновенно (хотя сам вычислительный процесс должен продолжаться и общее его время может оказаться весьма значительным).

Мы доказываем также, что если общее время работы таких машин ограничивать полиномами от размера входных данных, то эти машины смогут распознавать все языки из пересечения классов NP и co-NP. Эти результаты говорят о том, что предложенная модель не только теоретически интересна, но и обладает значительной вычислительной мощностью.

Мы также рассматриваем модификацию машины, которая позволяет осуществлять параллельную обработку данных. Показано, что в этом случае класс задач, решаемых такими машинами за полиномиальное

время, соответствует PSPACE (то есть полиномиальное время на новых машинах эквивалентно PSPACE на обычных машинах Тьюринга).

Другая модификация предполагает добавление механизма прямого доступа ко входным данным через адресацию ячеек входной ленты с помощью специальной адресной ленты. В этом случае, при ограничении времени работы логарифмом от размера входных данных, класс задач, решаемых нашим распараллеленным компьютером, включает LOGSPACE (здесь LOGTIME также преобразуется в соответствующую емкостную сложность). Стоит отметить, что такой режим работы не влечёт чрезмерно высокой сложности и является реализуемым на обычных компьютерах при разумных размерах входных данных.

С практической точки зрения, наша работа вводит новые механизмы программирования. Вместо сложных недетерминированных выборов, которые могут быть трудны для понимания, мы предлагаем использовать данные из будущего, что делает процесс вычислений интуитивно более понятным. Мы также исследуем возможные реализации этой модели, включая детерминированный поиск с возвратом и квантовый отжиг.

В первом подходе систематический поиск с возвратами назад гарантирует, что все возможные пути будут исследованы для предотвращения парадоксов. Во втором подходе квантовые адиабатические машины одновременно исследуют множество состояний вычислительного процесса, выбирая самосогласованную последовательность действий.

Таким образом, мы предлагаем модель, которая расширяет наше понимание вычислительных процессов и предоставляет новые подходы к программированию.

1. Связанные работы

За последние несколько десятилетий идея использования временных парадоксов и передачи данных во времени привлекла внимание в теоретической информатике и физике. Многие исследователи изучали возможность расширения классических вычислительных моделей за счёт включения механизмов, позволяющих получать информацию из будущего, что, как предполагается, может переопределить традиционные взгляды на вычислительные ресурсы и сложность.

В основополагающей работе Кука [1] было введено понятие NP-полноты, заложившее основу для дальнейших исследований в области вычислительной сложности и сводимости различных задач к универсальной проблеме. Развивая эти идеи, Р. М. Карп [2] проанализировал сводимость комбинаторных задач, показав, что многие на первый взгляд различные задачи взаимосвязаны.

Классическая работа Гэри и Джонсона [3] стала фундаментальным справочником для исследователей, изучающих вычислительную сложность, представив широкий спектр NP-полных задач и послужив отправной точкой для анализа ограничений традиционных вычислительных моделей.

В последние годы значительное внимание уделяется моделям, использующим физические эффекты для решения сложных вычислительных задач. Ааронсон [4] исследовал потенциал квантовых вычислений для решения традиционно сложных задач, продемонстрировав, что квантовые методы могут предоставить существенные преимущества перед классическими алгоритмами.

Дойч [5] ввёл концепцию универсального квантового компьютера, что стало важной вехой в квантовых вычислениях и повлияло на дальнейшие исследования гибридных вычислительных моделей.

Что касается моделей, нарушающих классическую причинность, существенный вклад был сделан в исследованиях, связанных с замкнутыми времениподобными кривыми (ЗВК). Брун [6] показал, что использование ЗВК может значительно ускорить решение задач. Кроме того, Ааронсон и Уотрус [7] установили эквивалентность квантовых и классических вычислений в присутствии ЗВК, подчеркнув потенциал таких моделей для расширения вычислительных возможностей.

Помимо классических подходов, идея интеграции передачи данных на основе путешествий во времени с концепциями теории сложности вызвала интерес. Работы Пападимитриу [8] и Ароры и Барака [9] предоставляют глубокий анализ структуры вычислительных задач, включая связь между классами NP и co-NP. Эти исследования служат основой для анализа того, как нестандартные вычислительные модели, использующие передачу данных из будущего, могут решать задачи, с которыми традиционные модели справляются с трудом.

Классические работы по квантовым вычислениям, такие как книга Нильсена и Чанга [10], а также алгоритмы Гровера [11] и Шора [12], иллюстрируют, как квантовые методы могут значительно ускорить решение определённых классов задач.

Более поздние исследования, такие как работы Харроу и Монтанаро [13] и исследования Икбала и др. [14], подтверждают практическую значимость квантовых алгоритмов для решения сложных вычислительных задач.

Важным аспектом этой области является концепция передачи данных на основе путешествий во времени, позволяющая машинам предсказывать свои будущие состояния и использовать эту информацию для корректировки текущих вычислений. В этом контексте работа Орлицки [15] исследует

модель машин с путешествиями во времени и рассматривает проблему самосогласованной остановки. Этот подход расширяет традиционные модели, предлагая новые взгляды на взаимодействие временных потоков в вычислениях.

Кроме того, вклад Гёделя [16] продемонстрировал теоретическую возможность замкнутых времениподобных кривых в космологических решениях, предоставив дополнительное физическое обоснование для вычислительных моделей, включающих передачу данных из будущего.

Исследования Бэкона [17] и Вольперта [18] указывают на то, что физические ограничения на предсказание будущего и обратную связь между временными состояниями значительно влияют на вычислительные процессы.

Наконец, экспериментальные исследования, такие как работа Кинга и др. [19], показывают, что квантовые системы с тысячами кубитов уже могут моделировать сложные физические явления, что вселяет надежды на практическую реализуемость вычислительных моделей, использующих эффекты путешествий во времени.

Таким образом, эти работы иллюстрируют, что идея передачи данных из будущего и эффектов временных петель в вычислениях имеет прочные теоретические и практические основания. Наша модель вычислений с путешествиями во времени естественным образом вписывается в этот контекст, предлагая новый взгляд на вычислительную сложность и позволяя решать задачи, недоступные для традиционных моделей.

2. Структура и особенности машины Тьюринга с каналом передачи данных во времени

В данном разделе описывается конструкция машины Тьюринга, оснащённой каналом передачи данных во времени. Эта модель расширяет классическую машину Тьюринга за счёт включения механизма, позволяющего передавать данные на ранее выполненные шаги вычислительного процесса и получать данные, переданные из будущих шагов. Рассмотрим структуру и особенности такой машины.

2.1. Основные компоненты машины

Машина Тьюринга с каналом передачи данных во времени является многоленточной машиной (с одной головкой на каждую ленту) и включает следующие элементы:

Входная лента: На этой ленте содержится входное слово, которое необходимо обработать. В начале работы машины входные данные записываются на эту ленту.

Выходная лента: На этой ленте записывается результат работы машины. Для распознающей машины результат может быть представлен в виде бинарного значения: 0 (нет) или 1 (да).

Несколько рабочих лент для выполнения необходимых вычислений.

Коммуникационная лента (лента связи): Специальная лента, предназначенная для передачи данных во времени. Машина может записывать данные на эту ленту для отправки в прошлое и получать данные из будущего. Для отправки и получения данных машина должна перейти в специальные состояния передачи и приёма данных соответственно.

Адресная лента для передаваемых данных: На этой ленте записываются метки, которые позволяют извлекать данные, ссылаясь на эти метки в запросах. Для передачи и извлечения данных используются специальные состояния машины.

В состоянии передачи содержимое коммуникационной ленты вместе с меткой отправляется в канал передачи данных во времени. Если данные с такой меткой уже были ранее отправлены, то считаем, что при работе машины возникла ошибка.

В состоянии приёма коммуникационная лента получает данные из канала передачи, связанные с заданной меткой. Если таких данных в будущем не оказалось, то также считаем, что при работе машины возникла ошибка. Для решения задач распознавания машины, при работе которых могут возникать ошибки, использовать запретим.

Алфавиты лент: Каждая лента имеет свой алфавит символов, которые могут быть записаны на неё.

Управляющее устройство: На каждом шаге работы управляющее устройство может находиться в одном из состояний алфавита состояний. Выделяются следующие специальные состояния:

начальное состояние, в котором машина начинает вычислительный процесс;

заключительное состояние, в котором работа машины завершается;

состояние передачи данных, при переходе в которое данные передаются с указанной меткой в канал передачи данных;

состояние приёма данных, при переходе в которое данные принимаются из канала передачи данных с указанной меткой;

состояние вывода, при переходе в которое содержимое выходной ленты отправляется во внешний мир, пока машина продолжает свою работу; считаем, что машина работает с однократной записью на выходную ленту, то есть повторный переход в состояние вывода — ошибка в работе машины.

Программа: Функция (таблица), которая отображает состояние машины и символы, видимые головками на каждой ленте, в новое состояние, новые символы, записываемые на ленты, и направления сдвигов головок.

Время вычислений: Последовательность шагов. Каждый шаг выполняет одну команду машины или передаёт/принимает один символ через канал передачи данных во времени (то есть на передачу длинного слова в канал передачи данных тратится столько шагов, какова длина этого слова). Для распознающей машины время вычислений должно быть конечным.

На диаграмме рисунка 1 показан процесс обработки данных с возможностью передачи информации из будущего в прошлое через временной сдвиг, реализуемый замкнутыми времениподобными кривыми в метрике Керра.



РИСУНОК 1. Схема работы машины Тьюринга с обратной временной связью

2.2. Особенности работы машины

Работа машины Тьюринга с каналом передачи данных во времени имеет несколько отличительных особенностей:

Вывод результата до завершения вычислений: Машина может вывести результат на выходную ленту до завершения процесса вычислений. Это приводит к разделению двух понятий времени:

время задержки ответа,
общее время вычислений.

Однозначность и определённость вычислений: Из-за возможности получения данных из будущего работа машины может стать неоднозначной или неопределённой. Например, неоднозначность возникает, если машина, получив некоторые данные из будущего, записывает те же данные с той же меткой, что может привести к недетерминированности процесса и, как следствие, к различным конечным результатам, в зависимости от того, какие данные окажутся прочитанными и записанными. Если машина, прочитав данные с одной меткой, затем обязательно записывает отличные от прочитанных данные с той же меткой, то возникает противоречие, делающее процесс невозможным, что приводит к неопределённости. Машина считается

распознающей некоторое свойство входных данных (т.е. вычисляющей характеристическую функцию некоторого множества), если её работа определена для всех входных данных и результат однозначен.

Программируемость: В конце статьи мы опишем инструменты алгоритмических языков, которые позволят программировать подобные действия не только для машин Тьюринга, но и для других компьютеров.

2.3. Временные аспекты работы машины

Ключевой особенностью машины является возможность получения данных из будущего, что приводит к следующим временным эффектам:

Постоянное время задержки ответа: Благодаря механизму передачи данных во времени машина может вывести простой результат (ДА или НЕТ) за постоянное время, независимо от вычислительной сложности.

Общее время вычислений: Несмотря на постоянное время задержки ответа, общее время вычислений может быть значительно больше, так как вычислительный процесс должен продолжаться даже после вывода результата.

2.4. Объяснение неоднозначности и неопределённости

Чтобы проиллюстрировать неоднозначность и неопределённость в работе машины, рассмотрим следующий пример. Предположим, что машина запрашивает один бит из будущего и выбирает ветвь вычислений на основе его значения. Если получено значение бита 1, то машина следует по первой ветви; если получено значение 0, то она следует по второй ветви. Неоднозначность процесса вычисления возникает, если в каждой ветви машина записывает тот же бит с той же меткой на коммуникационную ленту. Это создаёт два возможных пути вычисления. Следует отличать неоднозначность вычисления от неоднозначности результата. При неоднозначном процессе вычисления результат может оказаться одним и тем же.

Если машина считывает один бит из будущего и пытается записать противоречащий ему бит на коммуникационную ленту с той же меткой, то возникает противоречие, приводящее к невозможности такого вычисления.

Комбинации этих ситуаций могут приводить к тому, что некоторые ветви процесса вычисления, возникшие из-за внутренней неоднозначности, затем могут оказаться противоречивыми и быть по этой причине отсеяны. Тогда останутся только непротиворечивые ветви. Может оказаться так, что все оставшиеся непротиворечивые ветви приводят к одному окончательному результату. Если всегда происходит так, то будем считать такую машину корректной для решения задачи распознавания.

Таким образом, описанные ситуации могут привести к следующему:

- к неоднозначности*, если несколько ветвей вычислений завершаются разными результатами,
- к неопределённости*, если ни одна ветвь не может завершиться без противоречий.

Это демонстрирует, как использование данных из будущего может повлиять на вычислительный процесс.

3. Возможности машин с передачей данных во времени

Схема передачи данных во времени в машине Тьюринга с замкнутыми времениподобными кривыми представлена на рисунке 2. Показаны основные этапы обработки информации, механизм обратной связи через временные петли и направление передачи данных.

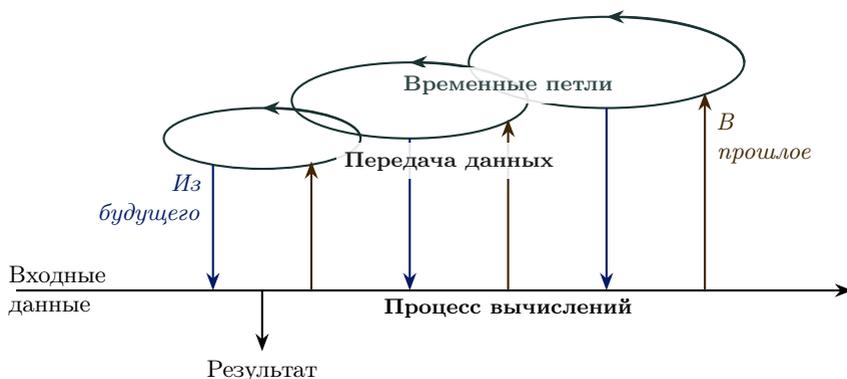


РИСУНОК 2. Схема временной передачи данных в машине Тьюринга с замкнутыми времениподобными кривыми

В данном разделе рассматриваются вычислительные возможности машины Тьюринга с каналом передачи данных во времени. Мы представим две теоремы, которые демонстрируют, как использование данных из будущего влияет на вычислительную сложность и классы задач, решаемых такой машиной. Эти теоремы показывают, что предложенная модель обладает свойствами, позволяющими решать задачи, недоступные для классических вычислительных систем.

3.1. Распознавание разрешимых множеств

ТЕОРЕМА 1. Машина Тьюринга с каналом передачи данных во времени распознаёт с постоянным временем задержки ответа в точности все рекурсивно разрешимые множества.

ДОКАЗАТЕЛЬСТВО. Доказательство основывается на способности машины использовать данные из будущего для предсказания результатов вычислений. Пусть нам надо распознать принадлежность входных слов к заданному рекурсивно разрешимому множеству. Для этого множества имеется обычная распознающая машина Тьюринга. Из этой машины построим требуемую машину Тьюринга с каналом передачи данных во времени, которая будет работать следующим образом.

В начале работы машина запрашивает содержимое коммуникационной ленты с меткой 0. Это значение и выдаётся на выход в качестве результата работы. Затем запускается процесс работы обычной машины Тьюринга, распознающей принадлежность заданному множеству. Когда распознавание закончится, его результат передаётся в прошлое с меткой 0. Таким образом обеспечивается, что в прошлом машина выдала верный результат.

Теперь пусть мы имеем исходную машину с каналом передачи данных по времени, распознающую принадлежность входных слов некоторому множеству. По определению, общее время работы такой машины всегда конечно. Построим новую обычную машину Тьюринга, распознающую принадлежность этому множеству, которая перебирает все возможные слова, претендующие на роль частичного протокола работы исходной машины (цепочки мгновенных описаний), в порядке их удлинения.

Машина проверяет для каждого слова, является ли оно правильным завершённым протоколом работы исходной машины. Как только такой протокол найден, результат работы исходной машины, выдаётся в качестве результата новой машины. Поскольку процесс работы исходной машины, по определению, определён и однозначен, новая машина всегда выдаёт правильный ответ о принадлежности входного слова заданному множеству. \square

3.2. Распознавание множеств в $NP \cap co-NP$

ТЕОРЕМА 2. Машина Тьюринга с постоянным временем задержки ответа и полиномиальным общим временем вычислений распознаёт в точности все множества в $NP \cap co-NP$.

ИДЕЯ ДОКАЗАТЕЛЬСТВА. интерпретируя приём данных из будущего как ветвление на несколько путей, мы строим машину, которая работает за полиномиальное время и обеспечивает согласованность результатов на всех ветвях вычислений.

4. Доказательство теоремы 2

В данном разделе мы изложим доказательство теоремы 2, сформулированной в предыдущем разделе. Сначала мы переформулируем работу нашей машины в терминах недетерминированных вычислений. Это позволит нам преобразовать работу нашей машины в работу обычных недетерминированных машин Тьюринга и наоборот. Поскольку механизм перехода к постоянному времени задержки описан при доказательстве предыдущей теоремы, то для начала достаточно доказать теорему 2 без упоминания об этом, что мы и сделаем.

4.1. Переход к недетерминированным вычислениям

Идея заключается в том, чтобы интерпретировать чтение каждого символа данных на коммуникационной ленте, полученного из будущего по некоторой метке, как ветвление процесса вычислений. Каждая ветвь соответствует одному из возможных значений этого символа. Когда машина достигает момента времени, обозначенного данной меткой, она проверяет, совпал ли предполагаемый считанный символ с переданным. Если значение символа, записанного в ячейке коммуникационной ленты, совпадает с ожидаемым значением в текущей ветви, вычисления продолжаются. В противном случае ветвь вычислений завершается из-за возникшего противоречия.

Корректная распознающая машина должна давать одинаковые ответы (либо 0 — НЕТ, либо 1 — ДА) на всех непротиворечиво завершающихся ветвях для данного входного слова. Далее мы преобразуем нашу исходную машину с передачей данных во времени в недетерминированную, как показано выше, а затем разделим её на две недетерминированные машины Тьюринга, обработав, упомянутые выше противоречия следующим образом.

- (1) Первая машина принимает распознаваемое множество, возвращая 1, когда исходная машина выводит 1, а во всех остальных случаях (включая противоречия) возвращает 0.
- (2) Вторая машина принимает дополнение этого множества, возвращая 1, когда исходная машина выводит 0, а во всех остальных случаях (включая противоречия) возвращает 0.

Таким образом, если исходная машина с передачей данных во времени выдаёт да, то первая из полученных машин на некоторых ветвях вычисления выдаёт ДА, а вторая — всегда НЕТ. Наоборот, если исходная машина выдаёт НЕТ, то первая всегда выдаёт НЕТ, а вторая — на некоторых ветвях вычисления выдаёт ДА. Таким образом, первая из построенных детерминированных принимает само распознаваемое множество, а вторая — его дополнение. Нетрудно видеть, что полиномиальное ограничение времени при указанном преобразовании сохраняется, так как все проведённые перестройки дают не более, чем полиномиальное замедление. Так что, если исходная машина распознавала данное множество за полиномиальное время, то такое же время работают и вновь построенные машины. Это означает, что исходно распознаваемое множество принадлежит $NP \cap co-NP$.

4.2. Построение машины с передачей данных во времени из недетерминированных машин

Обратное утверждение также верно: если существуют две недетерминированные машины, ограниченные полиномиальным временем, одна из которых принимает некоторое множество, а другая принимает его дополнение, мы можем из них построить одну машину с передачей данных из будущего, распознающую это множество и работающую полиномиальное время. Эта машина работает следующим образом:

- (1) В начале процесса она угадывает, какую из двух недетерминированных машин запустить.
- (2) Затем она выполняет выбранную машину.
- (3) Если первая выбранная машина выводит 1, ветвь вычислений завершается с результатом 1.
- (4) Если первая выбранная машина выводит 0, ветвь вычислений приводится к противоречию.
- (5) Если вторая выбранная машина выводит 1, ветвь вычислений завершается с результатом 0.

- (6) Если вторая выбранная машина выводит 0, ветвь вычислений приводится к противоречию.

По условию, ровно одна из этих машин должна дать положительный результат. Таким образом, наша машина всегда даёт определённый ответ: либо 1, либо 0.

4.3. Постоянное время задержки ответа

Постоянное время задержки ответа достигается благодаря тому, что решение о результате вычислений принимается в самом начале процесса. Машина запрашивает данные из будущего и использует их для немедленного вывода ответа. При этом процесс вычислений должен продолжаться. \square

5. Вопросы возможной реализации

В данном разделе обсуждаются возможные подходы к практической реализации машины Тьюринга с каналом передачи данных во времени. Хотя такая машина является теоретической конструкцией, её можно эмулировать с использованием существующих вычислительных методов и технологий. Основные подходы к реализации включают детерминированный поиск с возвратом, методы отжига, в частности, квантовый отжиг. Каждый из этих методов имеет свои особенности и ограничения, которые обсуждаются ниже.

5.1. Детерминированный поиск с возвратом

Один из самых простых способов реализации машины с передачей данных из будущего заключается в использовании детерминированного поиска с возвратом на обычном компьютере. В этом подходе процесс вычислений моделируется как дерево возможных состояний, где каждая ветвь соответствует одному из возможных значений символа, полученного из будущего.

Принцип работы:

- (1) На каждом шаге вычислений, при чтении символа данных с коммуникационной ленты, машина проверяет все возможные значения этого символа.
- (2) Если выбранное значение приводит к противоречию, система возвращается к точке принятия решения и пробует следующее значение.
- (3) Процесс продолжается до тех пор, пока не будет найдено согласованное решение или не будут исчерпаны все возможные значения.

Преимущества:

- Простота реализации на существующих вычислительных платформах.
- Возможность использования стандартных алгоритмов поиска с возвратом.

Недостатки:

- Высокая вычислительная сложность, особенно для задач с большим количеством ветвлений.
- Экспоненциальный рост времени вычислений с увеличением глубины дерева решений.

5.2. Методы отжига

Методы отжига представляют собой более сложный подход к реализации машины с передачей данных во времени. В этом случае процесс вычислений рассматривается как целостная структура (время разворачивается в пространстве). Затем выполняется поиск состояния этой структуры с наименьшей энергией, где энергия соответствует количеству противоречий в системе.

Принцип работы:

- (1) Изначально система находится в возбуждённом состоянии с высоким уровнем энергии, что соответствует наличию множества противоречий.
- (2) Постепенно система охлаждается, переходя в состояния с меньшей энергией.
- (3) Если в процессе отжига достигается состояние с наименьшей энергией (без противоречий), процесс считается успешным.
- (4) Если наименьшая энергия достигается только для одного ответа (0 или 1), результат считается окончательным.

Преимущества:

- Возможность нахождения глобального минимума энергии даже в сложных системах с множеством локальных минимумов.
- Эффективность для задач, где противоречия могут быть смягчены или устранены в процессе оптимизации.

Недостатки:

- Требует значительных ресурсов для достижения состояния с минимальной энергией.
- Не гарантирует нахождение оптимального решения за полиномиальное время.

5.3. Квантовый отжиг

Квантовый отжиг является перспективным методом реализации машины с передачей данных из будущего, основанным на использовании квантовых устройств. Этот метод использует эффект квантового туннелирования для нахождения состояний с наименьшей энергией.

Преимущества:

- Потенциально высокая скорость нахождения решений благодаря квантовым эффектам.
- Возможность решения задач, сложных для классических компьютеров.

Недостатки:

- Скорость и вычислительная эффективность квантового отжига ещё не до конца изучены.
- Существующие квантовые устройства ограничены по объёму памяти и уровню шума, что осложняет их использование для сложных задач.

Перспективы: Несмотря на текущие ограничения, квантовый отжиг остаётся перспективным направлением для экспериментов. Существующие машины для квантового отжига, такие как устройства компании D-Wave (см., например, [19]), уже подходят для исследований и могут быть использованы для изучения возможностей реализации машин с передачей данных из будущего.

5.4. Итог

Реализация машины Тьюринга с каналом передачи данных во времени возможна с использованием различных подходов, включая детерминированный поиск с возвратом, методы отжига и, в частности, квантовый отжиг. Каждый из этих методов имеет свои преимущества и ограничения, и выбор конкретного подхода зависит от характеристик решаемой задачи и доступных вычислительных ресурсов. Квантовый отжиг, в частности, остаётся перспективным направлением, хотя его практическая эффективность требует дальнейшего изучения.

6. Конструктивные распознаватели: пример решения задачи на машине с передачей данных во времени

6.1. Неконструктивность классических распознавателей

Классические алгоритмы распознавания обычно неконструктивны: они выдают только бинарный ответ (ДА или НЕТ) без объяснения результата. Например, на вопрос от том, существует ли y , удовлетворяющее условию $P(x, y)$, такие алгоритмы лишь подтверждают или опровергают существование такого y , не предоставляя сам объект y при положительном ответе или доказательство его отсутствия при отрицательном. Это ограничивает их использование в задачах, где требуется обоснование решения.

6.2. Конструктивные задачи в классе $NP \cap co-NP$

Машина Тьюринга с каналом передачи данных во времени позволяет реализовать конструктивные распознаватели, выдающие не только ответ, но и сертификат — доказательство или свидетельство результата. Мы рассматриваем симметричные задачи класса $NP \cap co-NP$, где свойство $Q(x)$ истинно, если существует сертификат y такой, что $Q_1(x, y)$, и ложно, если существует сертификат z такой, что $Q_0(x, z)$, причём размеры y и z полиномиально ограничены относительно $|x|$.

6.3. Пример: проверка наличия нетривиального делителя

Рассмотрим задачу определения, имеет ли число x нетривиальный делитель y такой, что $1 < y \leq d$ (порог d задан). Это свойство $Q(d, x)$:

- Если $Q(d, x) = 1$, то сертификат — делитель y , проверяемый делением за полиномиальное время.
- Если $Q(d, x) = 0$, то сертификат — факторизация x на простые множители, все $> d$, проверяемая за полиномиальное время (например, с использованием теста Миллера [20] при допущении обобщённой гипотезы Римана).

6.4. Реализация на машине с передачей данных во времени

Конструктивный распознаватель на машине с передачей данных во времени работает следующим образом:

(1) *Получение данных из будущего:*

- На первом шаге машина запрашивает с меткой 0 пару (b, C) , где $b \in \{0, 1\}$ — ответ, а C — сертификат (делитель или факторизация).
- Ответ b немедленно записывается на выходную ленту (если не требуется выдача ответа с фиксированной задержкой, то можно выдать и сертификат, тогда распознаватель получится конструктивным).

(2) *Проверка сертификата:*

- Если $b = 1$, то проверяется, что $C = y$ делит x и $1 < y \leq d$.
- Если $b = 0$, то проверяется, что C — факторизация x на простые множители $> d$, и их произведение равно x .

(3) *Самосогласование:*

- После проверки, если сертификат корректен, то (b, C) записывается на коммуникационную ленту с меткой 0.
- При некорректности создаётся противоречие, и ветвь вычислений отбрасывается.

Нетрудно видеть, что эта машина работает полиномиальное время.

Примеры выполнения:

Вход: $x = 15, d = 4$.

Шаг 1: Импорт $(1, 3)$ из будущего, вывод 1.

Шаг 2: Проверка: $15 \div 3 = 5, 1 < 3 \leq 4$, корректно.

Шаг 3: Передача $(1, 3)$ с меткой 0, самосогласование выполнено.

Вход: $x = 17, d = 4$.

Шаг 1: Импорт $(0, \{17\})$, вывод 0.

Шаг 2: Проверка: 17 — простое (тест Миллера), $17 > 4$, корректно.

Шаг 3: Передача $(0, \{17\})$, самосогласование выполнено.

6.5. Особенности и преимущества

Линейность: Процесс детерминирован.

Мгновенный отклик: Ответ доступен на первых шагах благодаря передаче из будущего.

Прозрачность: Если выдаётся сертификат, то это делает результат интерпретируемым, усиливая наглядность вычислений.

На рисунке 3 представлена схема работы конструктивного распознавателя: данные (b, C) передаются из конца вычислений в начало для немедленного вывода и последующей проверки. Блок-схема процесса от импорта результата до проверки и самосогласованной передачи данных дана на рисунке 4.

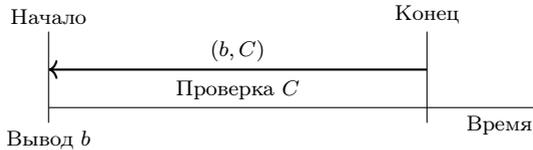


Рисунок 3. Схема работы конструктивного распознавателя

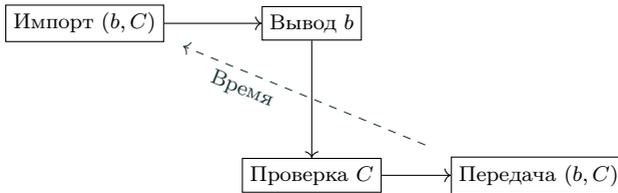


Рисунок 4. Блок-схема процесса

7. Действия для передачи данных во времени

Для преобразования обычного языка программирования в язык, поддерживающий передачу данных во времени, необходимо добавить две новые операции: передача данных в прошлое и приём данных из будущего. Эти действия позволяют машине взаимодействовать как с будущими, так и с прошлыми состояниями вычислительного процесса, предоставляя новые возможности для программирования. В данном разделе мы рассмотрим эти действия и их роль в организации вычислений с передачей данных во времени.

