UDC 004.7+004.056

[1]Zhukov I.A., doctor of engineering sciences,
orcid.org/0000-0002-9785-0233,
[1]Pechurin N.K., doctor of engineering sciences,
orcid.org/0000-000-1727-7455,
[2]Kondratova L.P., candidate of engineering sciences,
orcid.org/0000-0002-9170-4198

THE BIAS OF THE INITIAL INFORMATION FLOWS INTENSITIES ESTIMATES OF THE GLOBAL CYBER-PHYSICAL NETWORK

[1]National Aviation University
[2]National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute"

zhuia@ukr.net
nkpech@i.ua
ljupav@ukr.net

Introduction

To paraphrase a long-standing expression of Robert Mario Fano on the computer and communication systems, we say that. The marriage between the Internet of people and the Internet of things has come to pass, the honeymoon is over, and the spouses realized that they are a single whole, changed their names to Cyber-Physical Networks (CPS) and began to live and live and make good money (fig.1 based on data from [1-3]).
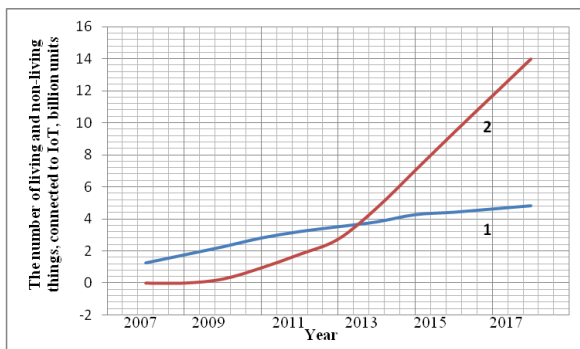

Fig. 1. The number of living things (curve 1) and non-living thing (curve 2), connected to Internet on depending from time

As you can see, the scale of the development of the global cyber-physical system is such that it makes sense to try to predict what will happen to us, and at least evaluate the quantitative parameters of this relatively new object of information technology. Many works are devoted to the estimation of the initial, from existing or potential subscribers, information load on a relatively new component of the CPS – the Internet of Things (IoT), caused by the great interest of society and the scientific community to predict the development of this relatively new (in the public sense) object, which is the birthplace of a completely new, but almost legalized phenomenon – global systems of the artificial intelligence as one of the components of the CPS. The horde of artificial intelligence systems occupies the space of the global information network.

The more accurately we assess the dynamics of the initial information from the horde of "brainless" sources, the more likely we will be able to correctly orient ourselves and develop an effective strategy of behavior (in particular, protection against this invasion), not being limited, as was the case with a horde of another type, washing hands and wearing a mask.

This is not the first time that the authors address the problem of forecasting the development of CPS and IoT, in particular, the problem of assessing initial information [4]. In this opus, we will describe some factors, the consideration of which can raise the accuracy of the estimates obtained by the methods proposed earlier.

The problem statement in detecting the bias (or lack thereof) of the estimates of the initial information flows intensities for the Internet of Things in the context of one method to evaluate

In [4], considerations are presented regarding (one) method of estimating the intensities of initial information flows of the Internet of Things, based on the use of the indicator – gross domestic product (GDP).

This indicator is widely used in socio-economic and other areas, is monitored at the level of individual manufacturers (generators of things), at the state level and globally ([5-7]); that is why in the mentioned work it is proposed to use the GDP indicator to solve the problem on estimating the intensities of the Internet of Things initial information flows. In this case, it is a natural assumption that the initial information flows intensity (to) the Internet of Things directly depends on the intensity of creation (production) of things themselves, which is estimated by GDP.

Indeed, the carriers of information coming into the network of things from the terminal devices of end users per unit time are the "things" themselves, i.e., material objects provided by i/o devices, the role of which can be played by various software and hardware, in particular, actuators and (intelligent) sensors [8]. You can even make assumptions about the direct linear dependence of the information amount generated by terminal devices and coming into the network per unit time (flow intensity), from the number of things-terminals: the more things are produced per unit time, the greater the information load on the network.

Indeed, in [4] it is proposed to estimate the flows intensities of "things", on the basis of which data on the intensities of information flows are obtained.

Next, when using time series tools to predict the intensity of initial information from the i-th object to the Internet of Things, to solve the problem it is necessary to identify the function – the dependence of the intensity of initial information to the Internet of Things on GDP_i.

In addition, it is necessary to determine the parameters of auto- and cross-correlation.