7.1. Действие передачи данных в прошлое

Действие передачи данных в прошлое позволяет машине записать данные в определённый момент времени, которые могут быть использованы в прошлых вычислениях. Это действие формализуется следующим образом:

Формальное описание:

- $transfer(l, d)$: передаёт объект d в момент времени l .
 - l — метка момента времени, для который передаются данные.
 - d — передаваемый объект (например, бит, число или строка).

Принцип работы:

- (1) Машина записывает объект d в канал передачи данных во времени с меткой времени l .
- (2) Данные становятся доступными для использования в прошлых шагах вычислений, связанных с моментом времени l .
- (3) Если запись с меткой l уже существует, возникает ошибка.

Пример: Если машина выполняет действие $transfer(1, 1)$, значение 1 передаётся с меткой 1. В прошлых вычислениях, связанных с 1, это значение может быть использовано для принятия решений.

7.2. Действие приёма данных из будущего

Действие приёма данных из будущего позволяет машине получить данные из определённого момента в будущем, что делает возможным использование будущей информации в текущих вычислениях. Это действие формализуется следующим образом:

Формальное описание:

- $accept(l)$: принимает данные из момента времени с меткой l .
 - l — имя момента времени, из которого запрашиваются данные.

Принцип работы:

- (1) Машина запрашивает данные из канала передачи данных во времени с меткой времени l .
- (2) Если данные существуют, они возвращаются и могут быть использованы в текущих вычислениях.
- (3) Если данные отсутствуют, возникает ошибка.

Пример: Если машина выполняет действие $accept(1)$, она получает значение d , которое было передано с меткой 1 с помощью действия $transfer(1, d)$. Это значение затем может быть использовано для принятия решений в текущих вычислениях.

7.3. Взаимодействие передач данных

Действия передачи и приёма данных работают совместно, обеспечивая связь между текущими и будущими состояниями вычислительного процесса. Это взаимодействие позволяет машине получать данные из будущего для принятия решений в текущих вычислениях. Важно отметить, что объём данных, передаваемых во времени, является важным показателем сложности вычисления. Например, при реализации процесса поиском с возвратом время работы замедляется экспоненциально с двоичным показателем объёма данных в битах, так как для каждого бита данных фактически приходится ввести две ветви вычислений. Подобные сложности возникают и в методах отжига.

7.4. Итог

Действия передачи и приёма данных во времени являются ключевыми элементами нашей вычислительной модели. В принципе, мы можем включить эти возможности в любой язык программирования, превратив его в язык управления машиной с передачей данных из будущего в прошлое. Эти инструменты предоставляют новые возможности для создания интуитивных вычислительных систем.

8. Распараллеливание вычислительного процесса

В данном разделе описывается параллельная машина с передачей данных во времени — расширение нашей машины с дополнительной возможностью, направленной на увеличение её вычислительной мощности: выполнение другого экземпляра машины с определёнными входными данными. Строка входных данных для нового процесса помещается на специальную ленту аргументов. Параллельный процесс запускается при переходе машины в специальное состояние запуска параллельного процесса. Новый процесс начинается с особого начального состояния подпроцесса. Изначально новый процесс выполняется до тех пор, пока не достигнет состояния, в котором выводит результат (в то время как старый процесс ожидает). Этот результат затем передаётся с выходной ленты нового процесса на ленту аргументов старого процесса. Затем оба процесса продолжают выполняться параллельно, при этом старый процесс больше не взаимодействует с новым процессом. Параллельное выполнение нового процесса необходимо для передачи данных в прошлое внутри него, что обеспечивает корректность результата, производимого этим процессом.

Новый процесс может быть запущен столько раз, сколько необходимо, основным процессом, либо порождёнными подпроцессами. Поскольку подпроцессы могут запускать свои собственные подпроцессы параллельно и независимо, количество подпроцессов может расти экспоненциально по мере работы машины в параллельном режиме.

Общее параллельное время работы машины считается временем до завершения всех параллельных процессов, как если бы они выполнялись синхронно, шаг за шагом.

ТЕОРЕМА 3. На параллельной машине с передачей данных во времени можно за полиномиальное время промоделировать работу обычной машины Тьюринга с полиномиальной памятью (что соответствует классу сложности PSPACE).

ДОКАЗАТЕЛЬСТВО. Рассмотрим машину Тьюринга с известной оценкой её емкостной сложности в классе PSPACE. Выполним следующее.

- (1) Сначала оценим время работы этой машины (экспоненциальное от объёма памяти).
- (2) Запустим подпроцесс для моделирования работы машины в течение половины этого времени (время моделирования передаётся подпроцессу вместе с начальной конфигурацией процесса).
- (3) Получив результат и позволив подпроцессу продолжить для его формирования в будущем, немедленно запустим новый подпроцесс для моделирования работы исходной машины в течение оставшейся второй половины выделенного времени, передав этому подпроцессу конфигурацию, полученную от первого подпроцесса.
- (4) Выведем результат.
- (5) Подпроцессы продолжают работу аналогичным образом, пока выделенное время не сойдётся к заранее определённой константе.

Нетрудно видеть, что глубина вызовов подпроцессов здесь полиномиальна, и общее параллельное время работы также полиномиально. \square

ТЕОРЕМА 4. Работа параллельной машины (с передачей данных во времени) за полиномиальное время может быть промоделирована обычной машиной Тьюринга с полиномиальным объёмом памяти.

ДОКАЗАТЕЛЬСТВО. Обратное утверждение следует из того, что работу такой параллельной машины можно промоделировать на обычной машине Тьюринга с использованием полиномиального объёма памяти. Это

достигается за счёт последовательного выполнения всех запущенных процессов, так как они не взаимодействуют друг с другом. Поскольку общее параллельное время моделируемой машины ограничено полиномиально, глубина вызовов подпроцессов также ограничена полиномиально. Таким образом, общий объём памяти для моделирования также остаётся полиномиальным. □

9. Прямой доступ к входным данным

В данном разделе мы модифицируем модель нашей машины, добавив возможность работы с большими данными через прямой доступ к ячейкам входной ленты. Под большими данными понимаются данные, которые невозможно последовательно просмотреть в рамках текущих временных ограничений. Механизм прямого доступа позволяет быстро обращаться к удалённым ячейкам, что позволяет преодолеть указанное ограничение. Это достигается следующим образом: вводится специальная адресная лента, на которой записывается двоичное представление номера ячейки на входной ленте. Головка входной ленты мгновенно перемещается к указанной ячейке. Другие механизмы перемещения этой головки не предусмотрены. Мы предполагаем, что все запущенные подпроцессы работают с одними и теми же входными данными. Таким образом, мы определили параллельную машину с передачей данных во времени и с прямым доступом ко входным данным.

Если заменить PSPACE на LOGSPACE в предыдущем разделе, мы получим следующую теорему.

ТЕОРЕМА 5. На параллельной машине с передачей данных во времени и с прямым доступом к входным данным работа обычной машины Тьюринга с логарифмическим объёмом памяти (LOGSPACE) может быть промоделирована за логарифмическое время (LOGTIME).

ДОКАЗАТЕЛЬСТВО. Рассуждение аналогично доказательству Теоремы 3, но с заменой полиномиальных ограничений на логарифмические. Поскольку доступ к входным данным происходит за постоянное время, общее время выполнения остаётся логарифмическим. □

Обратное утверждение проблематично, поскольку прямое моделирование логарифмического времени на нашей машине с использованием обычной машины Тьюринга требует объёма памяти, полиномиального от логарифма размера исходных данных. Однако аналогичный подход может быть использован для доказательства следующей теоремы.

ТЕОРЕМА 6. *Класс POLYLOGTIME на параллельной машине с передачей данных во времени и с прямым доступом к входным данным совпадает с классом POLYLOGSPACE для обычных машин Тьюринга.*

ДОКАЗАТЕЛЬСТВО. Утверждение основано на том, что работа нашей машины с прямым доступом может быть промоделирована на обычной машине Тьюринга с использованием полилогарифмической памяти. Таким образом, общее время моделирования остаётся полилогарифмическим. \square

Заключение

В данной работе предложена принципиально новая вычислительная модель, расширяющая классические представления о вычислительных процессах за счёт механизма передачи данных состояниями машины в различные моменты времени. Основные результаты исследования заключаются в следующем:

- (1) *Классы сложности и временные ограничения:* Доказано, что машина с передачей данных из будущего, без ограничений по времени, распознаёт в точности разрешимые множества с постоянной задержкой ответа. При полиномиальных ограничениях по времени модель соответствует классу задач из пересечения NP и co-NP. Эти результаты показывают, как контролируемые нарушения причинного порядка могут помогать вычислениям.
- (2) *Распараллеливание и емкостная сложность:* Расширение модели возможностью запуска подпроцессов, наследующих входные данные, позволяет охватить класс PSPACE за полиномиальное время, что демонстрирует, как временная сложность преобразуется в емкостную.
- (3) *Прямой доступ и эффективность обработки данных:* Модификация с адресуемым доступом ко входной ленте и логарифмической временной сложностью обеспечивает соответствие классу LOGSPACE, что особенно актуально для задач обработки больших данных. Этот результат поддерживается механизмом мгновенного перемещения головки входной ленты, устраняющим линейные задержки.
- (4) *Практическая реализация:* Предложены два подхода к практической реализации модели:
 - *Детерминированный поиск с возвратом* — систематический поиск самосогласованных вычислительных траекторий;

- *Методы отжига* (включая квантовые системы) — поиск состояний с минимальной энергией противоречий.

Эксперименты с квантовыми отжигателями, такими как D-Wave, могут стать следующим шагом для оценки физической реализуемости модели.

Перспективы дальнейших исследований охватывают как теоретические, так и практические аспекты:

- Оптимизация объёма передаваемых данных для снижения вычислительной нагрузки;
- Разработка специализированных языков программирования с операторами для передачи данных во времени;
- Анализ связи с квантовыми вычислениями.

Данная работа актуальна не только для теоретической информатики, но и для междисциплинарных исследований. Она демонстрирует, что связь состояний во времени может стать ключом к решению задач, недоступных для традиционных систем. Предложенная модель может служить мостом между теорией сложности и физическими реализациями, открывая новые возможности для алгоритмов.

Список использованных источников

- [1] Cook S. A. *The complexity of theorem-proving procedures // Logic, automata, and computational complexity: The works of Stephen A. Cook*, ed. Kapron B.C., New York: ACM.— 2023.— ISBN 979-8-4007-0779-7.— Pp. 143–152.  [↑31](#)
- [2] Karp R. M. *Reducibility among combinatorial problems // 50 Years of Integer Programming 1958–2008. From the Early Years to the State-of-the-Art*, eds. Jünger M., et al., Berlin–Heidelberg: Springer.— 2010.— ISBN 978-3-540-68274-5.— Pp. 219–241.  [↑31](#)
- [3] Garey M. R., Johnson D. S. *Computers and Intractability: A Guide to the Theory of NP-Completeness*, Series of Books in the Mathematical Sciences.— San Francisco: Freeman.— 1979.— ISBN 978-0716710455.— 340 pp. [↑32](#)
- [4] Aaronson S. *Guest column: NP-complete problems and physical reality // ACM Sigact News*.— 2005.— Vol. **36**.— No. 1.— Pp. 30–52.  [↑32](#)
- [5] Deutsch D. *Quantum theory, the Church–Turing principle and the universal quantum computer // Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*.— 1985.— Vol. **400**.— No. 1818.— Pp. 97–117.  [↑32](#)

- [6] Brun T. A. *Computers with closed timelike curves can solve hard problems efficiently* // Foundations of Physics Letters.– 2003.– Vol. **16**.– No. 3.– Pp. 245–253.  [↑₃₂](#)
- [7] Aaronson S., Watrous J. *Closed timelike curves make quantum and classical computing equivalent* // Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences.– 2009.– Vol. **465**.– No. 2102.– Pp. 631–647.  [↑₃₂](#)
- [8] Papadimitriou Ch. *Computational Complexity*.– Reading, Massachusetts: Addison-Wesley.– 1994.– ISBN 978-0201530827.– 523 pp. [↑₃₂](#)
- [9] Arora S., Barak B. *Computational Complexity: A Modern Approach*.– New York: Cambridge University Press.– 2009.– ISBN 978-0521424264.– 594 pp. [↑₃₂](#)
- [10] Nielsen M. A., Chuang I. L. *Quantum computation and quantum information*.– New York: Cambridge University Press.– 2010.– ISBN 9781107002173.– 702 pp.  [↑₃₂](#)
- [11] Grover L. K. *A fast quantum mechanical algorithm for database search* // STOC '96: Proceedings of the twenty-eighth annual ACM symposium on Theory of computing (22–24 May 1996, Pennsylvania, USA), New York: ACM.– 1996.– ISBN 978-0-89791-785-8.– Pp. 212–219.  [↑₃₂](#)
- [12] Shor P. W. *Algorithms for quantum computation: discrete logarithms and factoring* // Proceedings 35th Annual Symposium on Foundations of Computer Science (20–22 November 1994, Santa Fe, NM, USA).– IEEE.– 1994.– ISBN 0-8186-6580-7.– Pp. 124–134.  [↑₃₂](#)
- [13] Harrow A. W., Montanaro A. *Quantum computational supremacy* // Nature.– 2017.– Vol. **549**.– No. 7671.– Pp. 203–209.  [↑₃₂](#)
- [14] Iqbal K., Ahmad O., Vandika A. Y. *Quantum computing and its implications for complex system analysis* // Research of Scientia Naturalis.– 2024.– Vol. **1**.– No. 5.– Pp. 238–247.  [↑₃₂](#)
- [15] Orlicki J. I. *Time-travelling Turing Machines and the self-consistent Halting Problem*.– 2024.  [↑₃₂](#)
- [16] Gödel K. *An example of a new type of cosmological solutions of Einstein's field equations of gravitation* // Rev. Mod. Phys.– 1949.– Vol. **21**.– No. 3.– Pp. 447–450.  [↑₃₃](#)
- [17] Bacon D. *Quantum computational complexity in the presence of closed timelike curves* // Phys. Rev. A.– 2004.– Vol. **70**.– No. 3.– id. 032309.  [↑₃₃](#)
- [18] Wolpert D. H. *Physical limits of inference* // Physica D: Nonlinear Phenomena.– 2008.– Vol. **237**.– No. 9.– Pp. 1257–1281.  [↑₃₃](#)
- [19] King A. D., Raymond J., Lanting T., Harris R., Zucca A., Altomare F., Berkley A. J., Boothby K., Ejtemaee S., Enderud C., Hoskinson E., Huang Sh., Ladizinsky E., MacDonald A. J. R., Marsden G., Molavi R., Oh T., Poulin-Lamarre G., Reis M., Rich Ch., Sato Y., Tsai N., Volkmann M., Whittaker J. D., Yao J., Sandvik A. W., Amin M. H. *Quantum critical dynamics in a 5,000-qubit programmable spin glass* // Nature.– 2023.– Vol. **617**.– No. 7959.– Pp. 61–66.  [↑_{33, 43}](#)

- [20] Miller G. L. *Riemann's hypothesis and tests for primality // STOC '75: Proceedings of the seventh annual ACM symposium on Theory of computing* (5–7 May 1975, Albuquerque, New Mexico, USA), New York: ACM.— ISBN 978-1-4503-7419-4.— С. 234–239.  ↑44

Поступила в редакцию 15.03.2025;
 одобрена после рецензирования 02.04.2025;
 принята к публикации 02.04.2025;
 опубликована онлайн 21.04.2025.

Рекомендовал к публикации

д.ф.-м.н. Н. Н. Непейвода

Информация об авторах:



Милад Джудакизде

Аспирант в области искусственного интеллекта и машинного обучения, исследования сосредоточены на вычислительной сложности, машинном обучении и логике в компьютерных науках.



0000-0002-6167-6237

e-mail:



Анатолий Петрович Бельтюков

Доктор физико-математических наук, профессор. Научные интересы включают вычислительную сложность, классы сложности, субрекурсивные иерархии, интуиционистскую математику, конструктивные системы, механизацию доказательств, сложность доказательств, модели вычислений, субструктурные логики, слабые арифметики и логику в компьютерных науках.



0000-0002-3433-9067

e-mail:

Авторы внесли равный вклад в подготовку публикации.

Декларация об отсутствии личной заинтересованности: благополучие авторов не зависит от результатов исследования.

УДК 004.75

 10.25209/2079-3316-2025-16-2-55-79

Особенности организации единого научно-технологического пространства

Николай Сергеевич **Абрамов**¹, Виталий Петрович **Фраленко**²

^{1,2}Институт программных систем им. А.К. Айламазяна РАН, Веськово, Россия

Аннотация. Единое научно-технологическое пространство способно обеспечить общую платформу для более эффективного взаимодействия научного сообщества при реализации комплексных научно-технических проектов, поиска и привлечения к различным научным исследованиям и проектам необходимых ресурсов. В связи с актуальными межнациональными задачами современности в области науки и образования это могут быть, например, пространства России, Союзного государства, Арктической зоны, стран ближнего Зарубежья и др. В настоящей работе описаны требования и даны предложения по построению основных компонентов единого научно-технологического пространства в виде автоматических рабочих мест аналитиков и информационных порталов. Большое внимание уделено концептуальной модели информационного портала, приведена его архитектура. В статье затронуты важные вопросы в области стандартизации компонент единого научно-технологического пространства.

Ключевые слова и фразы: единое научно-технологическое пространство, автоматизированное рабочее место, интерфейс, интернет-портал, знания, требования к компонентам

Благодарности: Работа выполнена в рамках госбюджетной темы ИПС им. А.К. Айламазяна РАН № 125021302067-9 (срок выполнения: 2025–2027 гг.)

Для цитирования: Абрамов Н.С., Фраленко В.П. *Особенности организации единого научно-технологического пространства* // Программные системы: теория и приложения. 2025. Т. 16. № 2(65). С. 55–79. https://psta.pspiras.ru/read/psta2025_2_55-79.pdf

Введение

Задача построения единого научно-технологического пространства (ЕНТП) является актуальной, поскольку позволяет ускорить процессы информационного взаимодействия между распределенными системами управления и оптимизации. Важность создания ЕНТП отмечалась, например, в 2021 году в ходе пленарного заседания «Научно-техническое сотрудничество России и Белоруссии в эпоху цифровизации» [1]. Российская Федерация и Республика Беларусь давно и успешно взаимодействуют в области науки и техники и совместно решают стратегическую задачу создания единого научно-технологического и цифрового пространства Союзного государства. Ожидается, что единое пространство свяжет Россию и Беларусь до 2030 года [2]. Отметим, что одна из задач научно-технологического развития Российской Федерации, как заявлено в указе Президента Российской Федерации от 28 февраля 2024 г. № 145 «О Стратегии научно-технологического развития Российской Федерации», полностью определяет актуальность настоящего исследования: «г) сформировать эффективную систему управления в области науки, технологий и производства и осуществления инвестиций в эту область, обеспечив единое научно-технологическое пространство, ориентированное на решение государственных задач и удовлетворение потребностей экономики и общества» (см. раздел IV. Цель и основные задачи научно-технологического развития) [3].

ЕНТП – взаимосвязанная и эффективно функционирующая система организаций, инфраструктур и механизмов, обеспечивающая генерацию, распространение и применение знаний и технологий для достижения национальных целей социально-экономического развития, повышения конкурентоспособности страны и улучшения качества жизни населения. Основные характеристики ЕНТП:

- обеспечение доступа к научным публикациям, базам данных, патентной информации, результатам исследований и разработок;
- обмен оборудованием, программным обеспечением, опытом и экспертными знаниями;
- поддержка коллабораций между учеными, инженерами, исследователями и предприятиями;
- использование общих стандартов, протоколов и форматов данных для облегчения обмена информацией и совместной работы;
- создание благоприятной среды для генерации новых идей, разработок и их коммерциализации;

- обеспечение доступа к образовательным ресурсам, программам повышения квалификации и обмена опытом.

Выделим следующие плюсы, получаемые от использования ЕНТП, подтверждающие их актуальность и необходимость работы с ними:

(1) Ускорение научно-технического прогресса:

- объединение научных коллективов, исследовательских институтов, университетов и промышленных предприятий, имеющих разрозненные вычислительные, информационные, финансовые и кадровые ресурсы, создает синергетический эффект, когда совместная работа приводит к более быстрым и качественным результатам, чем работа по отдельности;
- обмен информацией о текущих проектах и результатах исследований позволяет избежать дублирования работ, экономить ресурсы и направлять их на перспективные направления;
- общее использование инфраструктуры (лаборатории, вычислительные центры, базы данных) снижает затраты на исследования и разработки;
- стимуляция конкуренции и сотрудничества.

(2) Повышение конкурентоспособности:

- быстрое информационное взаимодействие позволяет оперативно реагировать на изменения рынка, разрабатывать и внедрять новые продукты и услуги, опережая конкурентов;
- оптимизация производственных процессов на основе данных, полученных из разных источников, позволяет снизить издержки и повысить рентабельность;
- облегчение доступа к новым рынкам позволяет предприятиям более эффективно продвигать свою продукцию и услуги;
- обеспечение устойчивых темпов роста экономики в условиях быстрой цикличности экономических процессов.

(3) Решение глобальных проблем:

- обмен данными, знаниями и технологиями, необходимыми для решения проблем изменения климата, пандемий, нехватки ресурсов и т.п.;
- стимуляция разработки экологически чистых технологий, возобновляемых источников энергии и других решений, направленных на устойчивое развитие.

(4) Оптимизация управления и повышение эффективности принятия решений:

- сбор, анализ и обработка данных от различных распределенных систем для получения ценной информации, необходимой для принятия обоснованных управленческих решений;

- быстрый обмен информацией между различными системами позволяет оперативно реагировать на чрезвычайные ситуации, такие как стихийные бедствия, аварии и техногенные катастрофы;
- информационное взаимодействие между различными участниками логистических цепочек позволяет оптимизировать процессы транспортировки, хранения и доставки товаров, снижая издержки и повышая эффективность.

Основной глобальной целью является создание организационных, правовых, информационных и финансово-экономических основ функционирования ЕНТП в рамках Союзного государства. Для этого необходимо решить задачи по

- формированию единой системы приоритетов научно-технологического развития;
- координации научных исследований и разработок;
- построению организационной системы функционирования научно-технологического и инновационного взаимодействия на межгосударственном уровне;
- определению форм и условий совместной деятельности научных и образовательных структур России и Беларуси, форм и механизмов сотрудничества в области коммерциализации результатов научно-исследовательских и опытно-конструкторских работ и трансфера технологий;
- созданию системы информационного обеспечения сектора исследований и разработок Союзного государства;
- реализации совместных научно-технологических и инновационных программ, проектов и мероприятий [4].

На успешную реализацию ЕНТП влияет множество аспектов, среди которых можно отметить следующие:

- необходима устойчивая политическая поддержка со стороны руководства России и Белоруссии, последовательное исполнение принятых решений;
- важно создать эффективные механизмы управления и координации, которые будут обеспечивать согласованность действий различных министерств, ведомств и научных организаций;
- необходимы достаточные объемы финансирования для реализации совместных проектов и программ;
- важно упростить процедуры для ученых и исследователей, чтобы они могли свободно перемещаться между странами, обмениваться информацией и участвовать в совместных проектах;

- необходимо привлекать частный капитал к финансированию научно-технических разработок и внедрению инновационных технологий;
- проект должен быть гибким и способным адаптироваться к меняющейся экономической и политической ситуации в мире;
- принятие решений и реализация проектов должны быть прозрачными и подотчетными.

Для эффективного функционирования ЕНТП, а также для его успешного развития и адаптации к растущим потребностям, крайне важны аспекты масштабируемости и устойчивости. В контексте ЕНТП масштабируемость предполагает

- возможность добавления серверов, расширения хранилищ данных, использования облачных ресурсов для обработки больших объемов информации и выполнения сложных расчетов;
- возможность увеличения пропускной способности каналов связи, оптимизации сетевых протоколов для обеспечения быстрого и надежного обмена данными между участниками ЕНТП;
- обеспечение доступа к ресурсам ЕНТП для новых исследователей, научных групп, университетов и предприятий;
- возможность добавления новых программных и аппаратных средств, которые расширяют функциональность ЕНТП и удовлетворяют возникающие потребности научного сообщества;
- возможность внедрения эффективных методов сбора, хранения, обработки и анализа больших данных в научных исследованиях;
- гибкость архитектуры ЕНТП, возможность быстро адаптироваться к новым задачам, технологиям и научным направлениям.

Обеспечение устойчивости ЕНТП требует

- создания резервных копий данных и программного обеспечения, а также дублирования критически важных компонентов системы;
- размещения серверов и хранилищ данных в разных географических регионах для защиты от локальных аварий (пожары, землетрясения, отключения электроэнергии и пр.);
- внедрения механизмов автоматического обнаружения сбоев и переключения на резервные системы без прерывания работы ЕНТП;
- непрерывного мониторинга производительности, доступности и безопасности ресурсов ЕНТП для оперативного выявления и устранения проблем;
- разработки и тестирования планов восстановления системы после аварийных ситуаций, включающих процедуры восстановления данных, перезапуска сервисов и оповещения пользователей;

- внедрения мер защиты от кибератак, вирусов и других угроз, которые могут нарушить работу ЕНТП или привести к утечке конфиденциальной информации.

Для обеспечения масштабируемости и устойчивости ЕНТП используются различные методики и технологии, применяемые в сложных информационных системах. В том числе (но не ограничиваясь):

- увеличение мощности существующих серверов путем добавления процессоров, оперативной памяти и дискового пространства (подходит для небольших систем, но имеет ограничения по максимальной мощности одного сервера);
- распределение нагрузки между несколькими серверами, работающими параллельно (обеспечивает большую масштабируемость и отказоустойчивость, так как выход из строя одного сервера не приводит к полной остановке системы);
- использование вычислительных ресурсов, хранилищ данных и других сервисов, предоставляемых облачными провайдерами (облачные технологии обеспечивают высокую масштабируемость, отказоустойчивость и гибкость, позволяя быстро адаптироваться к изменяющимся потребностям);
- использование контейнеров (например, “Docker”) для упаковки приложений и их зависимостей в изолированные среды (контейнеры позволяют легко развертывать и масштабировать приложения, а также обеспечивают их переносимость между различными платформами);
- использование инструментов оркестрации контейнеров для автоматизации развертывания, масштабирования и управления контейнерами (обеспечивает высокую отказоустойчивость и эффективность использования ресурсов);
- разработка приложений в виде набора небольших, независимых сервисов, которые взаимодействуют друг с другом через API (микросервисная архитектура обеспечивает большую гибкость, масштабируемость и отказоустойчивость, так как каждый сервис может масштабироваться и восстанавливаться независимо от других);
- распределение нагрузки между несколькими серверами для обеспечения равномерной загрузки и высокой производительности (балансировка нагрузки повышает отказоустойчивость, так как при выходе из строя одного сервера нагрузка автоматически перераспределяется на другие).

Семантическая интероперабельность также является ключевым требованием для обеспечения эффективного функционирования ЕНТП.

Она означает способность различных систем, приложений и участников обмениваться и понимать данные и знания, не теряя их смысл и контекст. Требования к ЕНТП с точки зрения обеспечения семантической интероперабельности:

- (1) Разработка и использование общих, согласованных и поддерживаемых онтологий (формализованных описаний концепций и отношений между ними) и тезаурусов (словарей, содержащих синонимы, антонимы и другие смысловые связи между терминами) для описания научных и технических объектов, процессов и явлений. Обеспечение возможности совместного использования и расширения этих онтологий и тезаурусов.
- (2) Применение стандартизированных форматов данных и метаданных для описания научных данных, технических спецификаций, результатов моделирования и других объектов ЕНТП. Использование открытых стандартов, основанных на общепринятых принципах и практиках.
- (3) Разработка механизмов разрешения неоднозначности терминов и понятий, используемых в разных областях науки и техники. Обеспечение возможности контекстуализации данных и знаний, чтобы пользователи могли понимать их смысл в контексте конкретной задачи или области.
- (4) Использование семантических web-сервисов и API для обеспечения возможности автоматизированного доступа и обмена данными и знаниями между различными системами и приложениями. Использование семантических web-стандартов (RDF, OWL, SPARQL и др.) для описания данных и знаний в машиночитаемом формате.