In the aforementioned work, it is proposed, in fact, to calculate the cross-correlation parameters based on Vasily Vasilyevich Leontyev model of the interbranch balance, which is formed as follows.

We have n generators (producers) things as which can act as we have said, objects of different levels (individual producers, industries, etc.); for certainty, the state-level economy can be considered as such an object. The object generates a flow of homogeneous gross product with the intensity of X_i (conventional units/year); it is assumed that the flow intensity function of initial information from the values of X_i is known.

The so-called technological matrix of the classical V.Leontyev model is the carrier of estimates of the correlation between GDP_i and GDP_j ($i, \ j \in \{1,2,\dots,n\}$).

Thus, we have random components of the evaluation procedure, there is a correlation between variables, whence the suspicion of the possibility of making mistakes when using classical statistical research methods, which can make dangerous the use of good, old – since Simon Smith Kuznets – indicator.

Let's try to identify places where there may be shifts in the slender design of the procedure for estimating the intensities of initial information flows that it enters to the CPS and IoT.

Solving the problem of detecting the bias of estimates of the intensities of initial information flows

Estimation of input flow intensities measured by the number of devices connected to the CPS (accumulated, living things plus unliving things) as a function of accumulated product flows (AP).

According to data from reliable sources such as [6], every year during the period, say, from 1948 to 2020, an ordinary citizen of the country S produced a product (volume) X_s, that is 2 times more than a citizen of the country G.

Suppose, for simplicity, that at the beginning of 1948 (the beginning of the Bretton Woods calculation era for the product

produced per capita), the AP volume for all these citizens was zero. Then it is possible to state that the AP volume of the country S citizen in 2 times exceeds that of the citizen of the country G. In other words, at the end of 2020 the first created sources of information load on the CPS and IoT with an intensity twice that of the second.

We have statistics as in Table 1, where it is assumed that the "accumulated" GDP at the beginning of 2009 is equal to zero.

Table 1. Statistics on the number of devices connected to the CPS

Year	AP Volume of world, USD billion	Number of devices connected to CPS (accumulated, living things+non-living things), billion units
2009	59803,0	2,5
2010	125152,0	3,75
2011	197526,0	5,0
2012	271157,0	6,25
2013	346723,0	8,5
2014	424760,0	11,3
2015	498262,0	13,8
2016	573475,0	16,3
2017	653745,0	18,8

We have a close correlation, however, in reality, a citizen of country S has not accumulated twice as many things (recall, sources of initial information flows in IoT) than a citizen of country G: this is evident from the statistics of connecting things to the Internet.

The revealed bias is explained by the fact that the (additive) component of GDP – services are not accumulated in the form of material objects, except perhaps in the form of credit cards. For 2017, the world has accumulated a product in the amount of 653745.0 billion dollars, and this product consists of goods (the "source" of things for IoT) and services (measured and accumulated in the form of credit cards, and these services are not generating things for IoT).

However, here we are breaking into an open door, because the indicator "volume of AP" is not used so simply: two components of GDP are monitored – the volume of accumulated goods and separately – the volume of services provided.

Now we will try to localize the problem by eliminating the problem of accumulation and considering only the increment (say – annual) products (actual GDP) and are connected to the global CPS interface devices (from the living and non-living things).

To do this, we use the data given in Table 1 to assess the statistical correlation between the increment in the "accumulated" GDP of the world, i.e., the actual GDP (in billion dollars), and the annual increment of the total (summary) number of devices connected to the global computer network (billion units). Here [The annual increment of the total (summary) number of connected in 2009 devices to the global computer network] = [The total number of connected devices to the global computer network in 2009] - [The total number of connected devices to the global computer network in 2008]. Based on the data in Table 1, Table 2 is formed.

Here, and this is natural, there is a close statistical relationship (the value of the correlation coefficient = 0.86): the more product is produced per year (GDP per capita), the more interface devices connect this manufactured product to the global computer network.

All the above considerations regarding the accumulated GDP remain valid for the GDP itself.

Table 2. Statistics obtained on the basis of Table 1

Year	GDP of world, USD billion	The annual increment of the total (summary) number of connected devices to the global CPS, billion units
2007	57520,0	-
2008	63026,0	0,5
2009	59803,0	0,75
2010	65349,0	1,25
2011	72374,0	1,25
2012	73631,0	1,25
2013	75566,0	2,25
2014	78037,0	2,8
2015	73502,0	2,5
2016	75213,0	2,5
2017	80270,0	2,5

As we have already said, there are two types of interface devices of the global CPS: the first is the interface devices of the IoT itself (the interface of "non-living things"), and the second is the interface devices of the Internet of people ("living things", interface devices for subscribers – people are various gadgets type of phone, smart devices, etc.). And it is the first indicator that characterizes the annual increase in the flow intensity of the initial information from a citizen of any country in the IoT.