Рассмотрим несколько вариантов разработки ЕНТП – итеративный и реализация проекта сразу. Итеративный подход (пошаговое создание) отличается тем, что позволяет выявлять и устранять проблемы на ранних этапах, прежде чем будут вложены значительные ресурсы; позволяет адаптироваться к меняющимся условиям и корректировать стратегию в процессе реализации; позволяет постепенно вовлекать различных участников в процесс и обеспечивать их поддержку. Минусы: может занять больше времени, чем реализация проекта сразу; постепенное движение может привести к потере фокуса на конечной цели. Второй подход – реализация проекта сразу, он позволяет быстро достичь поставленных целей; обеспечивает четкий фокус на конечной цели. Минусы: в случае неудачи могут быть потеряны значительные ресурсы; сложно адаптироваться к меняющимся условиям; может вызвать сопротивление со стороны участников, которые не были вовлечены в процесс на ранних

этапах. В случае с созданием ЕНТП, пошаговый подход представляется более разумным и оправданным. Он позволяет постепенно интегрировать научные и технические потенциалы двух стран, выявлять и устранять проблемы на ранних этапах, а также адаптироваться к меняющимся условиям. Однако, важно понимать, что пошаговый подход требует четкой стратегии, эффективного управления и постоянного мониторинга прогресса. Необходимо четко определить цели и задачи каждого этапа, а также критерии успеха, чтобы можно было оценить эффективность предпринятых мер и своевременно скорректировать стратегию.

Интеграция ЕНТП в условиях разнообразия систем представляет собой сложную задачу, которая включает несколько ключевых проблем. Основные из них:

- (1) В современных научных и технологических исследованиях используются различные программные и аппаратные средства, которые часто несовместимы друг с другом. Различия в архитектуре систем могут осложнять обмен данными и сотрудничество в целом между участниками ЕНТП. Это приводит к необходимости разрабатывать «мосты» для обеспечения взаимодействия, что требует значительных ресурсов и времени.
- (2) Разработка и внедрение единых протоколов обмена данными и взаимодействия между системами является критически важным шагом для интеграции. Стандартизация протоколов часто сталкивается с сопротивлением со стороны участников, которые предпочитают использовать свои собственные решения, что замедляет процесс интеграции.
- (3) Разные системы могут использовать различные форматы данных и метаданных, что затрудняет их совместное использование. Отсутствие единого стандарта для описания данных приводит к возникновению «островов» данных, где информация оказывается недоступной для анализа и применения.
- (4) Даже если существуют технологии для интеграции, их реализация может потребовать больших инвестиций, как финансовых, так и временных. Разработка и поддержка совместимых систем часто выходят за рамки бюджета менее крупных научных организаций.
- (5) Интеграция данных между различными организациями может вызывать проблемы с правами на интеллектуальную собственность, защитой данных и соблюдением нормативных актов.

Перечисленные выше проблемы требуют комплексного подхода и взаимодействия среди ученых, разработчиков технологий и регулирующих органов для успешной интеграции ЕНТП.

Приведем несколько ярких примеров внедрений ЕНТП:

- (1) *Horizon Europe*^{URL} (Европейский союз) — программа, направленная на создание единого научного и инновационного пространства. Объединяет исследователей, университеты и компании из 27 стран ЕС и стран-партнеров. Финансирует междисциплинарные проекты в области климата, здравоохранения, цифровых технологий.
- (2) *CERN*^{URL} — международный проект, объединяющий ученых из 23 стран для исследований в области физики частиц. Инфраструктура включает Большой адронный коллайдер и общие базы данных (например, “Open Data Portal”).
- (3) *Национальная сеть производственных инноваций*^{URL} (Manufacturing USA, США) — сеть из 16 институтов, объединяющая промышленность, университеты и государственные органы для развития передовых производственных технологий.
- (4) *Инновационный центр «Сколково»*^{URL} (Россия) — технопарк, созданный для интеграции науки, образования и бизнеса. Включает кластеры IT, биомедицины, энергетики и космических технологий.
- (5) *Международная космическая станция (МКС)*^{URL} — пример глобального сотрудничества в течение более 25 лет: 16 стран участвуют в совместных экспериментах в области микрогравитации, биологии и материаловедения.

Настоящая работа развивает направление, обозначенное в исследованиях [5–8], в которых даны практические предложения, лежащие в основе построения элементов ЕНТП.

1. Требования к интеграции интеллектуальных технологий и систем обработки знаний в ЕНТП

Согласно современным принципам построения информационной инфраструктуры, взаимодействие должно быть организовано через интернет-порталы аналитических центров, представляющие «единую точку доступа» ко всем необходимым ресурсам и сервисам, необходимым для научно-исследовательской, инновационной и образовательной деятельности. Порталы агрегируют, структурируют и предоставляют персонализированный доступ к информации, данным, экспертным знаниям, инструментам и сервисам, способствуя обмену знаниями, сотрудничеству и ускорению инновационных процессов. Модель ЕНТП предназначена для предоставления пользователям различных категорий доступа, соответствующего их полномочиям, к интегрированным информационным системам, сервисам и ресурсам (базам данных, знаний и т.д.). Это обеспечивает возможность консолидированной обработки знаний и принятия обоснованных решений.

Интернет-портал в рамках ЕНТП обладает рядом ключевых преимуществ, делающих его удобным средством для интеграции интеллектуальных технологий и систем обработки знаний:

- портал становится единой точкой входа ко всем ресурсам ЕНТП: научным публикациям, базам данных, каталогам оборудования, информации о проектах, экспертным профилям и многому другому; это устраняет необходимость искать информацию на разрозненных сайтах и в различных системах;
- портал может быть оснащен интеллектуальными поисковыми системами, использующими методы обработки естественного языка, машинного обучения и семантического анализа; это позволяет искать не только по ключевым словам, но и по смыслу запроса, контексту и связям между понятиями; выявлять тенденции, закономерности и скрытые связи в данных, которые были бы незаметны при ручном анализе; организовывать данные в соответствии с различными критериями и представлять их в удобном для пользователя виде;
- портал может быть интегрирован с различными системами обработки знаний, такими как системы управления знаниями (Knowledge Management Systems, KMS) для сбора, хранения, систематизации и распространения знаний внутри ЕНТП; экспертные системы для предоставления консультаций и помощи в принятии решений в различных областях науки и техники; системы Data Mining для анализа данных и выявления скрытых закономерностей; системы визуализации и представления информации в наглядной и понятной форме;
- портал может предоставлять инструменты для совместной работы: форумы и дискуссионные площадки для обмена идеями, обсуждения проблем и поиска решений; инструменты для совместной работы над документами для одновременного редактирования и обсуждения научных статей, отчетов и других документов; виртуальные лаборатории и симуляции для проведения экспериментов и исследований в виртуальной среде; средства видеоконференцсвязи для проведения онлайн-встреч, семинаров и конференций;
- портал может быть настроен в соответствии с индивидуальными потребностями и интересами каждого пользователя; например, пользователь может выбрать интересующие его научные области, указать свои навыки и опыт, а портал будет предлагать ему релевантную информацию, проекты и возможности для сотрудничества.

Интернет-портал для ЕНТП гораздо больше, чем просто хранилище и средство доступа к данным, потому что он

- обеспечивает контекст – данные становятся информацией, когда они представлены в контексте; портал помогает связать данные из разных источников, предоставляя пользователю целостную картину;
- позволяет быстро найти специалистов в нужной области для консультаций, сотрудничества или обмена опытом;
- помогает организациям и отдельным лицам эффективно управлять своими знаниями, сохранять, систематизировать и делиться ими;
- создает благоприятную среду для генерации новых идей, разработки новых технологий и их коммерциализации;
- использует интеллектуальные технологии для автоматизации процессов поиска, анализа и обработки информации, освобождая время для более творческой работы;
- предоставляет доступ к образовательным ресурсам, программам повышения квалификации и обмена опытом, способствуя развитию кадрового потенциала;
- становится «живой» платформой – благодаря интерактивным инструментам и возможностям для сотрудничества, портал становится местом, где ученые и инженеры могут общаться, обмениваться идеями и совместно работать над решением сложных задач; он перестает быть просто статичным хранилищем данных, а превращается в динамичную и развивающуюся среду.

Портал знаний в интернете – web-сайт, созданный для конкретной аудитории и представляющий собой сложную систему взаимосвязанных информационных ресурсов, сервисов и баз знаний. Функциональная структура портала включает, как правило, следующие подсистемы:

- базовые подсистемы (авторизации и аутентификации; настройки пользовательского интерфейса и пр.) и средства взаимодействия;
- средства организации и доступа к данным и знаниям (хранение данных и работа с информационной базой портала, работа с метаинформацией);
- средства управления (управление производительностью и администрирование, средства обеспечения безопасности портала; управление кластерами, средства кэширования контента и др.);
- средства интеграции (обеспечение межпортальной интеграции баз данных, метаданных, поисковых процедур, систем безопасности, приложений, событийных и справочных систем; формирование кооперативной системы «зеркал» и виртуальных серверов, CDN-сети доставки контента, grid-структуры);
- средства коммуникации, средства порталных приложений и профильных сервисов, проблемные информационные системы.

Пример архитектуры единого научно-технологического пространства дан на рис. 1.

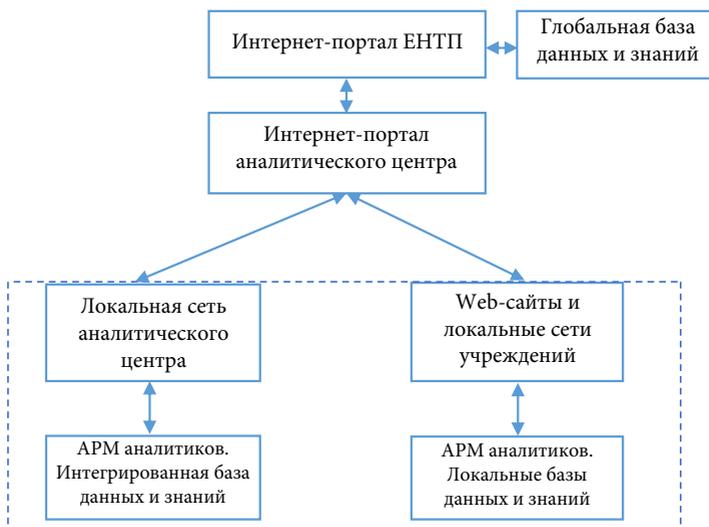


Рисунок 1. Архитектура единого научно-технологического пространства

Порталы предназначены для обеспечения обмена информацией между специалистами, формирования баз знаний, автоматизированного процесса приобретения, обработки знаний и принятия решений с применением автоматизированных рабочих мест (АРМ) аналитиков. Внедрение информационных компьютерных технологий начинается с построения абстрактной модели единого информационного пространства государства, региона, учреждения науки, образования, производства. В этой модели должны быть описаны внутренние процессы (объекты), связи между ними. Под связями между процессами понимаются информационные потоки, циркулирующие между элементами системы. Реализация модели требует большей, чем имеется сегодня, стандартизации и унификации интерфейсов для достижения интеграции отдельных служб и обеспечения координации в работе, развития и поддержки концепции виртуальных данных и вычислительных мощностей.

1.1. Требования к порталам знаний

Порталы знаний становятся массовым информационно-программным продуктом в едином научно-технологическом пространстве. Это придает

актуальность решению комплекса методических, технологических и инструментальных проблем создания порталов знаний и АРМ аналитиков различного назначения. Наиболее сложной задачей является обеспечение высокого уровня совместимости между приложениями, разработанными различными производителями и работающими на различных платформах, способности интеллектуальных систем эффективно взаимодействовать друг с другом для решения прикладных задач.

Работа порталов и АРМ аналитиков должна основываться на применении WWW и серверов БД для представления и хранения информации; использовании технологии «клиент-сервер» с минимальными техническими требованиями к клиентской части; поддержке распределенных баз данных. Использование интернет-технологий при построении интеллектуальных систем как общего назначения, так и специализированных является доминирующим принципом в едином научно-технологическом пространстве. Функциональные возможности современных информационных порталов должны позволять обрабатывать большой объем данных, эффективно доставлять пользователю необходимую информацию, обеспечивать связь между базами знаний, осуществлять поддержку принятия решений.

Первоочередными задачами в области стандартизации информационного обеспечения порталов знаний являются

- поддержание единой системы классификации и кодирования информации;
- обеспечение совместимости систем посредством стандартизации форматов обмена данными и протоколов взаимодействия;
- унификация технических требований, функциональных стандартов компонент информационных систем;
- внедрение общей методологии для информационного обеспечения.

При построении ЕНТП особое внимание стоит уделить использованию существующих стандартов, таких как ONIX, ISO/IEC, ISO 8000 и др. Рассмотрим кратко данные стандарты:

- (1) ONIX (Online Information Exchange) — стандарт обмена информацией о других стандартах, в т.ч. о книгах и учебных материалах. Он позволяет более эффективно передавать и обрабатывать информацию о товарах и услугах в научно-техническом пространстве. Стандарт ONIX используется для создания баз данных и каталогов, что делает его важным для интеграции научных данных и публикаций.
- (2) Стандарты ISO/IEC (Международная организация по стандартизации и Международная электротехническая комиссия) охватывают

широкий спектр областей, связанных с информационными технологиями и системами. Эти стандарты помогают обеспечивать совместимость и взаимосвязанность технических систем в научно-технологических пространствах, что способствует эффективному обмену данными и исследованиям.

- (3) Стандарт ISO 8000 — стандарт, связанный с управлением качеством данных и их интеграцией. Он акцентирует внимание на важности качественных данных для достижения успешных результатов в науке и технологии. Стандарт определяет принципы управления данными, которые могут быть использованы для формирования надежных ЕНТП.

Эти и другие стандарты могут значительно способствовать созданию и поддержанию ЕНТП. Их использование обеспечивает структурированный подход к обмену информацией, гарантируя качество и совместимость данных. Однако для более эффективного функционирования ЕНТП необходима разработка новых методологических подходов и дальнейшая адаптация существующих стандартов к специфическим нуждам научного сообщества и конкретных решаемых задач.

1.2. Основные требования к интеллектуальной компоненте

В настоящее время существует множество нормативных документов, регулирующих вопросы разработки и содержания систем, использующих технологии ИИ, например, [9–12].

Интеллектуальная система в составе единого научно-технологического пространства реализуется на основе технологий обнаружения, приобретения, хранения и обработки знаний с помощью инструментальных средств и опирается на технологию “Internet”. Интеллектуальная компонента системы должна быть реализована с использованием технологии интеллектуальных агентов и других современных достижений искусственного интеллекта. Интеллектуальная компонента должна реализовывать поисковые средства, средства анализа данных, базу данных и базу знаний. Модули интеллектуальной компоненты системы должны быть

- адаптивными – способными делать свою работу лучше с течением времени;
- обучаемыми – обладающими способностью усваивать новую информацию и использовать ее для совершенствования своих функций;
- коммуникативными – способными взаимодействовать с другими модулями и тем самым, при необходимости, в ходе решения своей задачи выполнять действия вне рамок собственных функций.

Система поддержки принятия решений должна обеспечивать работу удаленных пользователей. Корпоративная система поддержки решений включает

- формирование списка проблем (или целей);
- формирование экспертного совета или комитета;
- привлечение независимых экспертов и аудиторов для обсуждения проблем;
- систему анкетирования, в том числе по результатам принятых решений (экспертное заключение);
- ранжирование важности проблем и мнений экспертов;
- возможность подключения специализированных справочников;
- систему оповещений;
- систему администрирования;
- возможность поиска, фильтрации и группировки информации;
- терминологическую базу знаний.

Интерфейс интеллектуальной системы должен предусматривать

- доступ к данным по требуемому формату, формулируемому при запросе;
- доступ к вторичным данным (знаниям);
- фильтрацию и первичную обработку знаний.

1.3. Требования по назначению и составу инструментального комплекса

Инструментальный комплекс – набор интеллектуальных инструментов, не зависящих от конкретной прикладной области. Он используется для решения задач, требующих создания динамических моделей предметной области, правил вывода и принятия решений, организованных в виде баз данных и знаний с соответствующим человеко-машинным интерфейсом. Программные и аппаратные средства в составе инструментальных комплексов для ЕНТП обеспечивают поддержку всего жизненного цикла научно-исследовательских и инновационных проектов. В состав комплекса инструментально-программных средств для решения задач инженерии знаний входят

- инвариантная часть, представляющая собой универсальное ядро интеллектуальной системы, содержащее базу знаний, интегрированную с базой данных; универсальную моделирующую среду; интерфейсы;

- прикладная технологическая часть, содержащая комплекс специализированных моделирующих программных средств, в том числе модели систем распознавания и анализа текущей ситуации; модели систем отображения текущей ситуации.

Результатом применения инструментального комплекса должна стать разработка экспериментальных прикладных программ и систем, использующих методы искусственного интеллекта.

1.4. Требования к защите данных, аутентификации, идентификации и шифрованию при построении ЕНТП

Как и при проектировании любой сложной информационной системы, создание ЕНТП требует особого внимания к вопросам защиты данных, аутентификации, идентификации и шифрования. Ниже представлены критические требования, которые следует учитывать при проектировании и реализации таких пространств.

(1) Защита данных:

- данные должны быть защищены от несанкционированного доступа, необходимо применять механизмы шифрования как для хранимых данных, так и для передаваемых данных;
- данные должны быть защищены от изменений и подделки, необходимо использовать цифровые подписи и контрольные суммы для проверки целостности;
- необходимо обеспечить доступность данных для уполномоченных пользователей в любое время, а также устойчивость к отказам и кибератакам;
- необходимо определить, как долго данные будут храниться, и установить правила утилизации по истечении этого срока.

(2) Аутентификация:

- необходимо обеспечить поддержку использования нескольких факторов аутентификации (таких, как пароль, токен, биометрические данные);
- необходимо контролировать активные сессии и обеспечивать автоматический выход из системы после определенного времени бездействия;
- необходимо обеспечить ведение журналов аутентификации для мониторинга и последующего анализа попыток доступа к системе.

(3) Идентификация:

- каждому пользователю или устройству должен быть назначен уникальный идентификатор для предотвращения коллизий и проблем с доступом;

- необходим регулярный аудит используемых идентификаторов для обеспечения их актуальности и безопасности;
- пользователям должен предоставляться лишь тот уровень доступа, который им необходим для выполнения служебных обязанностей.

(4) Шифрование:

- для защиты данных должны использоваться актуальные и проверенные алгоритмы шифрования (такие, как AES-256);
- необходимо шифровать данные как на уровне передаваемых сообщений (TLS), так и на уровне хранения в базах данных;
- должна использоваться надежная система управления криптографическими ключами, включая их генерацию, распределение, хранение и утилизацию.

(5) Комплаенс и регуляция:

- должны выполняться требованиям законодательства о защите данных, включая право пользователей на доступ к своим данным и их удаление;
- должны соблюдаться международные стандарты безопасности информации, такие как ISO/IEC 27001.

(6) Обучение и культура безопасности:

- необходимо постоянное обучение и информирование пользователей о лучших практиках безопасности данных, а также о последствиях их нарушения;
- необходимо обеспечить формирование у пользователей понимания значимости безопасности данных на всех уровнях ЕНТП.

Эти и другие требования помогут создать безопасную среду для работы в ЕНТП, обеспечивающую защиту данных и пользователей.

2. Концептуальная модель информационного портала

В построении единого научно-технологического пространства на основе интеллектуальных систем преобладает метод создания набора подпространств в виде порталов, баз знаний и АРМ с консолидацией информации в центральном хранилище – обычно на сервере.

Предлагается стандартная трехзвенная архитектура клиент-сервер (см. рис. 2), состоящая из хранилища данных, сервера приложений и клиентской части. Информация хранится в базе данных, административная часть обеспечивает управление содержимым базы данных, а извлеченная информация отображается в клиентской части. Открывая



Рисунок 2. Схема портала с интеллектуальной компонентой

архитектура портала позволяет достаточно легко подключать к нему новые модули, расширяющие его функциональность в соответствии с новыми задачами. Пользователь информационного узла (портала) запрашивает определенную html-страницу с некоторыми специфическими данными. Система управления обращается к базе данных, получает нужную информацию и трансформирует ее в требуемый для передачи пользователю формат.

Базовым компонентом является ядро портала, представляющее собой программно-аппаратный комплекс. В число функциональных блоков, входящих в состав ядра, включены, в том числе, «Блок управления базой данных», «Интерфейс администратора», «Системный интерфейс».

Схема включает следующие компоненты:

- клиентская часть в составе: рабочие станции аналитиков с системой аутентификации и предоставления доступа к приложениям, имеющие сетевое соединение с серверной частью, web-браузер, вспомогательные утилиты (плагины, редакторы, фильтры форматов);

- закрытая корпоративная сеть, объединяющая компьютеры клиентской и серверной части системы;
- серверная часть системы, в составе: аппаратные средства – кластер; общесистемный уровень ПО серверной части, средства поддержки параллельного выполнения программ; ПО web-сервера; интеллектуальная компонента ПО системы, включающая поисковые средства, средства анализа данных, базу данных и базу знаний.

Информационная база системы – структурированное хранилище данных, знаний и информации, относящихся к различным аспектам науки, технологий и инноваций. Она может включать в себя научные публикации, патенты, отчеты об исследованиях, техническую документацию, данные о проектах, информацию об экспертах и организациях, а также другие релевантные ресурсы. В том числе информационная база системы предназначена для хранения

- результатов чтения информации: единичные документы, актуальные копии источников информации, версии (исторические срезы) источников информации, сведения о выполненном чтении;
- данных, полученных в процессе адаптации (обучения) интеллектуальной компоненты системы, – профили аналитиков, поисковых машин и источников информации;
- информации, полученной в процессе интеллектуального анализа снятой информации, – результаты анализа документов, тематических подборок, выявленные скрытые отношения, связи и знания.

Средства анализа и предоставления информации предназначены для интеллектуальной обработки информационной базы системы и внешней информации и предоставления аналитикам результатов этой обработки. Данные средства должны обеспечивать

- возможность анализа смыслового содержания отдельных документов и тематических подборок, связанных массивов на предмет выделения значимых ключевых слов, выявления скрытых отношений и связей;
- возможность работы с информацией на различных иностранных языках;
- накопление и пополнение знаний, полученных в результате анализа информации по интересующей тематике;
- предъявление по требованию различных форм одной и той же информации;
- обеспечение различных режимов использования предъявляемой информацией – совместное использование рабочей группой аналитиков и/или снятие копии для передачи заинтересованным лицам.

Аппаратные средства серверной части должны обеспечивать аппаратную платформу для эффективной реализации всех функций ПО системы.

Рассмотренная модель портала предоставляет возможности по организации клиентского места с целью устранения информационной перегрузки и оптимизации доступа к необходимой информации, отслеживания и протоколирования действий пользователей для поддержания портала в рабочем состоянии, оценки эффективности его работы и оперативно-го устранения сбоев, обеспечению активного доступа и отображению информации из множественных источников данных.

Интеллектуальная система строится с учетом следующих базовых требований:

- масштабируемость;
- открытость системы (API);
- надежность;
- независимость от аппаратной платформы;
- общесистемный уровень ПО серверной части;
- программное обеспечение web-сервера;
- интеллектуальная компонента ПО системы, включает поисковые средства, средства чтения данных, средства анализа данных, базу данных и базу знаний;
- формирование специализированного банка формализованных решений по функциональным моделям информационных порталов и их подсистем с учетом выделенных типовых классов.

Заключение

Активизация работ по формированию ЕНТП является ключевым фактором обеспечения конкурентоспособности, безопасности и технологического суверенитета. Его реализация способствует расширению приоритетных направлений научного и научно-технического сотрудничества на основе долгосрочных партнерских и равноправных международных отношений по направлениям, обеспечивающим решение стратегических задач развития государства.

В работе рассмотрены общие требования для формирования ЕНТП. Сделаны оценки актуальности работ по созданию ЕНТП. Обоснована роль интернет-порталов как средства для управления знаниями и интеграции технологий. Выявлены требования к ЕНТП с точки зрения обеспечения семантической интероперабельности. Приведена концептуальная модель информационного портала.

Список использованных источников

- [1] Владимир Путин отметил важность создания единого научно-технологического и цифрового пространства России и Республики Беларусь.– Минпросвещения России.– 2021. [URL](#) ↑⁵⁶
- [2] Сазонова Д. П. *Направления и механизмы формирования единого научно-технологического пространства Союзного государства // Проектирование будущего. Проблемы цифровой реальности*, Труды 5-й Международной конференции (Москва, 3-4 февраля 2022 г.), Москва: ИПМ им. М.В.Келдыша.– 2022.– ISBN 978-5-98354-063-7.– С. 163-170. [doi](#) ↑⁵⁶
- [3] *Указ Президента Российской Федерации от 28 февраля 2024 г. № 145 «О Стратегии научно-технологического развития Российской Федерации»*.– 2024. [URL](#) ↑⁵⁶
- [4] *Постановление Совета Министров Союзного государства от 16 июня 2017 г. № 19 «Об основных направлениях формирования единого научно-технологического пространства Союзного государства» с изменениями и дополнениями*.– 2017. [URL](#) ↑⁵⁸
- [5] Талалаев А. А., Фраленко В. П., Хачумов В. М., Шишкин О. Г. *Сеть автоматизированных рабочих мест для высокопроизводительной обработки данных дистанционного зондирования арктической зоны // Материалы IV Национального Суперкомпьютерного Форума, НСКФ-2015 (Переславль-Залесский, 24-27 ноября 2015 г.)*, Переславль-Залесский: ИПС им. А.К. Айламазяна РАН.– 2015.– 3 с. [URL](#) ↑⁶³
- [6] Емельянова Ю. Г., Хачумов В. М. *Концепция архитектуры автоматизированного рабочего места для дистанционного мониторинга арктической зоны // Информатика, управление и системный анализ*, Труды IV Всероссийской научной конференции молодых ученых с международным участием.– Т. 1, ИУСА-2016 (Тверь, 8-11 июня 2016 г.), Тверь: Тверской государственный технический университет.– 2016.– ISBN 978-5-905156-26-7.– С. 39-53. * ↑⁶³
- [7] Хачумов В. М., Емельянова Ю. Г., Ватутин В. М., Донцов С. А., Ивашина А. В. *Анализ баз данных ДЗЗ для решения задач мониторинга Арктической зоны // Актуальные проблемы ракетно-космического приборостроения и информационных технологий*, Материалы VIII Всероссийской научно-технической конференции (Москва, 1-3 июня 2016 г.).– 2016.– 15 с. [URL](#) ↑⁶³
- [8] Абрамов Н. С., Емельянова Ю. Г., Талалаев А. А., Фраленко В. П., Хачумов М. В., Хачумов В. М., Шишкин О. Г., Шустова М. В. *Автоматизированное рабочее место «АРМ-Арктика 1.0» для приема-передачи и обработки информации дистанционного зондирования Арктической зоны*, Свидетельство о государственной регистрации программы для ЭВМ № 2017617791, дата поступления заявки: 23.05.2017, дата регистрации: 12.07.2017.– 2017. ↑⁶³
- [9] *ГОСТ Р 59276-2020. Системы искусственного интеллекта. Способы обеспечения доверия. Общие положения*.– 2021. [URL](#) ↑⁶⁸

- [10] ГОСТ Р 59277-2020. Системы искусственного интеллекта. Классификация систем искусственного интеллекта.– 2021. [URL](#) ↑68
- [11] ПНСТ 835-2023 (ISO/IEC TS 4213:2022). Искусственный интеллект. Оценка эффективности моделей и алгоритмов машинного обучения в задаче классификации.– 2024. [URL](#) ↑68
- [12] ПНСТ 838-2023/ИСО/МЭК 23053:2022. Искусственный интеллект. Структура описания систем искусственного интеллекта, использующих машинное обучение.– 2024. [URL](#) ↑68

Поступила в редакцию 13.03.2025;
одобрена после рецензирования 24.04.2025;
принята к публикации 02.05.2025;
опубликована онлайн 13.05.2025.

Рекомендовал к публикации

д.т.н., проф. В. М. Хачумов

Информация об авторах:



Николай Сергеевич Абрамов

к.т.н., ведущий научный сотрудник ИЦМС ИПС им. А.К. Айламазяна РАН. Область научных интересов: математические методы синтеза, обработки и анализа изображений и сигналов, искусственный интеллект и принятие решений, интеллектуальный анализ данных и распознавание образов, геометрия.



0000-0002-1612-3879

e-mail:



Виталий Петрович Фраленко

Кандидат технических наук, ведущий научный сотрудник ИЦМС ИПС им. А.К. Айламазяна РАН. Область научных интересов: интеллектуальный анализ данных и распознавание образов, искусственный интеллект и принятие решений, параллельно-конвейерные вычисления, сетевая безопасность, диагностика сложных технических систем, графические интерфейсы.



0000-0003-0123-3773

e-mail:

Авторы внесли равный вклад в подготовку публикации.

Декларация об отсутствии личной заинтересованности: благополучие авторов не зависит от результатов исследования.



Features of the organization of an unified scientific and technological space

Nikolai Sergeevich **Abramov**¹, Vitaly Petrovich **Fralenko**²

^{1,2}Ailamazyan Program Systems Institute of RAS, Ves'kovo, Russia

Abstract. An unified scientific and technological space can provide a common platform for more effective interaction of the scientific community in the implementation of complex scientific and technical projects, search for and attraction of necessary resources to various scientific research and projects. In connection with the current transnational tasks of our time in the field of science and education, these can be, for example, the spaces of Russia, the Union State, the Arctic zone, neighboring countries, etc. This paper describes the requirements and provides proposals for the construction of the main components of the unified scientific-technological space in the form of automatic workstations for analysts and information portals. Much attention is paid to the conceptual model of the information portal and its architecture is given. The article touches upon important issues in the field of standardization of the components of the unified scientific-technological space. (*In Russian*).

Key words and phrases: unified scientific-technological space, automated workplace, interface, internet-portal, knowledge, requirements for components

2020 *Mathematics Subject Classification:* 68M14; 68T30

Acknowledgments: The work was carried out within the framework of the State Budget Theme of the Ailamazyan Program Systems Institute of the Russian Academy of Sciences No. 125021302067-9 (completion deadline: 2025–2027).

For citation: Nikolai S. Abramov, Vitaly P. Fralenko. *Features of the organization of an unified scientific and technological space*. Program Systems: Theory and Applications, 2025, **16**:2(65), pp. 55–79. (*In Russ.*).

https://psta.psiras.ru/read/psta2025_2_55-79.pdf

References

- [1] Vladimir Putin noted the importance of creating a single scientific, technological and digital space between Russia and the Republic of Belarus, *Minprosveshheniya Rossii*, 2021 (in Russian). [URL](#)
- [2] D. P. Sazonova. “Trends and mechanism of the unified scientific and technological space of the Union State”, *Proektirovanie budushhego. Problemy cifrovoj real'nosti*, Trudy 5-j Mezhdunarodnoj konferencii (Moskva, 3-4 fevralya 2022 g.), IPM im. M.V.Keldysha, M., 2022, ISBN 978-5-98354-063-7, pp. 163-170 (in Russian). [DOI](#)
- [3] Decree of the President of the Russian Federation of February 28, 2024 No. 145 «On the Strategy for Scientific and Technological Development of the Russian Federation», 2024 (in Russian). [URL](#)
- [4] Resolution of the Council of Ministers of the Union State of June 16, 2017 No. 19 «On the main directions of formation of a single scientific and technological space of the Union State» with amendments and additions, 2017 (in Russian). [URL](#)
- [5] A. A. Talalaev, V. P. Fralenko, V. M. Xachumov, O. G. Shishkin. “Network of automated workstations for high-performance processing of remote sensing data of the Arctic zone”, *Materialy IV Nacional'nogo Superkomp'yuternogo Foruma*, NSKF-2015 (Pereslavl'-Zalesskij, 24-27 noyabrya 2015 g.), IPS im. A. K. Ajlamazyana RAN, Pereslavl'-Zalesskij, 2015, 3 pp. (in Russian). [URL](#)
- [6] Yu. G. Emel'yanova, V. M. Xachumov. “Concept of the architecture of an automated workplace for remote monitoring of the Arctic zone”, *Informatika, upravlenie i sistemnyj analiz*, Trudy IV Vserossijskoj nauchnoj konferencii molodyx uchenyx s mezhdunarodnym uchastiem. V. 1, IUSA-2016 (Tver', 8-11 iyunya 2016 g.), Tverskoj gosudarstvennyj texniceskij universitet, Tver', 2016, ISBN 978-5-905156-26-7, pp. 39-53 (in Russian).
- [7] V. M. Xachumov, Yu. G. Emel'yanova, V. M. Vatutin, S. A. Doncov, A. V. Ivashina. “Analysis of Earth remote sensing databases for solving Arctic Zone monitoring problems”, *Aktual'nye problemy raketno-kosmicheskogo priborostroeniya i informacionnyx texnologij*, Materialy VIII Vserossijskoj nauchno-texniceskoj konferencii (Moskva, 1-3 iyunya 2016 g.), 2016, 15 pp. (in Russian). [URL](#)
- [8] N. S. Abramov, Yu. G. Emel'yanova, A. A. Talalaev, V. P. Fralenko, M. V. Xachumov, V. M. Xachumov, O. G. Shishkin, M. V. Shustova. *Automated workstation «ARM-Arctic 1.0» for receiving, transmitting and processing remote sensing information of the Arctic zone*, Svidetel'stvo o gosudarstvennoj registracii programmy dlya EVM No 2017617791, data postupleniya zayavki: 23.05.2017, data registracii: 12.07.2017, 2017 (in Russian).
- [9] *GOST R 59276-2020. Artificial intelligence systems. Methods of ensuring trust. General provisions*, 2021 (in Russian). [URL](#)
- [10] *GOST R 59277-2020. Artificial intelligence systems. Classification of artificial intelligence systems*, 2021 (in Russian). [URL](#)

- [11] *PNST 835-2023 (ISO/IEC TS 4213:2022). Artificial intelligence. Evaluation of the effectiveness of machine learning models and algorithms in the classification task*, 2024 (in Russian). [URL](#)
- [12] *PNST 838-2023/ISO/IEC 23053:2022. Artificial intelligence. Framework for describing artificial intelligence systems using machine learning*, 2024 (in Russian). [URL](#)

УДК 004.272.45

 10.25209/2079-3316-2025-16-2-81-109

Скоростные ассоциативные операции и алгоритмы управления и обработки данных для распределенных цифровых систем, действующих в условиях неопределенности

Геннадий Георгиевич **Стецюра**[✉]

Институт проблем управления имени В. А. Трапезникова РАН, Москва, Россия

[✉]gstetsura@mail.ru

Аннотация. Рассмотрены ассоциативные операции и использующие их алгоритмы для скоростной сетевой обработки данных и управления в распределенных цифровых системах с непредвиденными действиями объектов системы, с неизвестными временем и местом появления в системе результатов или отказов, с неожиданным воздействием внешней среды. Эти операции и алгоритмы позволили создать излагаемые в статье способы ассоциативного взаимодействия объектов (по качественным признакам, без использования адресов объектов и данных) и распределенные ассоциативные операции с одновременным участием в каждой операции группы объектов. Выполнение операций не увеличивает время передачи данных. Для ускорения реакции системы на непредвиденные воздействия предложены три вида адаптивных структур связей объектов с разной скоростью реакции на воздействия. Все связи перестраиваются одновременно.

Ключевые слова и фразы: ассоциативное взаимодействие, безадресные операции, одноканальная синхронизация, распределенные цифровые системы, распределенная обратная связь, активные объекты, адаптивные структуры

Для цитирования: Стецюра Г. Г. *Скоростные ассоциативные операции и алгоритмы управления и обработки данных для распределенных цифровых систем, действующих в условиях неопределенности* // Программные системы: теория и приложения. 2025. Т. 16. № 2(65). С. 81–109. https://psta.psir.ru/read/psta2025_2_81-109.pdf

Введение

В предлагаемой статье рассмотрены скоростные ассоциативные операции, алгоритмы обработки данных и управления, предназначенные для использования в распределенных цифровых системах. Предполагается, что эти системы состоят из активных цифровых объектов – групп компьютеров, активных компонент таких компьютеров или физических устройств, использующих встроенные компьютеры.

Активный объект разрешает собственные проблемы без обращения к другим объектам. Однако при участии объектов в решении совместных задач каждый активный объект должен интенсивно взаимодействовать с другими объектами для оценки состояния последних и выработки совместных действий. Это повышает интенсивность взаимодействия объектов через средства связи, замедляя этим в известных системах скорость реакции на события. Объекты действуют в условиях неопределенности, которая создается заранее неизвестными внутренними или внешними событиями, появляющимися в неизвестное время в неизвестных местах. Как следствие, программные средства должны в динамике отслеживать состояние всех объектов системы и с возможно меньшей задержкой реагировать на возникающие события.

Неопределенность усложняет или не допускает обращение к объектам по статическим адресам, что снижает реактивность систем, особенно систем, работающих в режиме жесткого реального времени (ЖРВ).

Для ускорения реактивности статья позволяет заменять адреса наборами признаков, характеризующих объекты и данные в них, и применять работающие с признаками распределенные ассоциативные операции. Каждая ассоциативная операция при выполнении управления и вычислений одновременно обрабатывает множество признаков-аргументов операции, обычно не требуя дополнительного времени на обработку. Чтобы подчеркнуть одновременное участие в ассоциативной операции множества данных, в ряде мест будем такие операции называть групповыми. Неопределенность в действиях активных объектов требует, чтобы структура связей между ними быстро изменялась в ответ на возникающие события. Иначе говоря, система должна быть адаптивной, с переменной структурой, что обеспечивается в статье. Технические детали структур таких систем подробно изложены в статье [1], ориентированной на разработчиков технических средств.

Новая статья основана на результатах этой статьи, но она ориентирована на создателей алгоритмов и программ, используемых в условиях неопределенности. Технические детали в ней сведены к минимуму и ограничены §2, иллюстрирующим реализацию предлагаемых операций в быстро перестраиваемой структуре связей объектов.

Приведенные в статье решения имеют следующие основные особенности.

- (1) Применена группа способов взаимодействия активных объектов в стационарных и мобильных системах, позволяющая объектам быстро реагировать на непредвиденные события.
- (2) Управление взаимодействием объектов полностью децентрализовано с переходом из асинхронного состояния в синхронное посылкой указанными объектами единственного синхросигнала (однотактная синхронизация).
- (3) Средства сетевого взаимодействия либо полностью расположены в объектах, либо вне объектов состоят из регрессирующих сигналов, не содержащих логические элементы.
- (4) Применены распределенные ассоциативные операции, которые без внешнего вмешательства инициируют объединение автономных объектов в синхронно действующую систему объектов. Каждая операция одновременно обрабатывает множество данных, поступающих от участвующих в операции объектов.

В совокупности §1–§4 решают общую задачу: переводят распределенные асинхронно действующие цифровые объекты в единую синхронную сетевую структуру со скоростью реакции на непредвиденные события близкой к скорости реакции в сосредоточенных системах.

Для оценки новизны предлагаемых решений интеллектуальный поиск публикаций в Интернете выделил за последние 5 лет 16 публикаций, которые можно разделить на две группы.

В статьях [2, 3, 12, 14] приведены решения, влияющие на организацию технических средств системы. Из них общий интерес представляют работы [2, 3], использующие FPGA, которые позволяют перестраивать структуру связей в системе при переходе от одной решаемой задачи к другой. Известно много других публикаций с применением FPGA для тех же целей. Отметим, что в предлагаемой статье перестройка возможна в ходе решения конкретной задачи выполнением единственной команды.

В остальных найденных ссылках [4–7, 9–11, 13] рассмотрены алгоритмические и программные решения как для конкретных распределенных систем, так и для класса распределенных систем в целом.

Отметим распределенные системы с участием ИИ [5, 6]. Здесь выполняется взаимодействие центральной базы данных ИИ с удаленными распределенными потребителями ИИ. Со временем, с расширением технических возможностей удаленных потребителей такая связь дополнится связью между распределенными системами ИИ. В [4] дан обзор обеспечения отказоустойчивости распределенных систем. В [7, 9, 13] рассмотрено общее текущее состояние разработок распределенных систем и будущие тенденции.

В [8] рассмотрено применение DataFlow для связи удаленных Edge-систем со средствами облачных вычислений. Статьи [10, 11] выделяются среди остальных статей тем, что рассматривают распределенные системы управления, как правило, требующие соблюдения требований жесткого реального времени. Статья [13] рассматривают состояние исследований в области киберфизических систем широкого вида. В [10] рассмотрены исследования сверхбольших систем управления.

Основное отличие предлагаемой статьи от приведенных в обзоре состоит в разработке широкого набора средств низкого уровня:

- трех структур взаимодействия объектов,
- скоростных операций и необходимого распределенных ассоциативных операций,
- необходимого для них специального представления данных (логических шкал).

В статье показана эффективность этих средств на конкретных частных примерах. Их полезность на реальных больших задачах нуждается в отдельных оценках. Не рассмотрено создание на основе приведенных базовых ассоциативных операций новых макроопераций, ориентированных на конкретные области приложений.

В новой статье по сравнению со статьей [1] материал по скоростным ассоциативным операциям и алгоритмам обработки данных и управления частично переработан, добавлены новые решения, удалены второстепенные. Удалены технические подробности, не представляющие интерес для разработчиков приложений. Разделы §1, §7–§9 полностью содержат новый материал. Переработано введение.

Состав статьи следующий. В §1 представлены требования к взаимодействию распределенных объектов и области их применения. В §2 приведены три вида структур взаимодействующих объектов с разной скоростью реакции на события. В §3 описан децентрализованный переход объектов в синхронное состояние, а в §4 – синхронизация действий группы объектов-приемников. В §5 показан запуск периодической синхронизации однопакетной синхронизацией. Приведенные в §6 логические шкалы и распределенные групповые операции работают одновременно с множеством объектов данных. Управлению взаимодействием объектов, находящихся в разных кластерах посвящен §7. В §8 рассмотрено применение сетевых карт для синхронного скоростного взаимодействия объектов, а в §9 – влияние безадресных операций на свойства распределенных алгоритмов.

§1. Требования к взаимодействию распределенных объектов, области их применения

1.1. Современные требования к взаимодействию распределенных управляющих и вычислительных цифровых средств

- (1) Существенно расширяется область задач, решаемых системами взаимодействующих активных объектов, для которых получение решения задачи позже заданного граничного момента времени недопустимо (режим ЖРВ) или существенно ухудшает возможности системы. Объекты должны быть стационарными и мобильными.
- (2) Группам объектов из п. 1 приходится реагировать на непредвиденные события с заранее неизвестными местом, временем появления, спецификой событий и общей для всей группы, реакцией на события, формируемой в динамике. Отсюда, граничный момент и длительность процесса совместного решения исходно неизвестны и определяются в динамике.
- (3) Реагировать на события вся система и входящие в нее группы объектов должны как на единое целое, выполнять и завершать совместные действия одновременно или с заданными временными задержками между объектами. Так как в п. 2 неизвестны адреса обнаруживших событие объектов, то объектам необходимо взаимодействовать ассоциативно, используя только характерные для объектов наборы признаков.
- (4) Для ускорения реакции на события сведения о событии, обнаруженном любым объектом, должны как можно быстрее быть доставлены всем объектам. Сформированная группой объектов реакция на события также должна быть быстро доставлена всем объектам-исполнителям реакции.
- (5) Для обеспечения быстрой реакции на события в условиях пп. 1–4 необходимо организовать синхронное взаимодействие объектов, идентифицируя группу не перечислением адресов объектов группы, а в целом, совокупностью признаков, выделяющих конкретную группу среди других групп. Поэтому необходимо в отдельных действиях оперировать с множествами объектов как с единичными объектами. Как отмечено выше, операции поиска и обработки данных с одновременным синхронным участием в операции множеств распределенных в системе объектов и данных далее будем называть групповыми или ассоциативными операциями, подчеркивая, что в операции одновременно участвует группа объектов, характеризующаяся не их расположением (адресами), а совокупностью признаков-ассоциаций.
- (6) Одновременное и быстрое манипулирование множествами сообщений, посылаемых группами распределенных объектов, увеличивает нагрузку на связывающие объекты сетевые средства и выдвигает к ним новое требование: на вход любого объекта должны одновременно

поступать сообщения, синхронно посылаемые произвольной группой объектов.

Для этого поступающие сообщения должны быть объединены в единое сообщение, содержащее сообщения групп объектов поочередно или как одно короткое сообщение с совмещением одноименных разрядов группы сообщений. Для ускорения взаимодействия групп объектов желательно совмещать передачу сообщений с обработкой данных в сообщениях непосредственно сетевыми средствами, без задержки процесса передачи.

Требования 1–6 характерны для многих видов задач, решаемых в ЖРВ, и многих видов задач, не имеющих жестких временных ограничений, но с большой неопределенностью в выборе последовательности шагов решения. Соответствующие примеры даны в 1.2.

1.2. Основные области, предъявляющие требования из 1.1

Перечисленные в 1.1 требования характерны прежде всего для работающих в ЖРВ распределенных систем управления и систем информирования о состоянии внешней среды. В этих системах, требующих быстрой реакции, каждый следующий шаг действий объекта во многих случаях зависит не только от его состояния, но и от неизвестного для него состояния других объектов. В других областях применения цифровых систем временные ограничения менее значимы, но потребность в решении задач без обращения к адресам объектов или местам хранения данных, также востребована широко.

Среди областей, где полезны ассоциативные решения, отметим многие задачи с применением эвристики, любые задачи с взаимодействием переменного состава объектов или данных, в которых выбор очередного шага решения зависит от событий, возникающих на текущем и предыдущих шагах. Использование адресов существенно замедляет работу системы или даже практически невозможно в задачах с быстро изменяемым и заранее неизвестным составом участников. Ассоциативные взаимодействия ускоряют определение места и времени получения ожидаемых промежуточных результатов вычислений в задачах DataFlow.

Следует особо подчеркнуть отличие предложенных в статье распределенных ассоциативных систем от известных сосредоточенных систем, например, ассоциативного суперкомпьютера STARAN, который выпускался в США в разных модификациях, начиная с 1970 года до конца века [15] и оперативно управлял сложными системами. Распределенным ассоциативным системам приходится интенсивно использовать сетевые средства взаимодействия объектов, которые существенно ограничивают параллельное взаимодействие объектов. Уменьшению сетевого ограничения посвящена значительная часть статьи.

В следующих параграфах приведены основные архитектурные и структурные решения, ориентированные на удовлетворение требований 1.1 с ориентацией на области из 1.2

§2. Структуры синхронного взаимодействия распределенных объектов

В §2 рассматриваются три вида структур, позволяющих всем объектам синхронно получать сообщения от всех объектов и выполнять их ассоциативную обработку. Виды структур имеют разную точность синхронизации, пропускную способность, скорость перестройки всех связей в системе, сложность реализации. Как следствие, эти структуры ориентированы на использование в системах, существенно отличающихся в требованиях к приведенным параметрам. Все представленные виды структур применимы для систем, к которым можно обращаться ассоциативно. В 2.1 рассматривается структура взаимодействия с наиболее высокими требованиями к скорости взаимодействия объектов.

2.1. Неоднородная структура взаимодействия с группой ретрансляторов

Организация структуры дана на рисунке 1. Структура ориентирована

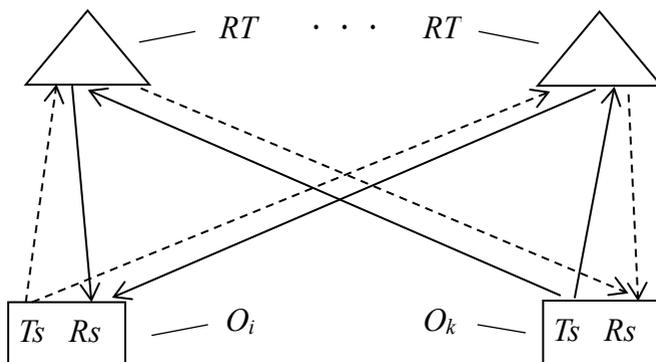


РИСУНОК 1. Структура с максимальными требованиями к параметрам взаимодействия

на связи объектов, компьютеры которых работают в жестких временных ограничениях и выполняют задачи с большим уровнем неопределенности. Объекты обмениваются оптическими сигналами по беспроводным каналам, оптоволоконным и смешанным. Структура неоднородная: объекты взаимодействуют через ретрансляторы сигналов. Все устройства структуры стационарны. На рисунке 1 O_i и O_k два произвольных объекта из N

объектов структуры, RT – группа ретрансляторов оптических сигналов, посылаемых объектами.

Каждый объект может выполнять совместно функции источника и приемника сообщений, передаваемых оптическими сигналами. Объекты имеют передатчик Ts и приемник Rs оптических сигналов. Если объект действует как источник, то его Ts передает в RT сигналы сообщения. Если объект действует как приемник, то его Ts передает в RT непрерывный сигнал, который RT возвратит в Rs модулированным сигналами источника, посланного сообщение в RT .

В каждом объекте есть коммутатор сигналов, переключающий связи источника и приемника объекта с любыми ретрансляторами из группы RT . Для выполнения указанных действий ретранслятор содержит ретрорефлектор (катафот), который поступающий в него сигнал от любого объекта возвращает этому объекту. Для связи группа источников и приемников направляет сигналы в выбранный ими ретранслятор по очереди или совмещая одноименные разряды.

Источник посылает сообщение, передавая единицу сигналом на частоте f_1 , ноль – на частоте f_0 . Приемник для приема сообщения источника посылает в ретранслятор непрерывный сигнал на частоте $*f_1$ для приема единицы и $*f_0$ для приема ноля. Ретранслятор принимает сигналы источника, переводит их в электрические сигналы и модулирует ими соответственно непрерывные сигналы приемника $*f_1$, и $*f_0$, отправляя этим сообщение приемнику. Разница в частотах f и $*f$ позволяет объектам принимать сигналы только от ретрансляторов. Энергии сигналов источника должно быть достаточно только для приема их ретранслятором, приемники получают сигналы, затрачивая свою энергию. Если энергия источника достаточна для действий модулятора в ретрансляторе, то последний может быть пассивным устройством.

Важные функциональные возможности:

- (1) Произвольные группы объектов-источников могут синхронно см. §3 послать отдельные сигналы и сообщения в ретранслятор RT – вторичный источник для всех приемников, следящих за данным RT . Точность синхронизации зависит только от ошибки технических средств, измеряющих расстояние между объектом, наилучший результат с ошибкой порядка пикосекунды предоставлен в [16, 17].
- (2) Синхронизация сообщений не зависит от расположения объектов и RT и позволяет создать ряд быстрых распределенных групповых операций, которые выполняются группой объектов за время, не зависящее от количества участвующих в операции объектов.
- (3) Синхронная сборка сообщений группы объектов в общий RT , объединенная с приведенными ниже групповыми операциями,

существенно ускоряет поиск данных в больших массивах и их обработку.

- (4) Структура позволяет экономно бороться с отказами ретрансляторов RT – единственных расположенных вне объектов сетевых объектов. Для замены m одновременно отказавших произвольных RT достаточно иметь только m резервных RT .
- (5) Соединения в структуре на рисунке 1 (и структурах из 2.2 и 2.3) позволяют на разных участках использовать беспроводные и волоконно-оптические связи. Такая реализация позволяет располагать беспроводные каналы в наиболее благоприятных условиях. В компьютерных центрах беспроводная исполнительная часть структуры на рисунке 1 может располагаться в одном или нескольких герметичных центрах коммутации (содержащих коммутаторы, источники, приемники сигналов и ретрансляторы). Управляющая средствами коммутации часть средств будет при этом находиться в объектах и соединяться с коммутирующей частью волоконно-оптическими каналами связи.
- (6) Отметим, в структуре оптические ретрансляторы с ретрорефлекторами и коммутаторы применяются в ряде областей, например, в средствах связи [18, 19], но не применяются в компьютерах. Далее используются и другие известные технические средства, также отсутствующие в компьютерах. Технические детали даны в [1].

2.2. Простая неоднородная структура с одним ретранслятором

По сравнению со структурой 2.1 структура на рисунке 2 предназначена для групп объектов, которым достаточна пропускная способность одного

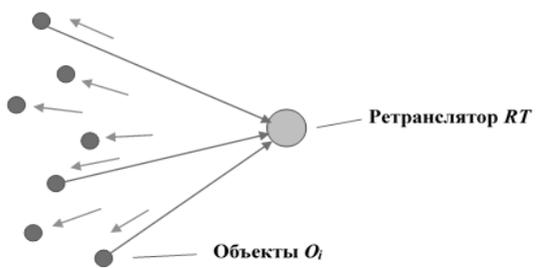


Рисунок 2. Неоднородная структура с одним ретранслятором

RT . Для обмена сообщениями могут использоваться оптические и радиосигналы. Наличие одного RT привело к удалению коммутаторов, необходимых в структуре на рисунке 1.

При обмене оптическими сигналами RT может использовать ретро-рефлектор, но доступно и обычное сочетание: RT принимает сигналы

объекта-источника, переводит их в электросигналы, которые действуют на генератор оптических или радиосигналов, направляемых объектам-приемникам.

Как и на рисунке 1 частоты сигналов объектов-источников и приемников различны для устранения помех приемникам от сигналов передатчиков. Наличие единственного ретранслятора позволяет источникам использовать ненаправленные сигналы для одновременной рассылки сообщений группе объектов-приемников.

2.3. Однородная структура взаимодействия объектов без ретрансляторов

В однородной системе ретранслятор отсутствует и объекты взаимодействуют непосредственно, могут быть стационарными и мобильными. Здесь существенно изменилось понятие одновременности событий по сравнению с неоднородными структурами.

В разделах 2.1 и 2.2 события считаются одновременными, если современные технологии не позволяют различить момент их появления, что означает ошибку порядка пикосекунды [17]. Для однородных систем события одновременны, если они появляются в пределах интервала времени T , равного времени переноса сигнала между наиболее взаимно удаленными взаимодействующими объектами. Разброс во времени начала синхронизации объектами достигает $T = L/c$, где L – максимальное расстояние между объектами, c – скорость света. Такая неопределенность мало влияет на обычный обмен длинными сообщениями, но значительно замедляет выполнение коротких распределенных операций в скоростных системах.

Однако при уменьшении размеров системы однородная структура взаимодействия в многих приложениях заменяет структуры из 2.1 и 2.2. Общий итог §2: в приведенных трех структурах передаваемая между любыми объектами информация одновременно поступает к группе объектов, непосредственно или через ретранслятор. Сообщения, посылаемые по конкретным адресам, видны всем объектам и ассоциативно могут вызвать незапланированные, но необходимые действия групп объектов. Сообщения могут быть безадресными, направленными объектам, имеющим только характерные наборы признаков.

§3. Децентрализация входа объектов-источников в синхронное состояние

В разделе §3 децентрализация исключает привлечение внешнего или внутреннего центра управления к выполнению перехода группы объектов в синхронное состояние.

3.1. Начальный синхронный вход объектов в структуру связи

Вход в процесс синхронизации одинаков для всех структур из §2.

Объекты, не передающие в текущий момент сигналы в структуру взаимодействия, для получения права синхронного доступа к структуре выполняют следующие действия, которые иллюстрирует рисунок 3.



Рисунок 3. Принцип децентрализованного входа в синхронизм

- (1) На рисунке 3 три произвольных объекта из группы n объектов ожидают момент отсутствия сигналов в сети (паузу) и посылают в сеть сигнал s заявки на вход в синхронизм. Сигнал s поступает в сеть любым из доступных в сети способов. Длительность сигнала s выбирается не меньше времени T перемещения сигнала между наиболее взаимно удаленными взаимодействующими объектами. В неоднородной сети сигнал отправляется в RT , в однородной сети непосредственно всем объектам. При наличии RT интервал T может быть заменен на интервал Trt времени перемещения сигнала между RT и наиболее удаленными от него объектами.
- (2) Сигналы s объектов в неоднородных и однородных системах объединятся с частичным совмещением в общий единственный (при выполнении указанного условия для длительности T) сигнал S , поступающий ко всем объектам. Момент завершения в объекте сигнала S формирует сигнал $*S$, который служит для объекта сигналом начала входа в синхронизм. Приход к объекту сигнала $*S$ начинается в нем процесс однократной синхронизации, как показано в 3.2

3.2. Синхронизация действий группы объектов-источников

Раздел 3.2 состоит из трех разделов. В разделе 3.2.1 рассмотрены варианты определения удаленности объектов-источников от RT . В разделе 3.2.2 – синхронная доставка сообщений группы объектов-источников в RT . В разделе 3.2.3 – корректировка источниками группы передаваемых ими сообщений.

3.2.1. Определение удаленности объектов от RT

Рисунок 4 иллюстрирует цель процесса синхронизации действий

объектов O_1 – O_3 относительно ретранслятора RT .

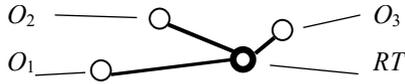


Рисунок 4. Синхронизация объектов-источников

Здесь объекты O_i расположены произвольно относительно RT , и требуется определить их удаленность от RT , которая позволит объектам посылать сигналы в RT одновременно или с заданным относительным сдвигом во времени. Наиболее точная синхронизация достигается при применении стационарных объектов, RT и поочередного измерения удаленности каждого объекта от RT . Наибольшую точность дает измерение удаленности с применением средств измерения стандарта IEEE 1588-2019 [16]. Измерения достаточно выполнять редко, только с учетом изменения влияния внешней среды на скорость переноса сигналов.

Поочередное измерение предполагает наличие упорядоченности объектов. При наличии адресов объекты упорядочат действия с учетом старшинства значений адресов (6.2). При отсутствии адресов объекты используют наборы качественных признаков, характерных для текущего состояния объекта. В этом случае может не выделиться единственный объект и действия продолжатся не с отдельным объектом, а с множеством объектов. Добавка к набору признаков случайного числа (вместо адреса) позволит выделить один объект.

ЗАМЕЧАНИЕ 1. Ретрансляторы, используемые объектами для взаимодействия, в отличие от объектов, не имеют индивидуальных отличительных наборов признаков, но порядковые номера им присвоить можно. Поэтому во многих случаях поиск группы взаимодействующих объектов можно существенно ускорить, сведя его к поиску используемого объектами общего RT (см. 6.3).