So, if indeed the annual increase in the "thing allowance" for a citizen of country S is 2 times greater than that for a citizen of country G (due to twice the "high" technologies available in country S than in country G), then the increase in the interface devices of a citizen S for IoT is more than the same from citizen G by the same 2 times.

Let's repeat our statement made at the beginning of the article: some interface devices for the Internet of people can also be considered as generators of the initial information in the IoT.

Now let's introduce an assumption: the number of devices connected to the IoT per year per one inhabitant of the planet is determined solely by the number of things produced by this inhabitant (one person) per year.

At one time, the State Planning Committee, to estimate the number of things produced in a year or five years, introduced a heterogeneous scale; homogeneous scale (c.u.) appeared as an alternative. If the global economic system were homogeneous, then it would be fair: if the number of products doubled, then the number of interface devices should also double (again, up to the pace of IoT installers).

A convenient indicator for estimating the number of interface devices connecting things to the global computer network could be the in everyday use indicator "the total number of devices (users) connected to the global CPS (Internet). The danger of erroneous conclusions follows from the fact that the modern Internet is a combination of two networks: the Internet of things and the old customary Internet of people with their (people) interface devices such as radio towers, laptops, tablets, iPads, and other gadgets. And we are interested in the part that unites all material objects (things, material products), but does not include the person himself.

Let's try to estimate the correlation when using not the total number of devices connected to the global network, but excluding smartphones, PCs, wearable devices (live Internet) connected to the Internet, leaving Smart TV, industrial interface devices, vehicle devices, household appliances in consideration (Table 3, assuming zero value of the connected interface devices of this type for 2008).

We have for a pair of (<world GDP, billion dollars>, <the annual increment in the number of Smart TVs connected to the global

computer network, industrial interface devices, vehicle devices, household appliances, etc., billion units>), a correlation coefficient equal to 0,84.

Table 3. Statistics on the number of non-living devices connected to the CPS

Year	GDP of world, USD billion	Annual increment in the number of connected devices
2007	57520,0	-
2008	63026,0	0,0
2009	59803,0	0,25
2010	65349,0	0,7
2011	72374,0	0,8
2012	73631,0	0,95
2013	75566,0	1,9
2014	78037,0	2,35
2015	73502,0	2,25
2016	75213,0	2,2
2017	80270,0	2,2

However, for a pair of (<world GDP, billion dollars>, <annual increment in the total number of connected devices to the global computer network, billion units>), the value of the correlation coefficient is 0.86 (see above).

Thus, if the independent variable represents the classical GDP, then statistically the amount of the product produced, measured by the number of corresponding interface devices of its material (product) part, to a lesser extent (0.84 < 0.86) depends on the amount of the product measured by the number of all interface devices; if the number of devices for non-living things were in a certain proportion with the number of devices for the set (non-living things + living things), then the corresponding coefficients of (linear) correlation would not differ so much from each other.

This raises further vague doubts about the truth of the generally accepted (in everyday life) assumption that S, which "gave out on the mountain in 2017" is twice the product measured by Bretton Woods (GDP), is more than G, exactly the same time more replenished its individual grocery warehouse.

Such doubts give rise to the task of identifying the data distortion on the dynamics of global and regional cyber-physical systems, and ways to correct these distortions (bias of estimates).

The first way is to establish fair, "correct" mechanisms for assessing the volume of the annual produced product (GDP), taking into account that the emission of any national currency should not directly, roughly influence the volume of the world universal currency: either establish a clear border, or even deduce the existing, issued in the national financial and economic system, a unit generally from consideration when evaluating things that the average citizen of the world has produced or accumulated – the beneficiary of information flows into the global CPS and IoT.

The second way is to use the good old natural indicators from the State Planning Commission, but when establishing reducing mappings, in particular – reducing coefficients, which can be indicators such as "the degree of ordering of things". In the paper [11] we proposed to use principle of material balance in the form of V.Leontyev' model:

$$(E-A)\cdot M - M^r = 0,$$

were M is the material resources values, M^r is the rest of material resources, after technological transformations, which is the main (stationary) sources of the initial information flows.

And, from this equation, if M takes a little value, then M^r takes a little value also; but really, we know vice versa examples.