ЗАМЕЧАНИЕ 2. В статье объекты и RT не используют часы с одинаковыми показаниями, установленными внешними средствами. Пусть такая возможность есть, объекты неподвижны и RT , отправляя сигнал $*S$ начала синхронизации, сопровождает его значением своего значения отсчета времени. Тогда при получении сигнала все объекты, вычитая из своего отсчета времени отсчет, полученный от RT , одновременно определяют свою удаленность от RT .

3.2.2. Синхронная доставка сообщений группы объектов в RT

Имея значение удаленности от RT , объекты-источники для синхронизации пошлют свои сообщения одно за другим в виде единого сообщения или в виде короткого сообщения, полученного совмещением одноименных разрядов группы сообщений. Для этого каждый объект O_i в момент получения сигнала $*S$ от RT посылает в RT сообщение с задержкой D_i , где $D_i = 2(T^* - T_i) + \alpha$, T^* — интервал времени, не меньше времени доставки сигнала в RT самым удаленным от RT объектам, T_i — время передачи сигнала от O_i к RT . Эти действия превращают произвольно отстоящие от RT объекты в равноудаленные. Варьируя α , объект при необходимости вносит дополнительный сдвиг прихода сигнала в RT . В результате объекты примут одно сообщение RT , состоящее из упорядоченных во времени сообщений объектов, или при $\alpha = 0$ примут одно короткое сообщение с совмещением одноименных разрядов в сообщениях объектов.

3.2.3. Корректировка объектами одновременно передаваемых сообщений

Применение α позволяет объектам также вносить изменения в сообщения других объектов, которые передаются синхронно в виде единого сообщения. Пусть до начала единого сообщения участвующим в его создании объектам известна структура сообщений объектов. Каждому объекту также известны позиции в едином сообщении, в которые объекту разрешено бесконфликтно вносить изменения. При выполнении всех указанных условий объекты смогут изменять общее сообщение, причем объекты, наблюдающие за ретранслирующим сообщением RT , получают информацию о всех групповых изменениях. Результат: объекты сэкономят время группового изменения, за счет исключения формирования группы адресных сообщений объектами. К способам перехода в синхронное состояние относится также оповещение системы объектов о завершении группы асинхронных действий (см. 3.2.4).

3.2.4. Оповещение о завершении группой объектов асинхронных действий

Пусть группа Oa объектов выполняет действия, завершаемые объектами в неизвестные заранее моменты времени, и требуется как можно быстрее оповестить все объекты о моменте завершения действий всеми объектами. Для этого каждый объект-источник из Oa , начиная действие, посылает в RT приемников сигнал u , подобно сигналу s (см. 3.1), и снимает его в момент завершения действия. Сигналы u поступают к объектам-приемникам в виде совмещенного сигнала U , подобного S .

В момент снятия последнего сигнала u появляется сигнал $*U$, аналогичный $*S$, который информирует объекты о завершении всех асинхронных действий. Приведенные действия наиболее существенны при решении задач 1.2, имеющих большую неопределенность в момент получения решения. Действия выполняются быстро, так как не требуют привлечения центра управления, и задержка в поступлении ко всем объектам сигнала $*U$ равна T или $2Trt$. Оповещение в 3.2.4 действует в распределенных структурах из §3 как старт-стопный режим взаимодействия объектов, подобный барьерной синхронизации: приход к объекту сигнала U – старт режима, приход сигнала $*U$ – его завершение.

Дополнение. Из троек вида «группа объектов – RT – группа объектов» могут создаваться цепочки, по которым будут перемещаться результаты обработки сообщений. Все перемещения могут выполняться через равные промежутки времени. Это неэкономно, если разные объекты выполняют свою часть обработки за разные интервалы времени. Применение сигналов $*U$ учтет реальные задержки.

§4. Синхронизация действий группы объектов-приемников

Приемники сигналов RT определяют свою удаленность во времени от RT , вычисляя D_i^* аналогично D_i , что позволяет им, получив сообщение источников, выполнять требуемые действия одновременно.

Каждый приемник с номером i предварительно определяет задержку $D_i = 2(T^* - T_i) + \alpha$. Получив в сообщении команду, каждый приемник O_i группы сделает задержку D_i^* и выполнит команду при $\alpha = 0$ одновременно или с разными задержками при задании объектам разных значений α .

Выполнение команды может заключаться в непосредственном воздействии на исполнительное устройство или в запуске через любой RT системы очередного синхронного этапа решения задачи. Здесь этап представляет собой не продвижение по обычному конвейеру с заранее известным путем обработки «изделия»: система быстро перестраивается, объекты могут ожидать установления новых связей (см. 6.3), и на следующем этапе обработка данных будет выполняться неизвестными заранее объектами.

§5. Запуск периодической синхронизации одноктной синхронизацией

Одноктная синхронизация (ОтС) позволяла не взаимодействующим между собой в некотором промежутке времени объектам перейти в режим

синхронного взаимодействия при получении сигнала $*S$ или $*U$ и, в частности, может инициировать запуск или внесение изменений в периодическую синхронизацию (ПС) действий объектов.

Взаимодействие ОтС и ПС полезно, если ПС достаточно поддерживать в течение короткого промежутка времени, от долей секунды до нескольких минут. При участии во взаимодействии компьютеров за такие интервалы может выполняться большое количество операций. По этой причине и при большой неопределенности в событиях часто не требуется длительная поддержка ПС, и достаточно использовать следующие действия ОтС.

Если объекты не синхронизированы, то, получив сигнал $*S$ или $*U$, готовые к синхронной работе объекты немедленно начнут синхронное взаимодействие. Их генераторы синхросигналов включатся синхронно по фазе и будут действовать до завершения заданного интервала выполнения ПС. Если есть объекты, не готовые к ПС, то они должны создать сигналы типа U и $*U$, что начнет переход всех объектов в ПС.

Если в системе есть объекты, находящиеся в режиме ПС, то они могут своими сигналами U задержать вход новых объектов в ПС. Альтернатива: при высоком приоритете запроса на вход в ПС он будет разрешен, если находящиеся в ПС объекты могут приостановить свои текущие действия.

Дополнительно к показанному действия объектов в интервале ПС могут быть дополнительно ускорены следующим образом. Интервал действия ПС разделяется на фрагменты, и перед переходом в следующий фрагмент один или несколько объектов (при предварительной согласованности в выполняемой программе) немедленно, без сигнала $*U$ имеют право выполнить вне очереди свои действия.

В такие моменты объекты, не затрачивая зависящее от размеров системы время на однократную синхронизацию, проводят совместные действия, например, групповую операцию.

Дальнейшее управление действиями объектов в синхронном состоянии требует использовать логические шкалы, рассмотренные в §6.

§6. Логические шкалы и распределенные групповые операции

6.1. Структура логической шкалы

Логическая шкала (далее для краткости – шкала) это конструкция, определяющая состояние объекта или находящейся у объекта группы данных и имеющая следующие особенности.

- (1) Шкала представляет собой последовательность двоичных разрядов, количество которых зависит от требований конкретной операции над сообщениями объектов.

- (2) При передаче шкал между объектами единица и ноль представлены активными сигналами разных частот. Будем единицу передавать сигналом с частотой f_1 , ноль – сигналом с частотой f_0 . При одновременной передаче ноля и единицы появляется смесь сигналов f_1 и f_0 . При необходимости сохранения объектом такой смеси активные единица и ноль кодируются парой битов 10 и 01 соответственно. Иначе будет сохранен один бит с единицей только для f_1 .
- (3) Разряды шкалы разделяются на блоки из различного, заранее известного количества разрядов. Семантика таких блоков также известна заранее. В простейшем случае каждый блок состоит из одного двоичного разряда, выделанного одному из логических параметров, характерных для сообщения. В этот разряд заносится единица, если параметр присутствует в сообщении и заносится ноль в противном случае.

Параметры также могут иметь численное значение. Каждой цифре числа в шкале выделяется блок с количеством двоичных разрядов, соответствующим основанию используемой системы счисления. Каждая такая шкала содержит единицу только в разряде, соответствующем значению цифры, остальные ноли. Например, цифра семь в десятичной системе счисления будет представлена шкалой 001000000.

- (4) Из п. 3 следует, что в зависимости от конкретной области применения шкал в них может быть внесена дополнительная информация, оперативно воздействующая на объект или сообщающая о его состоянии.

Покажем использование шкал в распределенных вычислительных и управляющих групповых операциях, используя результаты [1].

6.2. Распределенные групповые вычислительные операции

Пусть разряды шкалы соответствуют логическим признакам, характеризующим состояние объектов. Если признак есть в объекте, то в соответствующий ему разряд вносится единица, иначе – ноль. Шкалы объектов синхронно поступают в RT . Общая шкала из RT поступает к приемникам, которые выполняют над ней логические групповые операции.

Если объект выполняет поразрядную операцию AND, то для конкретного разряда общей шкалы его значение равно единице только если разряд содержит f_1 и не содержит f_0 , иначе результат – ноль; для OR значение операции равно нолю, если разряд содержит только f_0 , иначе – единица; для XOR значение операции равно единице, если разряд содержит f_1 и f_0 , иначе – ноль.

Для поиска в массиве чисел максимального и минимального числа введем представление цифр в системе счисления с основанием больше двух: цифру представим шкалой с количеством двоичных разрядов, равным основанию системы. В ней содержится единица в разряде, равном значению цифры, и ноли в остальных разрядах. Например, в десятичной системе совмещение шкал для цифр 7, 5, 4, 2 даст шкалу 001011010. Из нее видно, что в совмещении участвовали наибольшая цифра 7 и наименьшая 2. Это ускоряет поиск \max и \min по сравнению с поиском для цифр в двоичной системе, так как потребуются пересылать меньшее количество шкал.

Именно количество шкал, а не их разрядность замедляет операции. Каждую шкалу можно передать на высокой скорости, но до начала передачи следующей цифры чисел объекты должны получить общую шкалу для предыдущей цифры и определить, будут ли они участвовать в процессе дальнейшего определения \max или \min .

Рассмотрим операции суммирования, требующие одновременного суммирования большого количества цифр чисел [1]. Применим аналого-цифровые операции суммирования и вычитания к цифрам чисел, представленных шкалами. Добавим в RT аналого-цифровой преобразователь (АЦП). Он переведет уровень энергии одновременно приходящих в RT сигналов в число. Способы создания стабильных по энергии сигналов известны [20].

Рассмотрим пример:

ПРИМЕР 1. В RT приходит шкала группы цифр одного из разрядов десятичных чисел, принятых от объектов – 010110010. Шкала показывает наличие цифр 8, 6, 5, 2 в разряде. АЦП преобразует шкалу, например, в 030920070. Здесь цифры показывают: АЦП получил 3 сигнала для значения цифры 8, 9 сигналов для цифры 6, 2 для 5 и 7 для 2. Каждый объект одновременно выполнит подсчет $3 \times 8 + 9 \times 6 + 2 \times 5 + 7 \times 2 = 102$. Действуя аналогично с другими разрядами, объект вычислит сумму полученных чисел.

Дополнение: выполнение двух примеров вычислительных операций.

ПРИМЕР 2. Создание гистограмм для определения частоты встречаемости отдельных признаков в наборах признаков группы объектов. Действуя локально объекты определяют количество появлений заданных признаков во всех шкалах объекта (или заданной их части). Затем объекты посылают в RT с совмещением одноименных разрядов полученные шкалы, содержащие оценки количества признаков. RT

содержит АЦП и, как показано на примере выше, преобразует уровень общих сигналов в цифры, позволяющие оценить появление конкретных признаков в группе объектов.

ПРИМЕР 3. Сжатие и расшифровка данных, передаваемых через RT. На рисунке 5 показана передача в RT трех шкал с совмещением одноименных разрядов. Здесь шкалы, передаваемые тремя объектами, объединяются в RT и передаются всем приемникам в виде общей шкалы, которая в приемниках, выполняющих групповые поразрядные операции, расшифровывается различно в зависимости от вида операции AND, OR, XOR.

```

1 1 0 0 0 ... 0 1 0 – шкала первого объекта
1 0 0 1 0 ... 0 1 0 – второго объекта
1 1 0 1 0 ... 0 1 0 – третьего объекта
1 1/0 0 1/0 0 ... 0 1 0 – общая шкала
=====
1 0 0 0 0 ... 0 1 0 результат для группового поразрядного AND
1 1 0 1 0 ... 0 1 0 результат для группового поразрядного OR
0 1 0 1 0 ... 0 0 0 результат для группового поразрядного XOR

```

Рисунок 5. Сжатие и расшифровка групповых логических операций

Приведенные два вида операций не используются в компьютерах, ориентированных на решение классических вычислительных задач, но существенно ускоряют решение новых задач с включением эвристики. Возможно, эти примеры инициируют создание новых групповых операций.

6.3. Распределенные групповые операции управления действиями объектов

Операции управления в основном подобны вычислительным операциям, но содержимое шкал содержит информацию, необходимую для быстрого управления перестройкой структуры связей в системе и процессами взаимодействия объектов.

Примером типичной операции управления служит операция устранения конфликтов доступа при обмене сообщениями объектов, имеющих порядковые номера.

Претендующий на передачу сообщения объект, получив сигнал *S начала синхронизации, создает шкалу с количеством двоичных разрядов, равным количеству взаимодействующих объектов. В шкале в разряд, соответствующий номеру объекта, последний внесет единицу, остальные –

ноли. Шкалы всех претендентов синхронно, с совмещением одноименных разрядов, поступают в RT . Теперь совмещенная, общая для претендентов шкала содержит единицы в разрядах претендентов. Заявки объектов удовлетворяются поочередно с учетом старшинства разрядов, содержащих единицы.

Показанный способ устранения конфликтов позволил посылкой одного сообщения со шкалой упорядочить разрешение конфликтов в группе объектов, имеющих адреса. Но его можно применять более широко, позволяя без адресов любому набору признаков или произвольному сообщению выполнять функции устранения конфликтов, подобно приведенному выше.

При этом нет гарантии выбора единственного объекта и придется перейти к взаимодействию множеств объектов. Уменьшение размеров этого множества может быть достигнуто уточнением набора признаков объектов, возможно с добавкой к ним случайных чисел см. 3.2.1 и затем выбором объекта операциями \max или \min из оставшихся идентичных объектов.

§7. Управление взаимодействием объектов, находящихся в разных кластерах

Рассмотрим следующие виды взаимодействия в неоднородной структуре 2.1.

- (1) Объекты объединены в единственный кластер и решают общую для объектов задачу с учетом диктуемых задач ограничений на взаимодействие. Особого внимания требуют ограничения по изменению общей структуры связей. Кластер имеет центр управления кластером (ЦУК) для разрешения возможных сложных конфликтов при взаимодействии объектов в кластере. Объекты в кластере создают доступ к ЦУК, используя сигналы S и $*S$ (см. §3). ЦУК упорядочит взаимодействие объектов, послав им команду.
- (2) Объекты распределены на группе кластеров и решают общую задачу с взаимодействием объектов из разных кластеров. Система имеет центр управления системой (ЦУС), разрешающий конфликты при взаимодействии объектов в разных кластерах. ЦУК и ЦУС можно заменять в динамике.

Рассмотрим действия объектов с обращением к ЦУС.

Находящийся в одном из кластеров объект для организации взаимодействия с объектами в других кластерах обращается к ЦУК своего кластера, который обратится к ЦУС и последний посылкой команды

упорядочит взаимодействие объектов из разных кластеров. Детали организации взаимодействия могут существенно различаться в зависимости от выполняемой объектами задачи.

Отметим, в соответствии с принципом формирования структуры 2.1 ЦУК и ЦУС – могут быть не монолитным устройством, а связкой из RT и объектов.

Доставляемая от ЦУС объектам команда для регулировки их взаимодействия, будет включать: адреса RT , адреса используемых источником и приемником, набор признаков источника и приемника, имя канала барьерной синхронизации и другую служебную информацию. Команда поступит к объектам в разных кластерах. Теперь объекты могут связываться через помещенные в команду адреса RT , предоставленные центром управления.

Рассмотрим безадресное взаимодействие объектов в сложных системах. Практически во всех разделах статьи рассматриваются действия объектов, у которых адреса не используются или отсутствуют. Адреса заменяются наборами признаков, выделяющих объект или группу объектов с применением ассоциативного поиска. Поиск данных по наборам признаков вместо сканирования адресов существенно ускоряет не только сам поиск, но и обработку данных ассоциативными операциями, выполняемыми за время, не зависящее от количества участников операции (весь §6). Ассоциативные операции в сочетании с быстрой перестройкой структуры связей позволяют создавать графы произвольных связей между объектами и перестраивать их с каждой новой командой обработки данных.

Центры управления ЦУК и ЦУС, используя наборы признаков распределенных объектов и данных в них, отбирают данные для совместной обработки. Рассмотрим пример.

ПРИМЕР 4. Пусть объекту A требуется установить взаимодействие с неизвестными ему объектами B из других кластеров. От объектов B объект A требует, чтобы B содержали требуемые A совпадающие наборы признаков или имеющие отличия в них. Кроме этого, A необходимо выяснить, с какими объектами C одновременно взаимодействовали объекты B , т.е. указать те RT , которые B и C используют совместно. Это один из вариантов поиска общности в объектах и данных для дальнейшего их использования в новых сочетаниях. Так как объекты B находятся в разных кластерах, то установить с ними связь непосредственно, без центра управления объект не может. Кроме того, искомые объекты могли изменять используемые ими RT . Поэтому объект A обратится к центру управления. Центр управления по признакам найдет искомые объекты, определит используемые ими

RT, примет решение о предоставлении *A* доступа к *RT* и сообщит адреса *RT* объекту *A*. Далее выявлять детали *A* будет самостоятельно или с участием ЦУС.

Приведенный пример характерен для многих применений ассоциативных операций поиска и обработки данных в распределенных системах.

§8. Требования к сетевым картам при синхронном взаимодействии объектов

Современные компьютеры активно используют интерфейсные (сетевые) карты. В подавляющем большинстве это карты Ethernet, которые действуют как внешние по отношению к компьютеру устройства. В объектах систем, рассмотренных в статье, также желательно иметь сетевые карты, не влияющие на структуру объекта. Для обеспечения указанных в §1 новых требований к взаимодействию объектов распределенных цифровых систем в предъявляемые к картам требования должны входить следующие.

- (1) Протоколы, регулирующие взаимодействие карт между собой, должны быть детерминированными.
- (2) В передаваемые между картами сообщения должны входить короткие сообщения, типичные для обменов в пределах компьютера.
- (3) Скорость взаимодействия объектов должна быть близка к внутренней скорости взаимодействия в компьютерах объединяемых объектов – несмотря на распределенность система должна действовать как единый компьютер.
- (4) Должны поддерживаться распределенные ассоциативные операции.
- (5) Сетевая карта должна восприниматься компьютером, как один из специализированных процессоров, и скорость ее взаимодействия с компьютером не должна быть ниже скорости взаимодействия между собой других процессоров компьютера.
- (6) При многоэтапной обработке данных каждый следующий этап должен начинаться как можно быстрее после завершения действий текущего этапа. В 3.2.4 для этого каждый объект, начав действия на текущем этапе, посылает в отдельный канал непрерывный сигнал U , выключаемый в момент завершения объектом работы (фиксируется объектами как сигнал $*U$). Получив сигнал $*U$, объекты могут начинать следующий этап. Если объекты могут для всех этапов указать одинаковую длительность их выполнения, не потребуется использовать сигналы U и $*U$.

Все указанные действия должна отслеживать сетевая карта и посылать в компьютер объекта сигнал о возможности начала следующего этапа.

При наличии в карте памяти, объект ускорит действия, предварительно загрузив в память данные для следующего этапа, которые карта передаст самостоятельно.

В общем случае для очередного этапа данные в любой ретранслятор поставляет не отдельный объект, а группа объектов-источников и получает группа объектов-приемников.

- (7) Пункты 1–6 относятся к взаимодействию объектов, расположенных близко между собой. Но есть потребность наиболее быстро по выделенным каналам объединять объекты, расположенные в пределах здания, между вычислительным центром и организациями, потребителями его услуг. В этих структурах связей допустимо иметь меньшую скорость реакции на события, но более быструю, чем обеспечивает Интернет. Одно из возможных решений – взаимодействие также осуществляют сетевые карты, сохраняя логику близкой связи, но, используя отдельные каналы, позволяющие не замедлять локальные связи.

§9. Влияние безадресных операций на свойства распределенных алгоритмов

Все результаты статьи имеют конечную цель: предложить разработчикам алгоритмов и программ архитектурные и структурные решения в области компьютерных систем, ускоряющие решение новых распределенных задач, приведенных в 1.1. Выделим результаты, наиболее соответствующие этой цели.

- (1) Достигнута высокая скорость реакции систем на непредвиденные события – основная цель действий в режиме ЖРВ. В других системах из 1.1, работающих безадресно, быстродействие также повышается.
- (2) Распределенные объекты активны и работают синхронно: каждый объект имеет право влиять на работу других активных объектов, включая совместный переход в синхронное состояние. Синхронное состояние исключает многие временные задержки, в частности, при устранении конфликтов между объектами.
- (3) В статье структуры связи отличаются от известных возможностью одновременно подключать каналы группы объектов-источников сообщений к одному общему входу приемника сообщений, что необходимо для выполнения распределенных безадресных ассоциативных операций как основы борьбы с неопределенностью.

- (4) Доступны быстрая оценка состояния всей совокупности объектов системы и перевод ее в новое состояние. Обработка данных выполняется распределенной ассоциативной операцией без задержки передачи данных.
- (5) Достижима перестройка всей структуры связей между объектами в пределах одной ассоциативной операции. Перестройка выполняется в микросекундном и наносекундном диапазоне, в зависимости от взаимной удаленности объектов системы.
- (6) Совмещение обычной передачи данных с синхронной обработкой данных, выполняемой распределенной ассоциативной операцией и быстрой перестройкой структуры связей коренным образом повышают для новых задач пропускную способность средств взаимодействия объектов по сравнению с обычными системными сетями (как следует из закона Амдала).
- (7) В совокупности результаты статьи превращают обычную системную сетевую структуру объединения компьютеров в распределенный компьютер с быстрыми безадресными ассоциативными операциями, одновременно обрабатывающими множество данных, поступающих от множества объектов.

Ключевые для п.п. 1–7 решения статьи:

- адаптивные структуры со скоростной реорганизацией;
- передача парафазных сигналов при взаимодействии объектов;
- распределенные ассоциативные операции, одновременно работающие с множествами объектов и данных в них;
- логические шкалы для представления данных, обрабатываемых ассоциативными операциями.

Предложенные решения могут быть полезны для ускорения решения известных прикладных задач и для новых задач, порождаемых появлением адаптивных распределенных цифровых систем, требующих быстрой реакции на непредвиденные события.

Заключение

Результаты статьи позволяют выполнять быстрые ассоциативные поиск и обработку данных, размещенных в распределенных цифровых объектах.

Рассмотрены три вида структур взаимодействия объектов с различными требованиями к скорости реакции объектов на возникающие события, точности синхронизации совместных действий объектов и сложности реализации (см. §3).

Первая из трех структур, позволяет перестроить все связи между объектами системы посылкой объектам единственной команды. Перестройка выполняется в пределах микросекундного и субмикросекундного интервалов времени в зависимости от взаимной удаленности объектов.

Системы могут управляться полностью децентрализованно, согласованием между объектами совместных действий.

Приведены быстрые ассоциативные операции, каждая из которых обрабатывает множество данных, поступающих одновременно от распределенных объектов.

Значительная часть видов обработки данных выполняется непосредственно в сетевых средствах системы, не задерживая передачу данных между объектами.

Предлагаемые решения могут оказать влияние на конструкции алгоритмов, ориентированных на ассоциативную обработку данных, так как в решениях статьи в процессе передачи сообщений без задержки передачи выполняется параллельная обработка данных в одновременно посылаемых многими источниками сообщениях, что существенно ограничивает отрицательное влияние закона Амдала.

Допускается в пределах системы разделение объектов системы на кластеры, каждый из них решает свою часть общей задачи асинхронно, но в неизвестные заранее моменты времени кластеры взаимодействуют для продвижения в совместном решении.

Для такого взаимодействия кластеров использовано два вида центров управления: центра управления кластером объектов (ЦУК) и центра управления системой кластеров (ЦУС). Кластеры для выполнения взаимодействия обращаются к своим ЦУК, которые обращаются к ЦУС. Последний согласует действия ЦУК, согласующих действия объектов кластера. После этого несколько кластеров взаимодействуют между собой до момента нового обращения к ЦУС или вмешательства ЦУС.

Во многих применениях можно получить указанные результаты, не вмешиваясь в структуру компьютера, только на уровне специальных средств распределенного взаимодействия, например, сетевых карт.

Список использованных источников

- [1] Стецюра Г. Г. *Синхронное взаимодействие распределенных неупорядоченных цифровых объектов* // Программные системы: теория и приложения. – 2024. – Т. 15. – № 4 (63). – С. 3–26. doi URL ↑82, 84, 89, 96, 97
- [2] Johnson H., Fang T., Perez-Vicente A., Saniie J. *Reconfigurable distributed FPGA cluster design for deep learning accelerators*. – 2023. – 5 pp. arXiv:2305.18332 ↑83

- [3] Maidan M., Melnyk A. *Organization of FPGA-based devices in distributed systems* // International Journal of Computing.– 2023.– Vol. **22**.– No. 3.– Pp. 352–359.
⁸³
- [4] Hussein A. A., AL-zebari A., Omar N., Merceedi K. J., Ahmed A. J., Salim N. O. M., Hasan Sh. S., Kak Sh. F., Ibrahim I. M., Yasin H. M., Salih A. A. *State of art survey for fault tolerance feasibility in distributed systems* // Asian Journal of Research in Computer Science.– 2021.– Vol. **11**.– No. 4.– Pp. 19–34. ⁸³
- [5] Wei W., Liu L. *Trustworthy distributed AI systems: Robustness, privacy, and governance* // ACM Comput. Surv.– 2025.– Vol. **57**.– No. 6.– id. 143.– 42 pp. ⁸³
- [6] Gill S., Xu M., Ottaviani C., Patros P., Bahsoon R., Shaghaghi A., M. Golec, V. Stankovski, H. Wu, A. Abraham, M. Singh, H. Mehta, S. K. Ghosh, T. Baker, A. K. Parlikad, H. Lutfiyya, S. S. Kanhere, R. Sakellariou, S. Dustdar, O. Rana, I. Brandic, S. Uhlig *AI for next generation computing: Emerging trends and future directions* // Internet of Things.– 2022.– Vol. **19**.– id. 100514.– 43 pp.
 2203.04159 ⁸³
- [7] Dai F., Hossain Md A., Wang Y. *State of the Art in Parallel and Distributed Systems: Emerging Trends and Challenges* // Electronics.– 2025.– Vol. **14**.– No. 4.– id. 677.– 39 pp. ⁸³
- [8] Chini F., De Martini L., Margara A., Cugola G. *FlowUnits: Extending dataflow for the edge-to-cloud computing continuum*.– 2025.– 6 pp. 2504.11400 ⁸⁴
- [9] Lindsay D., Gill S., Smirnova D., Garraghan P. *The evolution of distributed computing systems: from fundamental to new frontiers* // Computing.– 2021.– Vol. **103**.– No. 8.– Pp. 1859–1878. ⁸³
- [10] Pedrosoa L., Batistab P., Heemelsa W. P. M. H. *Distributed design of ultra large-scale control systems: Progress, Challenges, and Prospects* // Annual Reviews in Control.– 2025.– Vol. **59**.– id. 100987.– 22 pp. 2503.23416 ^{83, 84}
- [11] Hamzah M. M., Islam M. M., Hassan S., Akhtar M. N., Ferdous M. J., Jasser M. B., Mohamed A. W. *Distributed control of cyber physical system on various domains: A critical review* // Systems.– 2023.– Vol. **11**.– No. 4.– id. 208.– 31 pp. ^{83, 84}
- [12] Park T., Choi I., Lee M. *Distributed associative memory network with memory refreshing loss* // Neural Networks.– 2021.– Vol. **144**.– Pp. 33–48.
 2007.10637 ⁸³
- [13] Ismail S., Shah K., Reza H., Marsh R., Grant E. *Toward management of uncertainty in self-adaptive software systems: IoT case study* // Computers.– 2021.– Vol. **10**.– No. 3.– id. 27.– 19 pp. ^{83, 84}
- [14] Dhilleswararao P., Boppu S., Manikandan M. S., Cenkeramaddi L. R. *Efficient hardware architectures for accelerating deep neural networks: Survey* // IEEE Access.– 2022.– Vol. **10**.– Pp. 131788–131828. ⁸³
- [15] Фостер К. *Ассоциативные параллельные процессоры*.– М.: Энергоиздат.– 1981.– 240 с.; Foster C. C. *Content Addressable Parallel Processors*.– Van Nostrand Reinhold.– 1976.– ISBN 978-0442224332.– 233 pp. ⁸⁶
- [16] *IEEE 1588-2019 IEEE Standard for a Precision Clock Synchronization Protocol for Networked Measurement and Control Systems*.– IEEE Instrumentation and Measurement Society.– 2020. ^{88, 92}

- [17] Dierikx E. F., Wallin A. E., Fordell T., Myyry J., Koponen P., Merimaa M., Pinkert T. J., Koelemeij J. C. J., Peek H. Z. *White Rabbit Precision Time Protocol on Long-Distance Fiber Links* // IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control.– 2016.– Vol. **63**.– No. 7.– Pp. 945–952. doi ↑88, 90
- [18] Rabinovich W. S., Goetz P. G., Mahon R., Swingen L. A., Murphy J. L., M. Ferraro, H. R. Burris, Ch. I. Moore, M. R. Suite, G. C. Gilbreath, S. C. Binari, D. J. Klotzkin *45-Mbit/s cat's-eye modulating retroreflectors* // Optical Engineering.– 2007.– Vol. **46**.– No. 10.– Pp. 104001–1004008. doi ↑89
- [19] Zhu Y., Wang G. *Research on retro-reflecting modulation in space optical communication system* // IOP Conf. Series: Earth and Environmental Science.– 2018.– Vol. **108**.– No. 3.– id. 032060.– 8 pp. doi ↑89
- [20] Bosiljevac M., Downing J., Babic D. *<100 ppm/K output intensity temperature stability with single-color light-emitting diodes* // Applied Optics.– 2016.– Vol. **55**.– No. 3.– Pp. 9060–9066. doi ↑97

Поступила в редакцию	08.02.2025;
одобрена после рецензирования	02.04.2025;
принята к публикации	12.04.2025;
опубликована онлайн	18.06.2025.

Рекомендовал к публикации

д.ф.-м.н. Н. Н. Ненейвода

Информация об авторе:



Геннадий Георгиевич Стецюра

д. т. н., профессор, гл. н. с. ИПУ РАН. Область интересов: информатика, computing. Основные работы в области распределенных многокомпонентных цифровых систем. Более 150 публикаций и более 20 патентов и авторских свидетельств на изобретения.



0000-0003-4606-4424

e-mail:

Декларация об отсутствии личной заинтересованности: *благополучие автора не зависит от результатов исследования.*

UDC 004.272.45

 10.25209/2079-3316-2025-16-2-81-109

High-speed associative interaction of distributed digital objects operating under uncertainty

Gennady Georgievich **Stetsyura**

V. A. Trapeznikov Institute of Control Sciences of RAS, Moscow, Russia

 gstetsura@mail.ru

Abstract. Associative operations and algorithms utilizing them are considered for high-speed network data processing and management in distributed digital systems with unpredictable actions of system objects, unknown time and location of result appearance within the system, system failures, and unexpected external environment influences. These operations and algorithms enabled the creation of interaction methods (based on qualitative features without using object addresses and data) and distributed associative operations involving a group of objects simultaneously in each operation. The execution of these operations does not increase data transmission time. To accelerate the system's response to unexpected influences, three types of adaptive connection structures with different reaction speeds to external effects are proposed. All connections are restructured simultaneously. (*In Russian*).

Key words and phrases: associative interaction, address-free operations, single-cycle synchronization, distributed digital systems, distributed feedback, active objects, adaptive structures

2020 *Mathematics Subject Classification:* 65Y05; 68Q10

For citation: Gennady G. Stetsyura. *High-speed associative interaction of distributed digital objects operating under uncertainty*. Program Systems: Theory and Applications, 2025, **16**:2(65), pp. 81–109. (*In Russ.*).

https://psta.psiras.ru/read/psta2025_2_81-109.pdf

References

- [1] G. G. Stetsyura. “Synchronous interaction of distributed disordered digital objects”, *Program Systems: Theory and Applications*, **15:4** (63) (2024), pp. 3–26 (in Russian).  
- [2] H. Johnson, T. Fang, A. Perez-Vicente, J. Saniie. *Reconfigurable distributed FPGA cluster design for deep learning accelerators*, 2023, 5 pp.  2305.18332
- [3] M. Maidan, A. Melnyk. “Organization of FPGA-based devices in distributed systems”, *International Journal of Computing*, **22:3** (2023), pp. 352–359. 
- [4] A. A. Hussein, A. AL-zebari, N. Omar, K. J. Merceedi, A. J. Ahmed, N. O. M. Salim, Sh. S. Hasan, Sh. F. Kak, I. M. Ibrahim, H. M. Yasin, A. A. Salih. “State of art survey for fault tolerance feasibility in distributed systems”, *Asian Journal of Research in Computer Science*, **11:4** (2021), pp. 19–34.  
- [5] W. Wei, L. Liu. “Trustworthy distributed AI systems: Robustness, privacy, and governance”, *ACM Comput. Surv.*, **57:6** (2025), id. 143, 42 pp. 
- [6] S. Gill, M. Xu, C. Ottaviani, P. Patros, R. Bahsoon, A. Shaghghi M. Golec, V. Stankovski, H. Wu, A. Abraham, M. Singh, H. Mehta, S. K. Ghosh, T. Baker, A. K. Parlikad, H. Lutfiyya, S. S. Kanhere, R. Sakellariou, S. Dustdar, O. Rana, I. Brandic, S. Uhlig. “AI for next generation computing: Emerging trends and future directions”, *Internet of Things*, **19** (2022), id. 100514, 43 pp.   2203.04159
- [7] F. Dai, Hossain Md A., Y. Wang. “State of the Art in Parallel and Distributed Systems: Emerging Trends and Challenges”, *Electronics*, **14:4** (2025), id. 677, 39 pp. 
- [8] F. Chini, De Martini L., A. Margara, G. Cugola. *FlowUnits: Extending dataflow for the edge-to-cloud computing continuum*, 2025, 6 pp.  2504.11400
- [9] D. Lindsay, S. Gill, D. Smirnova, P. Garraghan. “The evolution of distributed computing systems: from fundamental to new frontiers”, *Computing*, **103:8** (2021), pp. 1859–1878. 
- [10] L. Pedrosoa, P. Batistab, W. P. M. H. Heemelsa. “Distributed design of ultra large-scale control systems: Progress, Challenges, and Prospects”, *Annual Reviews in Control*, **59** (2025), id. 100987, 22 pp.   2503.23416
- [11] M. M. Hamzah, M. M. Islam, S. Hassan, M. N. Akhtar, M. J. Ferdous, M. B. Jasser, A. W. Mohamed. “Distributed control of cyber physical system on various domains: A critical review”, *Systems*, **11:4** (2023), id. 208, 31 pp. 
- [12] T. Park, I. Choi, M. Lee. “Distributed associative memory network with memory refreshing loss”, *Neural Networks*, **144** (2021), pp. 33–48.   2007.10637
- [13] S. Ismail, K. Shah, H. Reza, R. Marsh, E. Grant. “Toward management of uncertainty in self-adaptive software systems: IoT case study”, *Computers*, **10:3** (2021), id. 27, 19 pp. 

- [14] P. Dhilleswararao, S. Boppu, M. S. Manikandan, L. R. Cenkeramaddi. “Efficient hardware architectures for accelerating deep neural networks: Survey”, *IEEE Access*, **10** (2022), pp. 131788–131828. 
- [15] K. Foster. *Associativnyye parallel'nye processory*, Energoizdat, M., 1981, 240 pp.; C. C. Foster. *Content Addressable Parallel Processors*, Van Nostrand Reinhold, 1976, ISBN 978-0442224332, 233 pp.
- [16] *IEEE 1588-2019 IEEE Standard for a Precision Clock Synchronization Protocol for Networked Measurement and Control Systems*, IEEE Instrumentation and Measurement Society, 2020. 
- [17] E. F. Dierikx, A. E. Wallin, T. Fordell, J. Myyry, P. Koponen, M. Merimaa, T. J. Pinkert, J. C. J. Koelemeij, Peek H. Z.. “White Rabbit Precision Time Protocol on Long-Distance Fiber Links”, *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, **63**:7 (2016), pp. 945–952. 
- [18] W. S. Rabinovich, P. G. Goetz, R. Mahon, L. A. Swingen, J. L. Murphy M. Ferraro, H. R. Burris, Ch. I. Moore, M. R. Suite, G. C. Gilbreath, S. C. Binari, D. J. Klotzkin. “45-Mbit/s cat’s-eye modulating retroreflectors”, *Optical Engineering*, **46**:10 (2007), pp. 104001–104008. 
- [19] Y. Zhu, G. Wang. “Research on retro-reflecting modulation in space optical communication system”, *IOP Conf. Series: Earth and Environmental Science*, **108**:3 (2018), id. 032060, 8 pp. 
- [20] M. Bosiljevac, J. Downing, D. Babic. “<100 ppm/K output intensity temperature stability with single-color light-emitting diodes”, *Applied Optics*, **55**:3 (2016), pp. 9060–9066. 



Improving the accuracy of segmentation masks using a generative-adversarial network model

Igor Victorovich **Vinokurov**

Financial University under the Government of the Russian Federation, Moscow, Russia

igvinokurov@fa.ru

Abstract. Masks obtained using the deep learning model Mask R-CNN may in some cases contain fragmented contours, uneven boundaries, false fusions of adjacent objects, and areas with missed segmentation. The more detection objects in the image and the smaller their size, the more often various types of defects in their masks are encountered. Examples of such images include aerial photographs of cottage and garden associations and cooperatives characterized by high building density. To correct these defects, it is proposed to use a generative adversarial network model that performs post-processing of the predicted Mask R-CNN masks.

A qualitative assessment of the model formed in the work demonstrated that it is capable of restoring the integrity of contours at an acceptable level, filling in missing areas, and separating erroneously merged objects. Quantitative analysis using the IoU, precision, recall, and F1-score metrics showed a statistically significant improvement in the segmentation quality compared to the original Mask R-CNN masks. The obtained results confirmed that the proposed approach allows to increase the accuracy of the formation of object masks to a level that satisfies the requirements of their practical application in automated aerial photograph analysis systems. (*Linked article texts in English and in Russian*).

Key words and phrases: Computer vision, image segmentation, object masks, generative adversarial networks, Mask R-CNN, PyTorch

2020 *Mathematics Subject Classification:* 68T20; 68T07, 68T45

For citation: Igor V. Vinokurov. *Improving the accuracy of segmentation masks using a generative-adversarial network model*. Program Systems: Theory and Applications, 2025, **16**:2(65), pp. 111–152. (*In English, in Russian*). https://psta.psiras.ru/read/psta2025_2_111-152.pdf

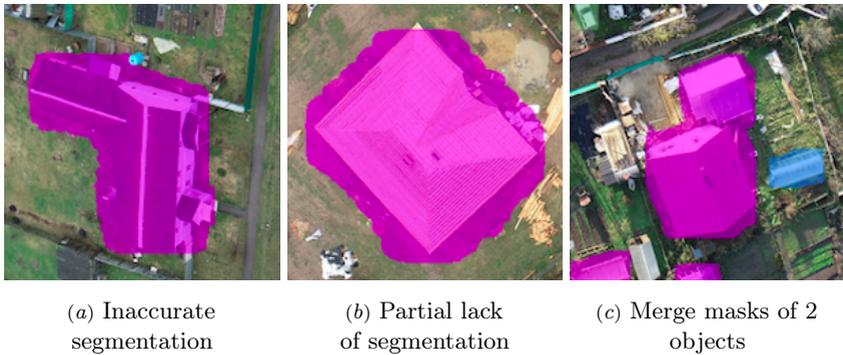


FIGURE 1. Errors in generated object masks

Introduction

Detecting correct property masks in dense aerial images is a challenge for deep learning models, Mask R-CNN, YOLO, and others. The main reason is that buildings are located close to each other, which makes it difficult to accurately separate objects, since their contours may intersect or merge in the image.

This is especially problematic in high-detail settings, where small architectural elements such as extensions, balconies, or common walls between buildings may be perceived by the model as a single object. In addition, complex building shapes, such as polygonal outlines or non-standard structures, require deeper analysis and adaptation from the model, which increases the likelihood of errors in mask generation. Several examples of such masks are shown in Figure 1.

Another reason is the presence of external factors, such as vegetation and shadows, which partially obscure objects in images. Trees, bushes, or even temporary objects such as cars can cover parts of buildings, creating gaps in the data available to the model. Shadows caused by different shooting angles or time of day also distort visual information, reducing contrast and making it difficult to distinguish clear boundaries of objects. These factors lead to the model either underestimating the area of an object, ignoring hidden areas, or overestimating it, including extraneous elements in the mask. As a result, there is an increased need for additional processing and improving the quality of predictions to ensure the correctness of the final data.

The problem of refining the boundaries of objects is especially relevant for the «Roskadastr» PLC information system (IS), where increased

requirements are placed on the accuracy of spatial data. Traditional post-processing methods, such as morphological operations or active contour algorithms, are often not effective enough for complex cases of dense development. In this regard, there is a need to solve the problem of automatic correction of segmentation results, while maintaining the topological correctness of objects.

The presented work proposes one of the possible solutions to this problem, based on the use of a generative adversarial network (GAN) for post-processing the segmentation results obtained using Mask R-CNN. The main idea of the study is that the GAN architecture, properly trained on a representative sample of development objects, typical for the «Roskadastr» PLC software package, will be able to effectively restore the integrity of broken boundaries, eliminate internal gaps in masks, separate erroneously merged objects and maintain the exact position of corner (turning) points of buildings.

The research implementation consisted of the following stages:

- (1) Formation of a representative and balanced dataset containing correspondences of the original masks from Mask R-CNN to reference masks extracted from files with object annotations and masks obtained as a result of their expert correction.
- (2) Development of a GAN architecture that takes into account the features of dense development in cottage and garden associations and cooperatives, requirements for the accuracy of boundaries for cadastral registration and the need to preserve the topology of objects.
- (3) Conducting experimental studies with an assessment of the qualitative characteristics of the resulting masks and quantitative indicators (IoU, precision, recall, F1-score).

1. Overview of work on using GAN to improve objects detection

There is currently a significant amount of research devoted to the use of generative adversarial networks (GANs) to improve the efficiency of image recognition and transformation.

The paper [3] discusses conditional GAN (cGAN) models that can be used to transform images of one type into another. For example, segmentation masks can be transformed into realistic images, or vice versa. The described models demonstrate high accuracy of object segmentation and improve the quality of their masks.

The Pix2Pix model described in [4] and based on cGAN is used to synthesize high-quality images based on input data such as segmentation masks. The paper shows that GAN can improve the quality of masks by training on (mask, exact mask) pairs.

The study [5] presents the MaskGAN model designed for interactive editing of face images using masks. The model improves the quality of masks and allows manipulation of segmented areas of images.

In [6], a SegAN model with a multi-level discriminator and L1 loss is proposed for medical segmentation. The model uses a multi-level discriminator to improve the quality of masks, especially in tasks with fuzzy object boundaries. The GAN model improves object boundaries, reduces noise, and eliminates object detection inaccuracies.

The attention mechanism for generating masks in Attention-GAN described in [7] is applied to transform objects in wildlife photographs. The model can eliminate noise and detection inaccuracies while preserving fine object details.

In [8], an image transformation method is proposed using CycleGAN and cyclic loss during training. The model can be adapted to improve segmentation masks, especially in tasks with inconsistent data. It eliminates noise and detection inaccuracies while preserving fine object details.

The Boundary-Aware GAN model with a multi-level discriminator described in [9] is designed to improve object boundaries in semantic segmentation problems. The discriminator is trained to distinguish between real and predicted boundaries. Improvements in object detection include identifying clear boundaries, reducing noise, and eliminating inaccuracies.

The paper [10] explores the application of GANs for binary semantic segmentation on imbalanced datasets. The authors propose a new architecture and approach to training GANs to improve the quality of object mask prediction under conditions of strong class imbalance. The main focus is on the problems of low prediction accuracy for the minority class, loss of fine object details, noise, and artifacts in the predicted masks. The authors demonstrate that GAN can improve the quality of masks due to more accurate modeling of object boundaries and restoration of fine details.

The article [11] describes the Power-Line GAN (PLGAN) model for segmentation of thin and long objects against a background of complex textures and noise. The main goal is to improve the accuracy of power line detection, which is important for infrastructure monitoring, mapping, and environmental analysis. The key ideas proposed in the paper include adapting GAN for thin object segmentation, improving detail, handling imbalanced data, and smoothing edges using active contours and spline interpolation.

The problem of unbalanced data distribution between the source domain (a set of images) and the target domain in the tasks of semantic segmentation of aerial photographs is considered in [12]. The authors propose a method for training a model without using labels for the target domain. For this purpose, a GAN is used that adapts the stylistic features of the source domain to the target one. The main goal of this work is to improve the segmentation accuracy on the target domain, minimizing the influence of differences between domains, such as changes in lighting, shooting angles, or object textures. The authors note the following improvements achieved using the proposed approach: better generalization of the model, eliminating the influence of differences between domains, preserving fine details, and reducing noise.

In [13], the application of cGAN to the task of semantic segmentation of medium-resolution satellite images is investigated. The authors focus on improving the accuracy of object extraction in such images, which often suffer from noise, lack of detail, and complex textures. The main goal of the work is to demonstrate how cGAN can be used to improve the quality of segmentation compared to traditional approaches such as U-Net or DeepLab. Particular attention is paid to the tasks of land use monitoring, urban planning, and environmental change analysis. The improvements achieved by the proposed approach include high accuracy for minority classes, sharp object boundaries, preservation of fine details, and noise removal.

In [14], the authors propose a new approach that uses GAN to improve the quality of predicted masks by more accurately modeling object boundaries and restoring fine details. The main focus is on tasks related to the analysis of high-resolution images such as aerial photographs, medical images, and urban landscapes. The aim of the work is to show that using

GAN can significantly improve the accuracy and detail of object instance segmentation compared to traditional methods such as Mask R-CNN or DeepLab. Improving the detail and smoothing the contours of masks are the main results of this work.

2. Dataset generation

The dataset used in the experimental studies is a set of (predicted mask, exact mask) pairs. The predicted mask is a binary mask of an object obtained using the Mask R-CNN model. The exact mask is a binary mask formed from a polygon of the same object extracted from a json file.

The types and quantities of objects presented in the dataset, as well as the features of forming json files for image files, are described in [1]. For polygons of objects missing from the json files, the corresponding exact masks were formed (corrected) manually. This made it possible to create a representative and consistent dataset, including both automatically generated and manual masks. All masks in the dataset were reduced to a size of 256×256 pixels. Larger mask sizes lead to a sharp increase in computational load and, accordingly, computation time, smaller ones lead to uneven contours of predicted masks, after increasing their size. Increasing to the size of an image with real estate objects is necessary at the stage of visualization of the result in order to combine the mask of the object and the image containing it.

3. Model creation and exploration

3.1. GAN architecture features

GANs consist of two main components, a generator and a discriminator, which operate in an adversarial mode. The main goal of a GAN is to train the generator to produce data that is indistinguishable from real data.

The generator takes random noise (usually from a normal distribution) as input during training and transforms it into some data, such as an image. The generator's task is to produce data that looks real. To do this, it learns to minimize the losses that occur when the discriminator tries to detect a fake. The main generator architectures are the following:

DCGAN (Deep Convolutional GAN) [15] is simple and effective, but may produce less detailed results on complex data. The generator uses transposed convolutions to gradually increase the image resolution. It starts with a random vector (latent space) and successively increases the dimensionality of the data to the target resolution.

U-Net [16] is suitable for tasks where it is important to preserve spatial details, such as segmentation or mask enhancement. The generator is based on the U-Net architecture, which includes an encoder for dimensionality reduction and a decoder for image restoration. It uses a skip connection mechanism, which allows preserving fine details.

StyleGAN [17] provides high detail and control over style, but requires large computational resources. Allows control over different levels of image detail (e.g. overall shape and textures). The generator is divided into two parts for style matching and synthesis.

The discriminator evaluates the quality of the generated data. The goal of the discriminator is to help the generator generate more reliable data (images) by providing correct error signals. The main discriminator architectures are:

PatchGAN [3] is effective for tasks where object boundaries and fine details are important. It analyzes the image by small local areas ("patches") instead of the entire image. Each patch is classified as real or fake, which allows focusing on local details.

Multi-level discriminator [18] improves the quality of generation due to a more comprehensive assessment of the data. It uses multiple discriminators, each operating at different levels of resolution. This allows evaluating both global and local image characteristics.

Conditional GAN (cGAN) [19] is the most effective for tasks that require directed generation, such as translating images between groups. The discriminator accepts additional information (such as class labels or conditional data) along with the image, allowing control over the type of data generated.

Considering the features of the above-described generator and discriminator architectures, the U-Net and PatchGAN architectures were chosen to improve the accuracy of small real estate masks, respectively.

3.2. Model creation

The generator, implemented on the basis of the U-Net architecture, implements two main stages: compression (downsampling) and expansion (upsampling).

The compression stage reduces the spatial resolution of the data. Each of the 5 groups of elements reduces the image size by 2 times and consists of a convolutional layer (`Conv2d`) with a kernel size (`kernel_size`) of 4x4, a stride (`stride`) of 2 and a padding (`padding`) of 1. Then a `BatchNorm2d` layer can follow with the activation function `LeakyReLU` and a coefficient of 0.2.

Dilation elements, on the contrary, increase the spatial resolution of the data. There are also 5 expansion elements in the proposed generator. Each such element increases the image size by 2 times by performing a transposed convolution (`ConvTranspose2d`) with a kernel size of 4x4, a stride of 2, and a padding of 1. Then a `BatchNorm2d` layer follows (except for the last block) with the activation function `ReLU` (except for the last block, which uses `Tanh` to limit the output to the range [-1, 1]). The organization of the GAN generator is shown in Figure 2.

The expansion element combines its output data with the data from the corresponding image compression element via skip connectors. This allows preserving details and improving the quality of generation. The generator output is a single-channel image of the same size as the input. The compression and expansion elements of the level 1 generator described above are shown in Figure 3 and Figure 4, respectively.

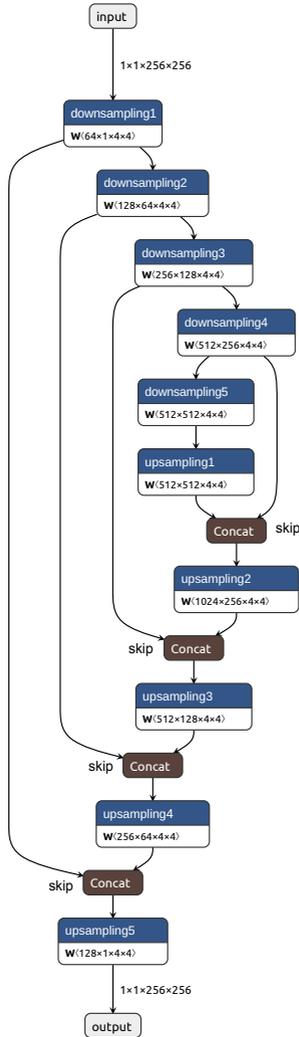
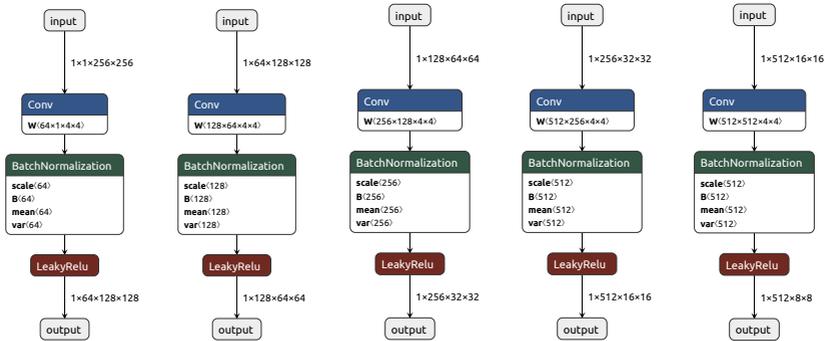


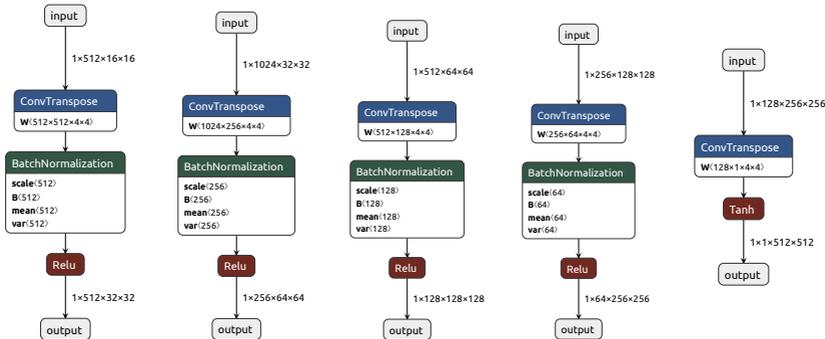
FIGURE 2. GAN generator

As noted above, the discriminator based on the PatchGAN architecture analyzes the image by small local regions (patches). Each patch is classified as real or fake, which allows focusing on local details. The discriminator



(a) downsampling1 (b) downsampling2 (c) downsampling3 (d) downsampling4 (e) downsampling5

FIGURE 3. GAN generator compression elements



(a) upsampling1 (b) upsampling2 (c) upsampling3 (d) upsampling4 (e) upsampling5

FIGURE 4. GAN generator extension elements

takes two input images: the predicted image (generated by the generator) and the real image (in our case, they are accurate object masks). Then these images are concatenated by channels. The next few convolutional layers (Conv2d) with a kernel size of 4x4, a stride of 2, and a padding of 1 reduce the spatial dimensions of the data.

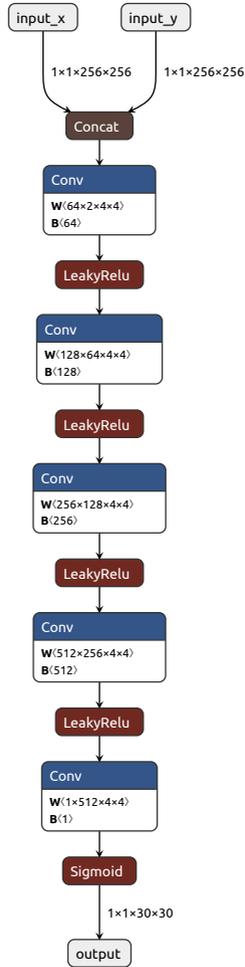


FIGURE 5. GAN discriminator

Starting from the second layer, BatchNorm2d is applied with the activation function LeakyReLU and a coefficient of 0.2. The last layer with the activation function Sigmoid calculates the probability that the input data is real. The structure of the discriminator is shown in Figure 5.

The generator uses the standard loss function for binary classification problems `BCEWithLogitsLoss` and `L1 Loss` to minimize the pixel difference between the generated and real masks. The first of these functions is used to “fool” the discriminator, the second helps the generator create masks that are close to the real ones in pixel values. An important parameter for `L1 Loss` is λ_{L1} , which controls the importance of the pixel difference between the generated and real masks. This allows you to balance between “fooling” the discriminator and the accuracy of detail reproduction. The generator loss function is `BCEWithLogitsLoss`.

Training GANs is associated with certain problems, such as instability of the training process, mode collapse, and low generation quality, which require the use of special methods to address them. One such approach is spectral normalization, which limits the spectral norm of the weights in the generator and discriminator layers, thereby stabilizing training and preventing gradient explosion [20]. Another method is to use a gradient penalty, which adds regularization to the discriminator outputs, providing a smoother distribution of values and improving model convergence [21]. Self-attention, which allows the model to take into account global dependencies in the data, is also widely used, which is especially useful when working with high-dimensional images [22]. However, since restoring simple binary masks is a relatively simple task, and as the results of the study of the GAN model described above showed, the use of these approaches in this case is redundant.

3.3. Model Study Results

The GAN model was developed and studied using the `PyTorch` framework in Google Colab Pro. The Jupyter notebook (MIT License) is available at this link—  `jupyter`. The main accuracy metrics were studied for different values of learning rate, number of epochs, and batch sizes. The accuracy metrics of the GAN generator were studied at $\lambda_{L1} = 100$, since

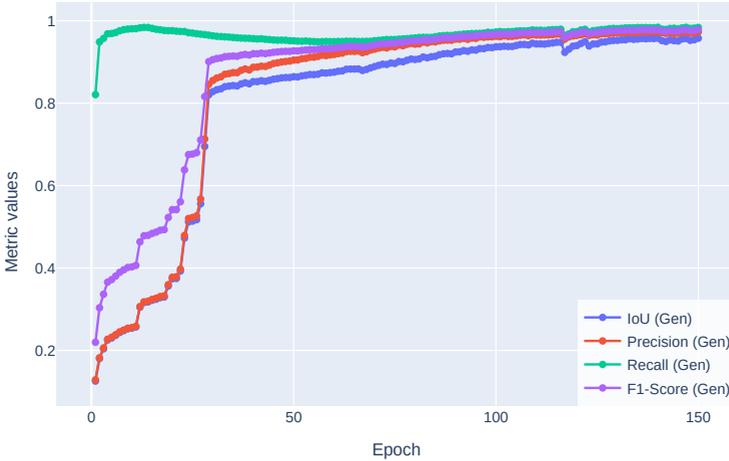


FIGURE 6. GAN generator accuracy metrics for $\lambda_{L1}=80$

at this value their curves have a sufficient degree of smoothness, which indicates the absence of explosive or decaying gradients. In Figure 6, Figure 7, and Figure 8 show the accuracy metrics curves for $\lambda_{L1} = 80, 100$ and 120 .