Consequently, $M^r = M^{r1} + M^{r2}$, where M^{r1} represents the own (inside) rest of the material resources, and M^{r2} represents the outside rest of the material resources, which is not the sources of the initial, from the determined object, information flows. The variables for balancing are in A; one of the ways to form the balance for the i-th object is the

variable a_{ii} raising, this is equivalents to (artificial) lowing of the i-th object technological effectiveness (the citizen S technological effectiveness is by 2 times higher, than the same for citizen G – see above).

The third way to take into account the bias is as follows.

In order to reduce the error in assessing things included in the CPS, the value of GDP should be shifted downward, but how much it is a separate task.

The considered methods on accounting for the displacements were based on the assumption of the independence of objects – generators of things. However, in reality the generators (manufacturers) are in communication. The carrier of these bonds is the technological matrix (A), which also takes into account the hidden part of the things flows that are formed in the actual production process, as well as consumed inside the object; element a_{ij} its essence is the volume of goods and services generated by the i-th object, which is used to create a unit of volume of the goods and the services generated by the j-th object. Expected, for time t, values of X_i ($i \in \{1, 2, \ldots, n\}$), that is, indirect estimates of the intensities of the initial information flows for the Internet of Things are calculated by solving SLAE, where the coefficients of the matrix A are usually presented as the expected value of a_{ij} on the basis of relevant statistics.

Regularities in the dynamics of the values of a_{ij}, except perhaps the downward trend due to technological improvements, are difficult to determine, especially given the "Bretton Woods" method of (comparative) evaluation of technology efficiency. (Incidentally, the trend of declining a_{ij} values due to "technological progress" has also failed in recent times due to the pandemic).

Also, keep in mind that GDP_i, that means – and X_i, usually considered as consisting of two parts: actually, the things and services, true, the additive nature of the convolution allows us to identify the component of GDP associated with the production of things, but to measure the volume of goods produced, semi-voluntaristic determination of the things

price in GDP reduces the accuracy in the assessment.

The authors believe that the main (only to what extent – they do not know) factor is the flow of things, although the "service" of mining from an essentially "mindless device" can also create a significant load on the network.

These circumstances make us consider the values of the components of the technological matrix to be random rather than deterministic.

Then the calculation (estimation) of the predicted values of X_i takes place in the sequence: forming a sample to estimate the elements of the technological matrix, calculating the expected values of the elements of the technological matrix, solving a system of (algebraic) equations (in the simplest, classical, case – linear) or solving the corresponding extreme problem (in the simplest case – linear programming).

Solutions (roots) are random variables with unknown a priori distribution laws, and there is no basis for claiming that their values are always unbiased: instead of really random values a_{ij} we used the estimated coefficients of the variables in the technological matrix.

In reality, such an assumption is not acceptable, but is forced, applied for simplicity. The simplicity is achieved by the fact that instead of considering the design parameters as random (stochastic) quantities, we use their expected (arithmetic mean, mode, median) value. Sometimes even – limit values (time of the greatest loading – TGL).

Such a simplification can generate – and in our case, a shift in the value of the (target) function for estimating the intensity of the initial flow due to a known fact: the value of the function (reflection) from the mathematical expectation of arguments does not always coincide with the mathematical expectation of function values from random arguments.

Numerous calculations using available data from sources [1-3,9,10] confirm the concern: there is a deviation from the estimate. From here we can make a suggestion: to taking into account the identified shift we must

go from the proposed in [4] simple deterministic models of the mathematical programming based on GDP and the appropriate methods on estimating and forecasting [12] to the models and the methods of the stochastic programming.

Conclusion

1. The use of the existing official statistical reporting for forecasting the development of the network, which provides deterministic estimates of the initial parameters, creates the danger of distorting the corresponding estimates of the initial information flows intensities generated by subscribers.

2. The mapping (extreme balancing function in the proposed estimation method) of the value set of initial parameters to the set of the estimates values is such that there is a bias in the intensities estimates of the initial information flows obtained by the method on using the indicator of GDP.

3. In order to improve the accuracy of the Internet of Things development forecast in the context of the initial information flows dynamics, for taking into account the bias of estimates, it is appropriate to move from those proposed in [4] simple deterministic models of the mathematical programming based on GDP and corresponding estimation methods to the models and the methods of the stochastic programming. The question of the feasibility of this suggestion, i.e., the possibility of finding a way to generate relevant statistics on GDP, is beyond the scope of this article.

References

1. Dynamics of the world's GDP from 1970 to 2016. [Electronic resource]. – Available at: https://seosait/com/dinamika-vvp-mira-1970-2016.