Table 1 present the results of studies of the main accuracy metrics IoU, IoU0.5-95, Precision, Recall and F1-Score. As the studies have shown, the use of GAN can significantly improve the accuracy of real estate segmentation masks obtained using the Mask R-CNN [1] model.

Let's look at some examples of improving real estate masks obtained using the Mask R-CNN model. Figure 9 shows masks predicted by the Mask R-CNN model and improved masks generated by GAN. Based on the accuracy metrics, it can be concluded that using GAN can significantly improve the quality of masks by correcting errors such as contour discontinuities, noise, and inaccuracies in the boundaries of objects.

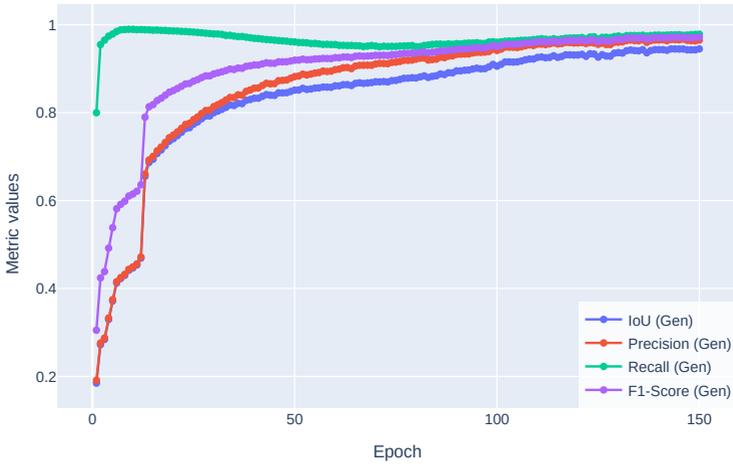


FIGURE 7. GAN generator accuracy metrics for $\lambda_{L1}=100$

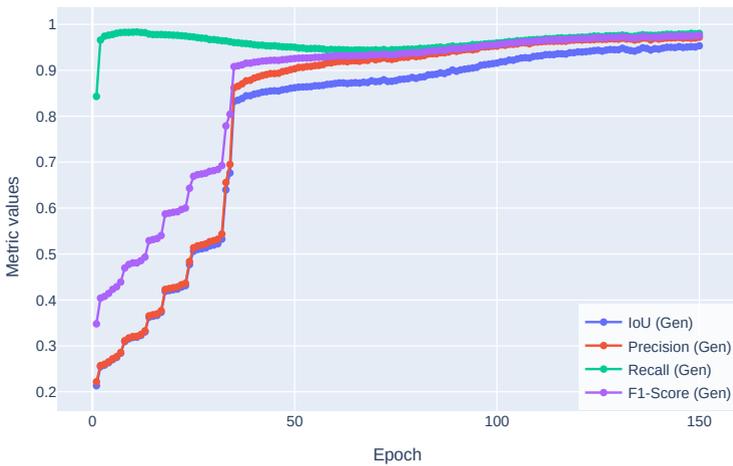


FIGURE 8. GAN generator accuracy metrics for $\lambda_{L1}=120$

TABLE 1. Evaluation of accuracy metrics for predicted GAN masks under different values of model training parameters

Learning Speed	Batch Size	Learning Epochs	IoU (Pred)	Precision (Pred)	Recall (Pred)	F1-Score (Pred)	IoU0.5-95 (Pred) (Mask R-CNN)	IoU0.5-95 (Pred) (GAN)
1e-6	10	500	0.87	0.92	0.91	0.91	0.67	0.90
5e-6	10	400	0.85	0.90	0.89	0.90	0.65	0.88
5e-6	20	350	0.84	0.89	0.88	0.88	0.64	0.87
1e-5	10	300	0.83	0.88	0.87	0.87	0.63	0.86
1e-5	20	250	0.84	0.89	0.88	0.88	0.64	0.87
5e-5	8	200	0.79	0.84	0.83	0.83	0.59	0.81
5e-5	10	300	0.82	0.87	0.86	0.86	0.62	0.85
5e-5	16	250	0.81	0.86	0.85	0.85	0.61	0.84
4e-5	8	150	0.80	0.85	0.84	0.84	0.60	0.82
4e-5	16	200	0.78	0.83	0.82	0.82	0.58	0.80
1e-4	8	100	0.75	0.80	0.78	0.79	0.55	0.78
1e-4	16	150	0.77	0.82	0.81	0.81	0.57	0.80
1e-4	20	200	0.78	0.83	0.82	0.82	0.58	0.81
2e-4	8	80	0.70	0.75	0.73	0.74	0.50	0.73
2e-4	16	100	0.72	0.78	0.76	0.77	0.52	0.76



FIGURE 9. The results of GAN enhancement are shown under the Mask R-CNN predicted real estate object masks

An example of improving object masks on an aerial photograph, where the complexity of the task increases due to the high density of objects and their diversity, is shown in Figure 10.

When generating masks using GAN, the problem of uneven and stepped edges often arises, see Figure 9. This is due to the specifics of GAN operation, which is associated with the discreteness of pixels when forming images and the complexity of accurately reproducing small details of object boundaries. To improve the contours of segmentation masks, various morphological smoothing functions can be applied, which allow you to eliminate noise, fill in gaps and make object boundaries smoother.

One of the basic approaches is to use morphological opening and closing operations with a given structuring element (for example, a square matrix (5×5)). Opening helps to remove small noise artifacts, and closing fills small gaps inside the mask. For additional smoothing of contours, you can apply Gaussian blur (`GaussianBlur`), which softens sharp transitions between the background and the object.

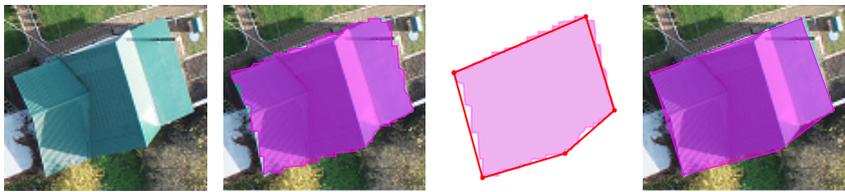


(a) Results of detecting real estate masks using the Mask R-CNN model



(b) The masks are improved by GAN generator

FIGURE 10. Improvement of objects masks



(a) Real estate (b) Jagged edges (c) Aligned edges (d) Resulting mask

FIGURE 11. Eliminating mask aliasing using the Douglas-Peucker algorithm

Another effective method is to use active contours (snakes), which adaptively adjust to the shape of the object by minimizing the elastic energy and smoothness. To achieve even smoother contours, spline interpolation can be used, which approximates the original contour using cubic splines, ensuring uniform distribution of points and smooth boundaries.

Another effective way to eliminate aliasing is to use the Douglas-Peucker algorithm, which specializes in simplifying polylines and smoothing contours [23]. The peculiarity of this algorithm is its ability to analyze each point of the mask boundary and evaluate its significance for the overall shape of the object. Points that make a minimal contribution to the contour configuration are removed, which allows you to get smoother and more natural boundaries without losing key geometric features. An example of using the Douglas-Peucker algorithm is shown in Figure 11.

These methods or their combinations allow in some cases to improve the quality of masks, making them more suitable for subsequent analysis or visualization. As studies have shown, the use of morphological smoothing methods and improvement of mask boundaries using the Douglas-Peucker algorithm in post-processing of masks makes it possible to increase the accuracy of their formation by an average of 5–7%, which is especially important for computer vision tasks and segmentation of objects on aerial photographs in particular.

Conclusion

In this paper, a GAN model was proposed and investigated to improve the accuracy of segmentation masks obtained using the Mask R-CNN model. Experiments have shown that the use of GAN can significantly improve the quality of detected real estate masks by correcting errors such as gaps in masks, noise artifacts, and jagged edges. The improvement is achieved by using the L1 loss function to minimize pixel differences between the target and generated masks, as well as by using a discriminator that contributes to more realistic mask generation.

The results of the studies demonstrate an improvement in accuracy metrics such as IoU, Precision, Recall, and F1-Score for the improved masks compared to the original Mask R-CNN predictions. The average IoU_{0.5-95} value increased by an average of 30-35%, which confirms the effectiveness of the proposed approach. It was additionally noted that the use of morphological operations and contour smoothing methods will further improve the smoothness and accuracy of mask boundaries.

The obtained results are of great practical importance and will be used in the «Roskadastr» PLC IS for automatic conversion of raster images with real estate objects into vector maps. This will improve the accuracy and speed of digital map generation, which is especially important for cadastral registration and territory management tasks. The proposed approach can be adapted for other applications related to image processing and computer vision, where high accuracy of object segmentation is required. The study conducted in the work confirms the prospects of using GAN to improve the quality of real estate masks, which allows it to be effectively used to automate the processes of data analysis and transformation in geographic information systems. Further research can be aimed at optimizing the model parameters, introducing new methods of mask post-processing and expanding the scope of GAN.

References

- [1] Vinokurov I. V.. “Using the Mask R-CNN model for segmentation of real estate objects in aerial photographs”, *Program Systems: Theory and Applications*, **16**:1(64) (2025), pp. 3-44.   [↑_{116, 123}](#)

- [2] G. Cohen, R. Giryes. *Generative adversarial networks*, 2024, 28 pp.  [arXiv:2203.00667](#) 
- [3] P. Isola, J.-Y. Zhu, T. Zhou, A. A. Efros. *Image-to-image translation with conditional adversarial networks*, 2016, 17 pp.  [arXiv:1611.07004](#) 113, 117
- [4] T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, A. Tao, J. Kautz, B. Catanzaro. *High-resolution image synthesis and semantic manipulation with conditional GANs*, 2017, 14 pp.  [arXiv:1711.11585](#) 114
- [5] C.-H. Lee, Z. Liu, L. Wu, P. Luo. *MaskGAN: Towards diverse and interactive facial image manipulation*, 2019, 20 pp.  [arXiv:1907.11922](#) 114
- [6] Y. Xue, T. Xu, H. Zhang, L. Rodney Long, X. Huang. *SegAN: Adversarial network with multi-scale L_1 loss for medical image segmentation*, 2017, 9 pp.  [arXiv:1706.01805](#) 114
- [7] X. Chen, C. Xu, X. Yang, D. Tao. *Attention-GAN for object transfiguration in wild images*, 2018, 18 pp.  [arXiv:1803.06798](#) 114
- [8] J.-Y. Zhu, T. Park, P. Isola, A. A. Efros. *Unpaired image-to-image translation using cycle-consistent adversarial networks*, 2017, 18 pp.  [arXiv:1703.10593](#) 114
- [9] J. Gong, J. Xu, X. Tan, J. Zhou, Y. Qu, Y. Xie, L. Ma. *Boundary-aware geometric encoding for semantic segmentation of point clouds*, 2021, 9 pp.  [arXiv:2101.02381](#) 114
- [10] L. Xu, M. Gabbouj. *Revisiting generative adversarial networks for binary semantic segmentation on imbalanced datasets*, 2024, 14 pp.  [arXiv:2402.02245](#) 114
- [11] R. Abdelfattah, X. Wang, S. Wang. *JPLGAN: Generative adversarial networks for power-line segmentation in aerial images*, 2022, 11 pp.  [arXiv:2204.07243](#) 115
- [12] B. Benjdira, Y. Bazi, A. Koubaa, K. Ouni. “Unsupervised domain adaptation using generative adversarial networks for semantic segmentation of aerial images”, *Remote Sens.*, **11**:11 (2019), id. 1369, 23 pp.  [arXiv:1905.03198](#) 115
- [13] A. Kulkarni, T. Mohandoss, D. Northrup, E. Mwebaze, H. Alemohammad. *Semantic segmentation of medium-resolution satellite imagery using conditional generative adversarial networks*, 2020, 7 pp.  [arXiv:2012.03093](#) 115
- [14] Q. H. Le, K. Youcef-Toumi, D. Tsetserukou, A. Jahanian. *GAN Mask R-CNN: Instance semantic segmentation benefits from generative adversarial networks*, 2020, 13 pp.  [arXiv:2010.13757](#) 115
- [15] A. Radford, L. Metz, S. Chintala. *Unsupervised representation learning with deep convolutional generative adversarial networks*, 2016, 16 pp.  [arXiv:1511.06434](#) 117

- [16] O. Ronneberger, Ph. Fischer, Th. Brox. *U-Net: Convolutional networks for biomedical image segmentation*, 2015, 8 pp. [doi](#) [arXiv](#) 1505.04597 ^{↑117}
- [17] T. Karras, S. Laine, T. Aila. *A style-based generator architecture for generative adversarial networks*, 2018, 12 pp. [doi](#) [arXiv](#) 1812.04948 ^{↑117}
- [18] T. Karras, T. Aila, S. Laine, J. Lehtinen. *Progressive growing of GANs for improved quality, stability, and variation*, 2017, 26 pp. [doi](#) [arXiv](#) 1710.10196 ^{↑117}
- [19] M. Mirza, S. Osindero. *Conditional generative adversarial nets*, 2014, 7 pp. [doi](#) [arXiv](#) 1411.1784 ^{↑117}
- [20] T. Miyato, T. Kataoka, M. Koyama, Y. Yoshida. *Spectral normalization for generative adversarial networks*, 2018, 26 pp. [doi](#) [arXiv](#) 1802.05957 ^{↑122}
- [21] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, A. C. Courville. *Improved training of Wasserstein GANs*, 2017, 20 pp. [doi](#) [arXiv](#) 1704.00028 ^{↑122}
- [22] H. Zhang, I. Goodfellow, D. Metaxas, A. Odena. *Self-attention generative adversarial networks*, 2019, 10 pp. [doi](#) [arXiv](#) 1805.08318 ^{↑122}
- [23] H. Chen. “An improved Douglas-Peucker algorithm applied in coastline generalization”, Fourth International Conference on Geology, Mapping, and Remote Sensing (ICGMRS 2023) (14–16 April 2023, Wuhan, China), Proc. SPIE, vol. **12978**, 2024, id. 1297804. [doi](#) ^{↑128}

Received	21.04.2025;
approved after reviewing	10.06.2025;
accepted for publication	11.06.2025;
published online	28.06.2025.

Recommended by

prof. A. M. Elizarov

Information about the author:



Igor Victorovich Vinokurov

Candidate of Technical Sciences (PhD), Associate Professor at the Financial University under the Government of the Russian Federation. Research interests: information systems, information technologies, data processing technologies

[ID](#) 0000-0001-8697-1032

e-mail:

The author declare no conflicts of interests.

УДК 004.932.75'1, 004.89

 10.25209/2079-3316-2025-16-2-111-152

Повышение точности сегментирования объектов с использованием генеративно-сопоставительной сети

Игорь Викторович **Винокуров**[✉]

Финансовый Университет при Правительстве Российской Федерации, Москва, Россия

[✉]igvvinokurov@fa.ru

Аннотация. Маски, полученные с использованием модели глубокого обучения Mask R-CNN, в ряде случаев могут содержать фрагментированные контуры, неровные границы, ложные сращения соседних объектов и участки с пропущенной сегментацией. Чем больше объектов детектирования на изображении и меньше их размер, тем более часто встречаются различного вида недостатки их масок. Примерами таких изображений могут являться аэрофотоснимки коттеджных и садовых товариществ и кооперативов, характеризующихся высокой плотности застройки. Для коррекции указанных недостатков предлагается использовать модель генеративно-сопоставительной сети, выполняющую постобработку предсказанных Mask R-CNN масок.

Качественная оценка сформированной в работе модели продемонстрировала, что она способна на приемлемом уровне восстанавливать целостность контуров, заполняет пропущенные области и разделять ошибочно объединенные объекты. Количественный анализ по метрикам IoU, precision, recall и F1-score показал статистически значимое улучшение качества сегментации по сравнению с исходными масками Mask R-CNN. Полученные результаты подтвердили, что предложенный подход позволяет довести точность формирования масок объектов до уровня, удовлетворяющего требованиям их практического применения в системах автоматизированного анализа аэрофотоснимков. (*Связанные тексты статьи на английском и на русском языках*)

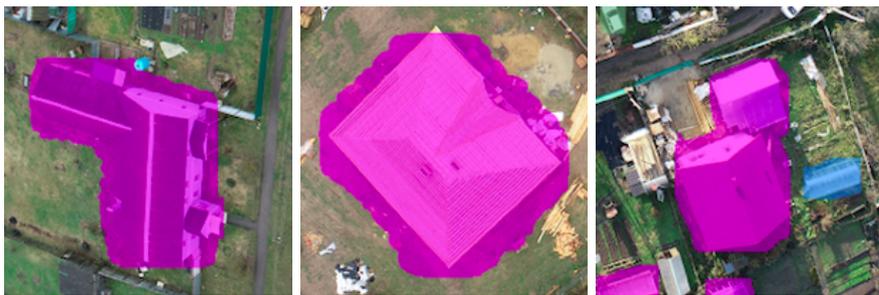
Ключевые слова и фразы: Компьютерное зрение, сегментация изображений, маски объектов, генеративно-сопоставительные сети, Mask R-CNN, PyTorch

Для цитирования: Винокуров И.В. *Повышение точности сегментирования объектов с использованием генеративно-сопоставительной сети* // Программные системы: теория и приложения. 2025. Т. 16. № 2(65). С. 111–152. (Англ.+русс.) https://psta.psiras.ru/read/psta2025_2_111-152.pdf

Введение

Выявление точных масок объектов недвижимости на аэрофотоснимках с плотной застройкой представляет определённую сложность для моделей глубокого обучения, Mask R-CNN, YOLO и других. Основной причиной является близкое расположение зданий друг к другу, что затрудняет точное разделение объектов, поскольку их контуры могут пересекаться или сливаться на изображении.

Это особенно проблематично в условиях высокой детализации, где небольшие элементы архитектуры, такие как пристройки, балконы или общие стены между строениями, могут быть восприняты моделью как единый объект. Кроме того, сложная форма зданий, например, многоугольные очертания или нестандартные конструкции требует от модели более глубокого анализа и адаптации, что увеличивает вероятность ошибок при формировании масок. Несколько примеров таких масок приведены на рисунке 1.



(а) Неточная сегментация (б) Частичное отсутствие сегментации (в) Слияние масок объектов

РИСУНОК 1. Ошибки при формировании масок объектов

Ещё одной причиной является наличие внешних факторов, таких как растительность и тени, которые частично скрывают объекты на снимках. Деревья, кустарники или даже временные объекты, такие как автомобили, могут перекрывать части зданий, создавая разрывы в данных, доступных модели. Тени, вызванные различными углами съёмки или временем суток, также искажают визуальную информацию, снижая контраст и затрудняя выделение четких границ объектов. Эти факторы приводят к тому, что модель может либо недооценивать площадь объекта, игнорируя скрытые участки, либо переоценивать ее, включая посторонние элементы в маску. В результате повышается потребность в дополнительной обработке и улучшении качества предсказаний, чтобы обеспечить корректность итоговых данных.

Проблема уточнения границ объектов особенно актуальна для информационной системы (ИС) ППК «Роскадастр», где к точности пространственных данных предъявляются повышенные требования. Традиционные методы постобработки, такие как морфологические операции или алгоритмы активных контуров, часто оказываются недостаточно эффективными для сложных случаев плотной застройки. В связи с этим возникает необходимость в решении задачи автоматического корректирования результатов сегментации, сохраняя при этом топологическую правильность объектов.

В представленной работе предлагается одно из возможных решений этой задачи, основанное на применении генеративно-сопоставительной сети (GAN) для постобработки результатов сегментации, полученных с помощью Mask R-CNN. Основная идея исследования заключается в том, что GAN-архитектура, должным образом обученная на репрезентативной выборке объектов застройки, характерной для ППК «Роскадастр», сможет эффективно восстанавливать целостность нарушенных границ, устранять внутренние разрывы масок, разделять ошибочно слившиеся объекты и сохранять точное положение угловых (поворотных) точек зданий.

Реализация исследований состояла из следующих этапов:

- (1) Формирование репрезентативного и сбалансированного датасета, содержащего соответствия исходных масок от Mask R-CNN эталонным маскам, извлечённым из файлов с аннотациями объектов и маскам, полученные в результате их экспертной коррекции.
- (2) Разработка архитектуры GAN, учитывающей особенности плотной застройки в коттеджных и садовых товариществах и кооперативах, требования к точности границ для кадастрового учета и необходимость сохранения топологии объектов.
- (3) Проведение экспериментальных исследований с оценкой качественных характеристик получаемых масок и количественных показателей (IoU, precision, recall, F1-score).

1. Обзор работ по использованию GAN для улучшения результатов распознавания объектов

В настоящее время существует значительное количество научных исследований, посвященных применению генеративно-сопоставительных сетей (GAN) для повышения эффективности распознавания изображений и их трансформации.

В работе [3] рассматриваются модели условно-сопоставительных сетей (conditional GAN, cGAN), которые могут быть применены для преобразования изображений одного типа в другой. Например, маски сегментации могут быть преобразованы в реалистичные изображения, или наоборот. Описанные модели демонстрируют высокую точность сегментации объектов и улучшают качество их масок.

Модель Pix2Pix, описанная в [4] и основанная на cGAN, используется для синтеза высококачественных изображений на основе входных данных, таких как маски сегментации. В статье показано, что GAN может улучшить качество масок путем обучения на парах «маска-точная маска».

В исследовании [5] представлена модель MaskGAN, предназначенная для интерактивного редактирования изображений лиц с использованием масок. Модель улучшает качество масок и позволяет манипулировать сегментированными областями изображений.

В [6] предлагается модель SegAN с многоуровневым дискриминатором и L1-потерями для медицинской сегментации. Модель использует многоуровневый дискриминатор для улучшения качества масок, особенно в задачах с размытыми границами объектов. GAN-модель улучшает границы объектов, снижает уровень шума и устраняет неточности детектирования объектов.

Механизм внимания для генерации масок в Attention-GAN, описанный в [7], применяется для трансформации объектов на фотографиях дикой природы. Модель позволяет устранять шумы и неточности детектирования, сохраняя мелкие детали объектов.

В статье [8] предлагается метод преобразования изображений с использованием CycleGAN и циклических потерь при обучении. Модель может быть адаптирована для улучшения масок сегментации, особенно в задачах с несоответствующими данными. Она устраняет шумы и неточности детектирования объектов, сохраняя их мелкие детали.

Модель Boundary-Aware GAN с многоуровневым дискриминатором, описанная в [9], предназначена для улучшения границ объектов в задачах семантической сегментации. Дискриминатор обучается различать реальные и предсказанные границы. Улучшения детектирования объектов заключаются в выявлении четких границ, снижении уровня шума и устранении неточностей.

В статье [10] исследуется применение GAN для бинарной семантической сегментации на несбалансированных датасетах. Авторы предлагают новую архитектуру и подход к обучению GAN для улучшения качества предсказания масок объектов в условиях сильной разбалансировки классов. Основное внимание уделяется проблемам низкой точности предсказания для миноритарного класса, потере мелких деталей объектов, шуму и артефактам в предсказанных масках. Авторы демонстрируют, что GAN позволяет улучшить качество масок за счёт более точного моделирования границ объектов и восстановления мелких деталей.

В работе [11] описывается модель PLGAN (Power-Line GAN), предназначенная для сегментации тонких и длинных объектов на фоне сложных текстур и шума. Основная цель — улучшить точность выделения линий электропередач, что важно для мониторинга инфраструктуры, картографирования и анализа окружающей среды. В статье предложены ключевые идеи, включая адаптацию GAN для сегментации тонких объектов, улучшение детализации, обработку несбалансированных данных и сглаживание контуров с использованием активных контуров и сплайн-интерполяции.

Проблема несбалансированного распределения данных между исходным доменом (совокупность изображений) и целевым доменом в задачах семантической сегментации аэрофотоснимков рассматривается в [12]. Авторы предлагают метод обучения модели без использования разметки для целевого домена. Для этого используется GAN, которая адаптирует стилистические особенности исходного домена к целевому. Основная цель этой работы — улучшить точность сегментации на целевом домене, минимизируя влияние различий между доменами, таких как изменения в освещении, ракурсы съемки или текстуры объектов. Авторы отмечают следующие улучшения, достигнутые с помощью предложенного подхода: лучшие обобщения модели, устранение влияния различий между доменами, сохранение мелких деталей и снижение шума.

В статье [13] исследуется применение cGAN для задач семантической сегментации спутниковых изображений среднего разрешения. Авторы фокусируются на улучшении точности выделения объектов на таких изображениях, которые часто страдают от шума, недостаточной детализации и сложных текстур. Основная цель работы — продемонстрировать, как cGAN может быть использована для повышения качества сегментации по сравнению с традиционными подходами, такими как U-Net или DeepLab. Особое внимание уделяется задачам мониторинга землепользования, городского планирования и анализа изменений в окружающей среде. Улучшениями, достигнутыми с помощью предложенного подхода, являются высокая точность для миноритарных классов, чёткие границы объектов, сохранение мелких деталей и устранение шума.

В [14] авторы предлагают новый подход, который использует GAN для улучшения качества предсказанных масок за счёт более точного моделирования границ объектов и восстановления мелких деталей. Основное внимание уделяется задачам, связанным с анализом изображений с высоким разрешением, таких как аэрофотоснимки, медицинские изображения и городские ландшафты. Цель работы — показать, что использование GAN может значительно повысить точность и детализацию

сегментации экземпляров объектов по сравнению с традиционными методами, такими как Mask R-CNN или DeepLab. Улучшение детализации и сглаживание контуров масок являются основными результатами этой работы.

2. Формирование набора данных

Использованный в экспериментальных исследованиях датасет представляет собой совокупность пар (предсказанная маска, точная маска). Предсказанная маска это бинарная маска объекта, полученная с использованием модели Mask R-CNN. Точная маска представляет собой бинарную маску, сформированную из полигона того же самого объекта, извлечённого из json-файла.

Типы и количества объектов, представленных в датасете, а также особенности формирования json-файлов к файлам с изображениями, описаны в [1]. Для отсутствующих в json-файлах полигонов объектов, соответствующие им точные маски формировались (корректировались) вручную. Это позволило создать репрезентативный и согласованный набор данных, включающий как автоматически сгенерированные, так и ручные маски. Все маски датасета были приведены к размеру 256×256 пикселей. Большой размер масок приводит к резкому увеличению вычислительной нагрузки и, соответственно, времени вычислений, меньший – к неровностям контуров предсказанных масок, после увеличения их размера. Увеличение до размера изображения с объектами недвижимости необходимо на этапе визуализации результата с целью совмещения маски объекта и содержащего его изображения.

3. Формирование и исследование модели

3.1. Особенности архитектуры GAN

GAN состоят из двух основных компонент: генератора и дискриминатора, которые работают в состязательном режиме. Основная цель GAN заключается в обучении генератора создавать данные, неотличимые от реальных.

Генератор в процессе обучения принимает случайный шум (обычно из нормального распределения) в качестве входных данных и преобразует его в некоторые данные, например, изображения. Задача генератора – сформировать данные, которые выглядят как реальные. Для этого он обучается минимизировать потери, возникающие при попытке дискриминатора определить подделку. Основными архитектурами генератора являются следующие:

DCGAN (Deep Convolutional GAN) [15]. Простой и эффективный, но может выдавать менее детализированные результаты на сложных данных. Генератор использует транспонированные свёртки для постепенного увеличения разрешения изображения. Начинает со случайного вектора (латентного пространства) и последовательно увеличивает размерность данных до целевого разрешения.

U-Net [16]. Подходит для задач, где важно сохранить пространственные детали, например, сегментация или улучшение масок. Генератор основан на архитектуре U-Net, которая включает энкодер для понижения размерности и декодер для восстановления изображения. Использует механизм пропусков (skip connections), что позволяет сохранять мелкие детали.

StyleGAN [17]. Высокая детализация и контроль над стилем, но требует больших вычислительных ресурсов. Позволяет контролировать различные уровни детализации изображения (например, общую форму и текстуры). Генератор разделен на две части для сопоставления стилей и для синтеза.

Дискриминатор оценивает качество сгенерированных данных. Задача дискриминатора – помочь генератору генерировать более правдоподобные данные (изображения), предоставляя корректные сигналы ошибки. Основные архитектуры дискриминатора:

PatchGAN [3]. Эффективен для задач, где важны границы объектов и мелкие детали. Анализирует изображение по небольшим локальным областям («патчам») вместо всего изображения. Каждый патч классифицируется как реальный или поддельный, что позволяет фокусироваться на локальных деталях.

Многоуровневый дискриминатор [18]. Улучшает качество генерации за счёт более комплексной оценки данных. Использует несколько дискриминаторов, каждый из которых работает на разных уровнях разрешения. Это позволяет оценивать как глобальные, так и локальные характеристики изображения.