2. Internet of Things, IoT, M2M world market. // TAdviser. [Electronic resource]. – Available at: www. Tadviser.ru/index.php/Статья:Интернет_вещей,_IoT,_M2M_(world_market).

3. Internet access (world market) // TAdviser. [Electronic resource]. – Available at: www.tadviser.ru/index.php/Статья:Интернет-доступ_(мировой_ рынок)]).

4. *Zhukov I.A., Pechurin N.K., Kondratova L.P., Pechurin S.N.* Estimation of intensities of the primary information flows generated by subscribers of Internet of things // Problems of informatization and management: a collection of scientific papers. – 2019. – Вип.1 (61). – Р.26-32.

5. State Statistics Service of Ukraine. [Electronic resource]. – Available at: www.ukrstat.gov.ua.

6. The World Factbook / Central Intelligence Agency. [Electronic resource]. – Available at: www.cia.gov.

7. World Trade Organization. – Electronic resource. [Electronic resource]. – Available at: stat.wto.org.

8. *Voitovich I.D.* Intelligent sensors // Ukraine. Science and culture. – 2009. – Issue 35. – P. 106-111.

9. List of countries by GDP sector composition: Difference between revisions. [Electronic resource] – Available at: https://en.wikipedia.org/wiki/List of Counries by GDP sector composition#List by Alphabetical_Order.

10. *Tipanov V.V.* Evolution of international trade: different time horizons, aspects and trends. [Electronic resource] – Available at: https://cyberleninka.ru/article/n/evolyutsiya-mezhdunarodnoy-torgovli-razlichnye-vremennye-gorizonty-aspekty-i-tendentsii-1/viewer.

11. *Igor Zhukov, Nickolay Pechurin, Lyudmila Kondratova.* IoT'S Primary Information Load Estimation // Proceedings 2020 IEEE 11th International Conference on Dependable Systems, Services and Technologies DESSERT, Ukraine, Kyiv May 14-18 2020. – P. 331-335.

12. *Pankratova N.D., Kondratova L.P.* System Strategy for Guaranteed Complex Engineering System Functioning in Real Operating Conditions//Journal of Automation and Information Sciences, 2019, Vol. 51, Issue 1. – P. 15-25.

Жуков І.А., Печурін М.К., Кондратова Л.П.

ЗМІЩЕНІСТЬ ОЦІНОК ІНТЕНСИВНОСТЕЙ ПОЧАТКОВИХ ІНФОРМАЦІЙ-НИХ ПОТОКІВ ГЛОБАЛЬНОЇ КІБЕРФІЗИЧНОЇ МЕРЕЖІ

Визначено, що використання для прогнозування розвитку мережі існуючої офіційної статистичної звітності, яка надає детерміновані оцінки висхідних параметрів, створює небезпеку спотворення відповідних оцінок інтенсивностей первинних інформаційних потоків, що вони генеруються абонентами. Відображення (функція екстремального балансування у запропонованому методі оцінювання) множини значень висхідних параметрів на множину значень оцінок є такою, що існує зміщення оцінок інтенсивностей первинних інформаційних потоків, отриманих по методу використання показника ВВП. Доцільним, з метою підвищення точності прогнозу розвитку інтернету речей в контексті динаміки потоків первинної інформації, для врахування зміщення оцінок, є перехід від запропонованих простих детермінованих моделей математичного програмування на основі ВВП і відповідних методів оцінювання до моделей і методів стохастичного програмування.

__Ключові слова:__ оцінка стану, прогнозування, киберфізичні мережі, розвиток інтернету, початкова інформація.

Zhukov I.A., Pechurin N.K., Kondratova L.P.

THE BIAS OF THE INITIAL INFORMATION FLOWS INTENSITIES ESTIMATES OF THE GLOBAL CYBER-PHYSICAL NETWORK

The use of the existing official statistical reporting for forecasting the development of the network, which provides deterministic estimates of input parameters, creates the danger of distorting the corresponding estimates of the initial information flows intensities generated by subscribers. The mapping (function of extreme balancing in the proposed estimation method) of the set of the input parameters values to the set of the estimates values is such that there is a bias in the estimates of the initial information flows intensities obtained by using the GDP indicator. In order to improve the accuracy of the Internet of Things development forecast in the context of the initial information flows dynamics, to taking into account the bias of estimates it is appropriate to move from the simple deterministic models of mathematical programming based on GDP and corresponding estimation methods to models and methods of stochastic programming.

__Keywords:__ condition assessment; forecasting; cyber-physical networks; development of the Internet; initial information.