Conditional GAN (cGAN) [19]. Наиболее эффективен для задач, где требуется направленная генерация, например, перевод изображений между их группами. Дискриминатор принимает дополнительную информацию (например, метки классов или условные данные) вместе с изображением, что позволяет контролировать тип генерируемых данных.

Учитывая особенности описанных выше архитектур генератора и дискриминатора, для повышения точности масок мелких объектов недвижимости были выбраны архитектуры U-Net и PatchGAN соответственно.

3.2. Формирование модели

Генератор, реализованный на основе архитектуры U-Net, реализует два основных этапа: сжатие (downsampling) и расширение (upsampling).

Этап сжатия уменьшает пространственное разрешение данных. Каждая из 5-ти групп элементов уменьшает размеры изображения в 2 раза и состоит из свёрточного слоя (Conv2d) с размером ядра (kernel_size) 4x4, шагом (stride) 2 и дополнением (padding) 1. За ним может следовать слой BatchNorm2d с функцией активации LeakyReLU и коэффициентом 0.2.

Элементы расширения наоборот, увеличивают пространственное разрешение данных. Элементов расширения в предлагаемом генераторе тоже 5. Каждый такой элемент увеличивает размеры изображения в 2 раза за счёт выполнения транспонированной свёртки (ConvTranspose2d) с размером ядра 4x4, шагом 2 и дополнением 1. Затем следует слой BatchNorm2d (кроме последнего блока) с функцией активации: ReLU (кроме последнего блока, где используется Tanh для ограничения выхода в диапазон [-1, 1]). Организация генератора GAN показана на рисунке 2.

Элемент расширения объединяет свои выходные данные с данными из соответствующего ему элемента сжатия изображения посредством скип-коннекторов. Это позволяет сохранять детали и улучшать качество генерации. На выходе генератора формируется одноканальное изображение размером, равным входному. Описанные выше элементы сжатия и расширения генератора 1-го уровня приведены на рисунках 3 и 4 соответственно.

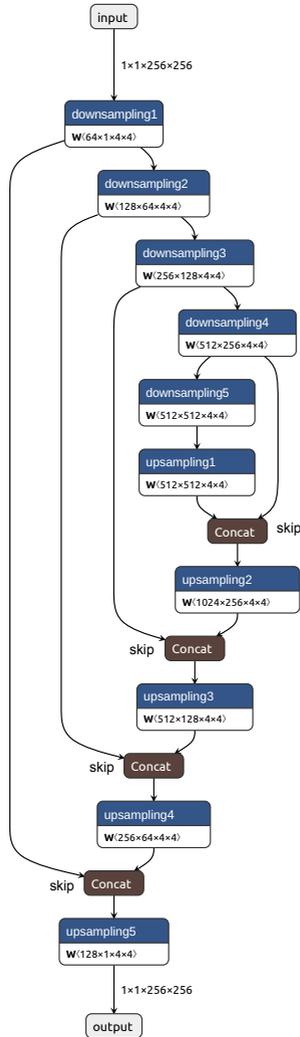


Рисунок 2. Генератор GAN

Как уже было отмечено выше, дискриминатор на основе архитектуры PatchGAN анализирует изображение по небольшим локальным областям (патчам). Каждый патч классифицируется им как реальный

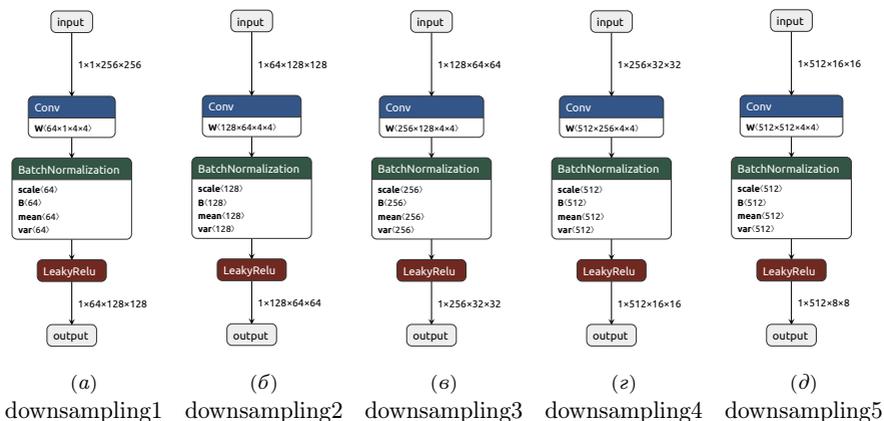


РИСУНОК 3. Элементы сжатия генератора GAN

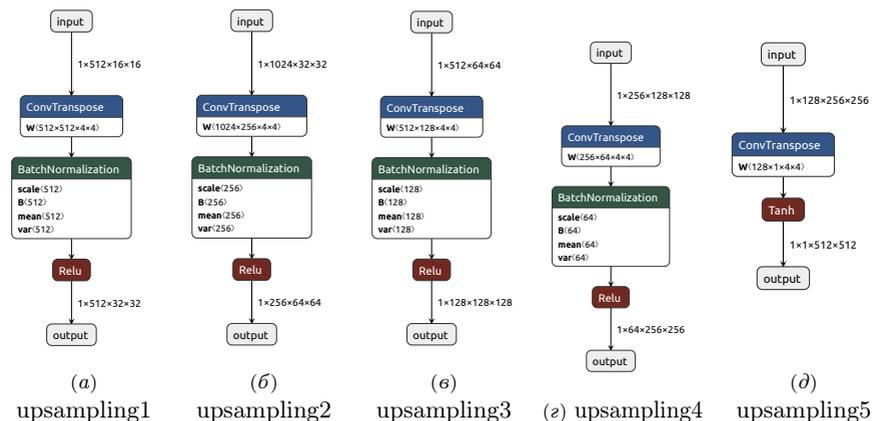


РИСУНОК 4. Элементы расширения генератора GAN

или поддельный, что позволяет фокусироваться на локальных деталях. Дискриминатор принимает два входных изображения: предсказанное изображение (сгенерированное генератором) и реальное изображение (в нашем случае они представляют собой точные маски объектов). Затем эти изображения объединяются по каналам. Последующие несколько свёрточных слоёв (Conv2d) с размером ядра 4x4, шагом 2 и дополнением 1 уменьшают пространственные размеры данных.

Начиная со второго слоя применяется BatchNorm2d с функцией активации LeakyReLU и коэффициентом 0.2. Последний слой с функцией активации Sigmoid вычисляет вероятность того, что входные данные являются реальными. Структура дискриминатора приведена на рисунке 5.

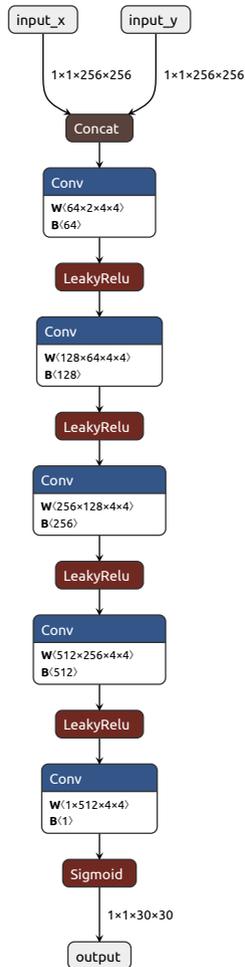


РИСУНОК 5. Дискриминатор GAN

В генераторе используется стандартная функция потерь для задач бинарной классификации `BCEWithLogitsLoss` и `L1 Loss` для минимизации пиксельной разницы между сгенерированными и реальными масками. Первая из этих функций используется для «обмана» дискриминатора, вторая помогает генератору создавать маски, которые близки к реальным по пиксельным значениям. Важным параметром при `L1 Loss` является λ_{L1} , который управляет важностью пиксельной разницы между сгенерированными и реальными масками. Это позволяет балансировать между «обман» дискриминатора и точностью воспроизведения деталей. Функция потерь генератора – `BCEWithLogitsLoss`.

Обучение GAN связано с определёнными проблемами, такими как нестабильность процесса обучения, коллапс мод и низкое качество генерации, что требует применения специальных методов для их решения. Одним из таких подходов является спектральная нормализация, которая ограничивает спектральную норму весов в слоях генератора и дискриминатора, тем самым стабилизируя обучение и предотвращая взрыв градиентов [20]. Другой метод заключается в использовании штрафа за градиент (`gradient penalty`), который добавляет регуляризацию к выходам дискриминатора, обеспечивая более плавное распределение значений и улучшая сходимость модели [21]. Также широкое применение нашёл механизм самовнимания (`self-attention`), позволяющий модели учитывать глобальные зависимости в данных, что особенно полезно при работе с высокоразмерными изображениями [22]. Однако, поскольку восстановление простых бинарных масок является относительно простой задачей, и как показали результаты исследования описанной выше модели GAN, использование этих подходов в данном случае является излишним.

3.3. Результаты исследования модели

Формирование и исследование модели GAN проводилось с использованием фреймворка `PyTorch` в среде `Google Colab Pro`. `Jupyter`-ноутбук (лицензия MIT) [доступен по этой ссылке](#):  `Jupyter`. Исследования основных метрик точности были выполнены для различных значений скорости обучения, количества эпох и размеров батча. Метрики точности генератора

GAN исследовались при $\lambda_{L1} = 100$, поскольку при таком значении их кривые обладают достаточной степенью гладкости, что свидетельствует об отсутствии взрывных или затухающих градиентов. На рисунках 6, 7 и 8 показаны кривые метрик точности для значений $\lambda_{L1} = 80, 100$ и 120 .

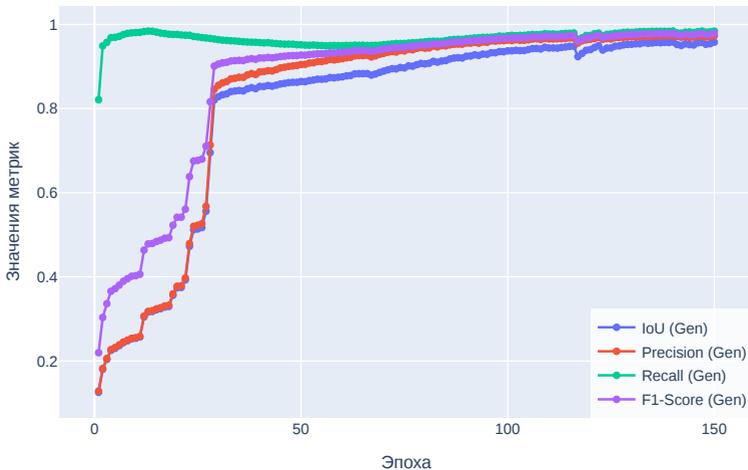
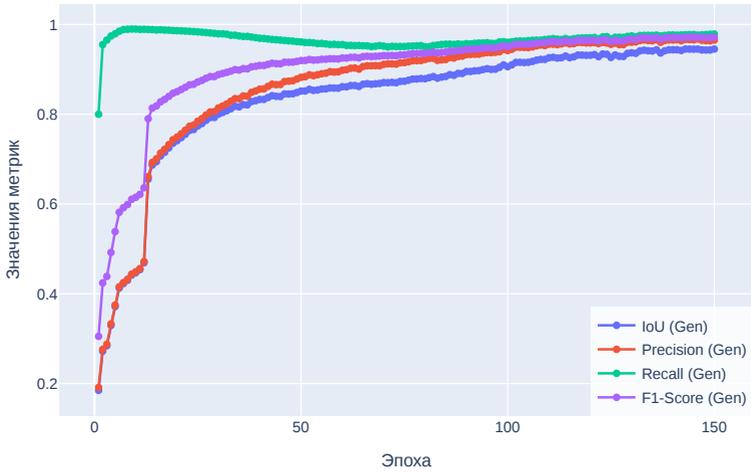


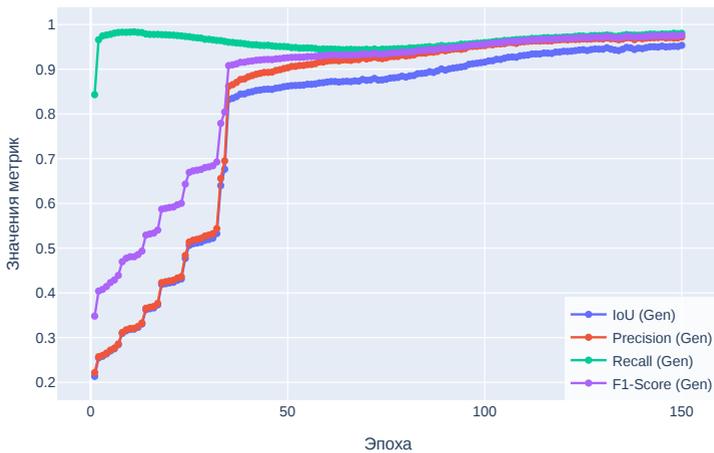
Рисунок 6. Графики метрик точности генератора GAN для $\lambda_{L1}=80$

В таблице 1 приведены результаты исследований основных метрик точности IoU, IoU0.5-95, Precision, Recall и F1-Score. Как показали проведённые исследования, использование GAN позволяет значительно улучшить точность масок сегментации объектов недвижимости, полученных с использованием модели Mask R-CNN [1].

Рассмотрим несколько примеров улучшения масок объектов недвижимости, полученных с использованием модели Mask R-CNN. На рисунке 9 показаны маски, предсказанные моделью Mask R-CNN, и улучшенные маски, сгенерированные с помощью GAN. По значениям метрик точности можно сделать вывод, что использование GAN позволяет значительно улучшить качество масок за счёт коррекции ошибок, таких как разрывы контуров, шум и неточности в границах объектов.



Рисунк 7. Графики метрик точности генератора GAN для $\lambda_{L1}=100$



Рисунк 8. Графики метрик точности генератора GAN для $\lambda_{L1}=120$

Таблица 1. Оценка метрик точности для предсказанных GAN масок в зависимости от параметров обучения модели

Скорость обучения	Размер батча	Эпох обучения	IoU (Pred)	Precision (Pred)	Recall (Pred)	F1-Score (Pred)	IoU0.5-95 (Pred) (Mask R-CNN)	IoU0.5-95 (Pred) (GAN)
1e-6	10	500	0.87	0.92	0.91	0.91	0.67	0.90
5e-6	10	400	0.85	0.90	0.89	0.90	0.65	0.88
5e-6	20	350	0.84	0.89	0.88	0.88	0.64	0.87
1e-5	10	300	0.83	0.88	0.87	0.87	0.63	0.86
1e-5	20	250	0.84	0.89	0.88	0.88	0.64	0.87
5e-5	8	200	0.79	0.84	0.83	0.83	0.59	0.81
5e-5	10	300	0.82	0.87	0.86	0.86	0.62	0.85
5e-5	16	250	0.81	0.86	0.85	0.85	0.61	0.84
4e-5	8	150	0.80	0.85	0.84	0.84	0.60	0.82
4e-5	16	200	0.78	0.83	0.82	0.82	0.58	0.80
1e-4	8	100	0.75	0.80	0.78	0.79	0.55	0.78
1e-4	16	150	0.77	0.82	0.81	0.81	0.57	0.80
1e-4	20	200	0.78	0.83	0.82	0.82	0.58	0.81
2e-4	8	80	0.70	0.75	0.73	0.74	0.50	0.73
2e-4	16	100	0.72	0.78	0.76	0.77	0.52	0.76



Рисунок 9. Результаты улучшения с помощью GAN показаны под предсказанными Mask R-CNN масками объектов недвижимости

Пример улучшения масок объектов на аэрофотоснимке, на которых сложность задачи возрастает из-за высокой плотности объектов и их разнообразия приведён на рисунках 10.

При генерации масок с помощью GAN часто возникает проблема ступенчатых краёв, см. рисунок 9. Это обусловлено спецификой работы GAN, которая связана с дискретностью пикселей при формировании изображений и сложностью точного воспроизведения мелких деталей границ объектов. Для улучшения контуров масок сегментации могут быть применены различные функции морфологического сглаживания, которые позволяют устранить шум, заполнить пробелы и сделать границы объектов более гладкими.

Одним из базовых подходов является использование операций морфологического открытия и закрытия с заданным структурирующим элементом (например, квадратной матрицей (5×5)). Открытие помогает удалить мелкие шумовые артефакты, а закрытие заполняет мелкие разрывы внутри маски. Для дополнительного сглаживания контуров



(а) Обнаруженные моделью Mask R-CNN маски объектов недвижимости



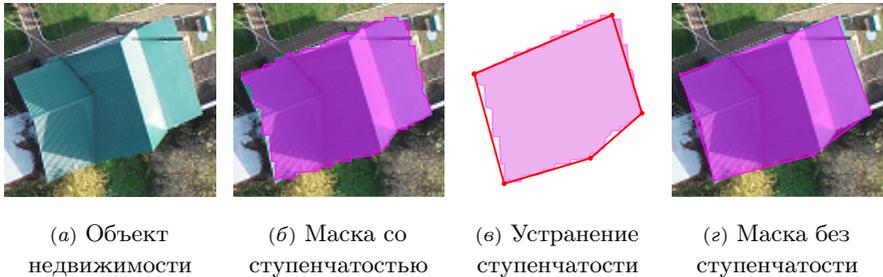
(б) Маски улучшены генератором GAN

РИСУНОК 10. Улучшение масок объектов недвижимости

можно применить Гауссово размытие (GaussianBlur), которое смягчает резкие переходы между фоном и объектом.

Также эффективным методом является использование активных контуров (snakes), которые адаптивно подстраиваются под форму объекта за счёт минимизации энергии упругости и гладкости. Для достижения еще более гладких контуров может быть применена сплайн-интерполяция, которая аппроксимирует исходный контур с помощью кубических сплайнов, обеспечивая равномерное распределение точек и плавность границ.

Ещё одним из эффективных способов устранения ступенчатости является использование алгоритма Дугласа-Пекера, который специализируется на упрощении полилиний и сглаживании контуров [23]. Особенность данного алгоритма заключается в его способности анализировать каждую точку границы маски и оценивать ее значимость для общей формы объекта. Точки, которые вносят минимальный вклад в конфигурацию контура, удаляются, что позволяет получить более плавные и естественные границы без потери ключевых геометрических особенностей. Пример использования алгоритма Дугласа-Пекера приведён на рисунке 11.



(а) Объект
недвижимости

(б) Маска со
ступенчатостью

(в) Устранение
ступенчатости

(г) Маска без
ступенчатости

Рисунок 11. Устранения ступенчатости маски с использованием алгоритма Дугласа-Пекера

Эти методы или их комбинация позволяют в ряде случаев улучшить качество масок, делая их более пригодными для последующего анализа или визуализации. Как показали проведённые исследования, использования методов морфологического сглаживания и улучшения границ масок с использованием алгоритма Дугласа-Пекера в постобработке масок, даёт возможность повысить точность их формирования в среднем на 5–7%, что особенно важно для задач компьютерного зрения и сегментации объектов на аэрофотоснимках в частности.

Заключение

В данной работе была предложена и исследована модель GAN, предназначенная для повышения точности масок сегментации, полученных с использованием модели Mask R-CNN. Эксперименты показали, что применение GAN позволяет существенно улучшить качество детектированных масок объектов недвижимости за счёт коррекции ошибок, таких как разрывы в масках, шумовые артефакты и неровные контуры. Улучшение достигается за счёт использования функции потерь L1 для минимизации пиксельных различий между целевыми и сгенерированными масками, а также за счёт работы дискриминатора, который способствует более реалистичной генерации масок.

Результаты проведенных исследований демонстрируют улучшение метрик точности, таких как IoU, Precision, Recall и F1-Score, для улучшенных масок по сравнению с исходными предсказаниями Mask R-CNN. Среднее значение IoU_{0.5-95} увеличилось в среднем на 30-35%, что подтверждает эффективность предложенного подхода. Дополнительно было отмечено, что использование морфологических операций и методов сглаживания контуров позволит дополнительно улучшить гладкость и точность границ масок.

Полученные результаты имеют важное практическое значение и будут использованы в ИС ППК «Роскадастр» для автоматического преобразования растровых изображений с объектами недвижимости в векторные карты. Это позволит повысить точность и скорость формирования цифровых карт, что особенно важно для задач кадастрового учета и управления территориями. Предложенный подход может быть адаптирован для других приложений, связанных с обработкой изображений и компьютерным зрением, где требуется высокая точность сегментации объектов. Проведённое в работе исследование подтверждает перспективность применения GAN с целью улучшения качества масок объектов недвижимости, что позволяет его эффективно использовать для автоматизации процессов анализа и преобразования данных в геоинформационных системах. Дальнейшие исследования могут быть направлены на оптимизацию параметров модели, внедрение новых методов постобработки масок и расширение области применения GAN.

Список использованных источников

- [1] Винокуров И. В. *Использование модели Mask R-CNN для сегментации объектов недвижимости на аэрофотоснимках* // Программные системы: теория и приложения.– 2025.– Т. 16.– № 1(64).– С. 3–44. [↑137, 144](#)
- [2] G. Cohen, R. Giryes *Generative adversarial networks.*– 2024.– 28 pp. [arXiv:2203.00667](#) ↑
- [3] P. Isola, J.-Y. Zhu, T. Zhou, A. A. Efros *Image-to-image translation with conditional adversarial networks.*– 2016.– 17 pp. [arXiv:1611.07004](#) [↑134, 138](#)
- [4] T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, A. Tao, J. Kautz, B. Catanzaro *High-resolution image synthesis and semantic manipulation with conditional GANs.*– 2017.– 14 pp. [arXiv:1711.11585](#) [↑135](#)
- [5] C.-H. Lee, Z. Liu, L. Wu, P. Luo *MaskGAN: Towards diverse and interactive facial image manipulation.*– 2019.– 20 pp. [arXiv:1907.11922](#) [↑135](#)
- [6] Y. Xue, T. Xu, H. Zhang, L. Rodney Long, X. Huang *SegAN: Adversarial network with multi-scale L_1 loss for medical image segmentation.*– 2017.– 9 pp. [arXiv:1706.01805](#) [↑135](#)
- [7] X. Chen, C. Xu, X. Yang, D. Tao *Attention-GAN for object transfiguration in wild images.*– 2018.– 18 pp. [arXiv:1803.06798](#) [↑135](#)
- [8] J.-Y. Zhu, T. Park, P. Isola, A. A. Efros *Unpaired image-to-image translation using cycle-consistent adversarial networks.*– 2017.– 18 pp. [arXiv:1703.10593](#) [↑135](#)
- [9] J. Gong, J. Xu, X. Tan, J. Zhou, Y. Qu, Y. Xie, L. Ma *Boundary-aware geometric encoding for semantic segmentation of point clouds.*– 2021.– 9 pp. [arXiv:2101.02381](#) [↑135](#)
- [10] L. Xu, M. Gabbouj *Revisiting generative adversarial networks for binary semantic segmentation on imbalanced datasets.*– 2024.– 14 pp. [arXiv:2402.02245](#) [↑135](#)
- [11] R. Abdelfattah, X. Wang, S. Wang *JPLGAN: Generative adversarial networks for power-line segmentation in aerial images.*– 2022.– 11 pp. [arXiv:2204.07243](#) [↑136](#)
- [12] B. Benjdira, Y. Bazi, A. Koubaa, K. Ouni *Unsupervised domain adaptation using generative adversarial networks for semantic segmentation of aerial images* // Remote Sens.– 2019.– Vol. 11.– No. 11.– id. 1369.– 23 pp. [arXiv:1905.03198](#) [↑136](#)
- [13] A. Kulkarni, T. Mohandoss, D. Northrup, E. Mwebaze, H. Alemohammad *Semantic segmentation of medium-resolution satellite imagery using conditional generative adversarial networks.*– 2020.– 7 pp. [arXiv:2012.03093](#) [↑136](#)
- [14] Q. H. Le, K. Youcef-Toumi, D. Tsetserukou, A. Jahanian *GAN Mask R-CNN: Instance semantic segmentation benefits from generative adversarial networks.*– 2020.– 13 pp. [arXiv:2010.13757](#) [↑136](#)
- [15] A. Radford, L. Metz, S. Chintala *Unsupervised representation learning with deep convolutional generative adversarial networks.*– 2016.– 16 pp. [arXiv:1511.06434](#) [↑138](#)
- [16] O. Ronneberger, Ph. Fischer, Th. Brox *U-Net: Convolutional networks for biomedical image segmentation.*– 2015.– 8 pp. [arXiv:1505.04597](#) [↑138](#)

- [17] T. Karras, S. Laine, T. Aila *A style-based generator architecture for generative adversarial networks.*– 2018.– 12 pp.  arXiv  1812.04948 ↑138
- [18] T. Karras, T. Aila, S. Laine, J. Lehtinen *Progressive growing of GANs for improved quality, stability, and variation.*– 2017.– 26 pp.  arXiv  1710.10196 ↑138
- [19] M. Mirza, S. Osindero *Conditional generative adversarial nets.*– 2014.– 7 pp.  arXiv  1411.1784 ↑138
- [20] T. Miyato, T. Kataoka, M. Koyama, Y. Yoshida *Spectral normalization for generative adversarial networks.*– 2018.– 26 pp.  arXiv  1802.05957 ↑143
- [21] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, A. C. Courville *Improved training of Wasserstein GANs.*– 2017.– 20 pp.  arXiv  1704.00028 ↑143
- [22] H. Zhang, I. Goodfellow, D. Metaxas, A. Odena *Self-attention generative adversarial networks.*– 2019.– 10 pp.  arXiv  1805.08318 ↑143
- [23] H. Chen *An improved Douglas-Peucker algorithm applied in coastline generalization, Fourth International Conference on Geology, Mapping, and Remote Sensing (ICGMRS 2023) (14–16 April 2023, Wuhan, China), Proc. SPIE.*– vol. **12978.**– 2024.– id. 1297804.  ↑149

Поступила в редакцию	21.04.2025;
одобрена после рецензирования	10.06.2025;
принята к публикации	11.06.2025;
опубликована онлайн	28.06.2025.

Рекомендовал к публикации

д.ф.-м.н. А. М. Елизаров

Информация об авторе:



Игорь Викторович Винокуров

Кандидат технических наук (PhD), ассоциированный профессор в Финансовом Университете при Правительстве Российской Федерации. Область научных интересов: информационные системы, информационные технологии, технологии обработки данных



0000-0001-8697-1032

e-mail:

Декларация об отсутствии личной заинтересованности: *благополучие автора не зависит от результатов исследования.*

Том 16

Выпуск 2(65)

2025 г.

АВТОРСКИЙ УКАЗАТЕЛЬ

Абрамов Николай Сергеевич*Особенности организации единого научно-технологического пространства* **55, 76****Бельтюков Анатолий Петрович***Сложность вычислений с путешествиями во времени* **3, 54****Винокуров Игорь Викторович***Повышение точности сегментирования объектов с использованием генеративно-состязательной сети* **111, 152****Джудакизаде Милад***Сложность вычислений с путешествиями во времени* **3, 54****Стецюра Геннадий Георгиевич***Скоростные ассоциативные операции и алгоритмы управления и обработки данных для распределенных цифровых систем, действующих в условиях неопределенности* **81, 106****Фраленко Виталий Петрович***Особенности организации единого научно-технологического пространства* **55, 76**

VOL. 16 ISSUE 2(65) 2025

AUTHOR INDEX

Abramov Nikolai Sergeevich

Features of the organization of an unified scientific and technological space . 55, 76
Beltiukov Anatoly Petrovich

Complexity of Computations with Time Travel 3, 28
Fralenko Vitaly Petrovich

Features of the organization of an unified scientific and technological space . 55, 76
Joudakizadeh Milad

Complexity of Computations with Time Travel 3, 28
Stetsyura Gennady Georgievich

High-speed associative interaction of distributed digital objects operating under uncertainty 81, 106
Vinokurov Igor Victorovich

Improving the accuracy of segmentation masks using a generative-adversarial network model 111, 131

VOL. 16 ISSUE 2(65) 2025

CONTENTS

Research Article

THEORETICAL FOUNDATIONS OF SOFTWARE SYSTEMS

MILAD JOUDAKIZADEH, ANATOLY P. BELTIUKOV[✉]. *Complexity of Computations with Time Travel (In Engl., In Russ.)* **3–28**, 29–54

Research Article

APPLIED SOFTWARE SYSTEMS

NIKOLAI S. ABRAMOV[✉], VITALY P. FRALENKO. *Features of the organization of an unified scientific and technological space (In Russ.)* 55–76, **77–79**

Research Article

HARDWARE AND SOFTWARE FOR DISTRIBUTED AND SUPERCOMPUTER SYSTEMS

GENNADY G. STETSYURA[✉]. *High-speed associative interaction of distributed digital objects operating under uncertainty (In Russ.)* 81–106, **107–109**

Research Article

ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

IGOR V. VINOKUROV[✉]. *Improving the accuracy of segmentation masks using a generative-adversarial network model (In Engl., In Russ.)* . **111–131**, 132–152

Author index **154**

Чтобы сменить язык страницы, кликните, пожалуйста, флаг в верхнем углу

Авторский указатель 153

Содержание 2

