BUSINESS INFORMATICS

HSE SCIENTIFIC JOURNAL

CONTENTS

Business Informatics is a peer reviewed interdisciplinary academic journal published since 2007 by National Research University – Higher School of Economics (HSE), Moscow, Russian Federation. The journal is administered by School of Business Informatics. The journal is published quarterly.

The mission of the journal is to develop business informatics as a new field within both information technologies and management. It provides dissemination of latest technical and methodological developments, promotes new competences and provides a framework for discussion in the field of application of modern IT solutions in business, management and economics.

The journal publishes papers in the areas of, but not limited to:

✦ data analysis and intelligence systems

✦ information systems and technologies in business

✦ mathematical methods and algorithms of business informatics

✦ software engineering

✦ Internet technologies

✦ business processes modeling and analysis

✦ standardization, certification, quality, innovations

✦ legal aspects of business informatics

✦ decision making and business intelligence

✦ modeling of social and economic systems

✦ information security.

The journal is included into the list of peer reviewed scientific editions established by the Supreme Certification Commission of the Ministry of Education and Science of the Russian Federation.

The journal is included into Russian Science Citation Index (RSCI) database on the Web of Science platform.

International Standard Serial Number (ISSN) 1998-0663.

Editor-in-Chief: Dr. Alexey Golosov – President of FORS Development Center, Moscow, Russian Federation.

БИЗНЕС-ИНФОРМАТИКА

НАУЧНЫЙ ЖУРНАЛ НИУ ВШЭ

СОДЕРЖАНИЕ

«**Б**изнес-информатика» — рецензируемый междисциплинарный научный журнал, выпускаемый с 2007 года Национальным исследовательским университетом «Высшая школа экономики» (НИУ ВШЭ). Администрирование журнала осуществляется школой бизнес-информатики НИУ ВШЭ. Журнал выпускается ежеквартально.

Миссия журнала — развитие бизнес-информатики как новой области информационных технологий и менеджмента. Журнал осуществляет распространение последних разработок технологического и методологического характера, способствует развитию соответствующих компетенций, а также обеспечивает возможности для дискуссий в области применения современных информационно-технологических решений в бизнесе, менеджменте и экономике.

Журнал публикует статьи по следующей тематике:

✦ анализ данных и интеллектуальные системы

✦ информационные системы и технологии в бизнесе

✦ математические методы и алгоритмы бизнес-информатики

✦ программная инженерия

✦ Интернет-технологии

✦ моделирование и анализ бизнес-процессов

✦ стандартизация, сертификация, качество, инновации

✦ правовые вопросы бизнес-информатики

✦ принятие решений и бизнес-интеллект

✦ моделирование социальных и экономических систем

✦ информационная безопасность.

В соответствии с решением президиума Высшей аттестационной комиссии Российской Федерации журнал включен в Перечень российских рецензируемых научных журналов, в которых должны быть опубликованы основные научные результаты диссертаций на соискание ученых степеней доктора и кандидата наук, по следующим группам научных специальностей: 05.13.00 — информатика, вычислительная техника и управление; 05.25.00 — документальная информация; 08.00.00 — экономические науки.

Журнал входит в базу Russian Science Citation Index (RSCI) на платформе Web of Science.

Журнал зарегистрирован Федеральной службой по надзору в сфере связи, информационных технологий и массовых коммуникаций (Роскомнадзор), свидетельство ПИ № ФС77-59805 от 07 ноября 2014 г.

Международный стандартный серийный номер (ISSN) 1998-0663.

Главный редактор: Голосов Алексей Олегович, кандидат технических наук, Президент компании «ФОРС — Центр разработки».

РЕДАКЦИОННАЯ КОЛЛЕГИЯ

IoT safety and security as shared responsibility

Vinton G. Cerf
Vice President and Chief Internet Evangelist, Alphabet Inc.
Address: 1600 Amphitheatre Parkway, Mountain View, CA 94043, USA
E-mail: vint@google.com

Patrick S. Ryan
Strategy & Operations Principal, Alphabet Inc.
Address: 1600 Amphitheatre Parkway, Mountain View, CA 94043, USA
E-mail: patrickryan@google.com

Max Senges
Research Program Manager, Alphabet Inc.
Address: 1600 Amphitheatre Parkway, Mountain View, CA 94043, USA
E-mail: maxsenges@google.com

Richard S. Whitt
Corporate Director for Strategic Initiatives, Alphabet Inc.
Address: 1600 Amphitheatre Parkway, Mountain View, CA 94043, USA
E-mail: whitt@google.com

Abstract

What happens when everyday standalone devices and machines acquire network interfaces? The somewhat obvious result will be an unprecedented number of "things" connected to the Internet. It is less obvious what this means for the governance of the Internet when this occurs. With the "Internet of Things" (IoT) the Internet's loosely coupled governance structures are already adapting to accommodate the evolution of the Internet's use. As the governance structure continues to develop, users' safety must be the first priority for all hardware and software providers. In the context of the Internet of Things, this paper proposes a definition of digital safety as distinct from security and discusses how multistakeholder governance can be applied to address safety challenges. The paper also considers the integration of "old" industries and the transformation of their governance into the multistakeholder model as their products and services are coming online. We consider how the thousands of manufacturers who traditionally produced analog, not-connected physical "things" adapt to become stakeholders in the Internet and how that changes the way that we think about Internet Governance. The particular interest of this paper is how to address safety issues that become much more prominent with the spread of Internet-enabled physical environments.

The authors of this paper have written this project in their personal capacities as an academic contribution. The views reflected may not be the official position of the authors' employer.

Introduction

What happens when everyday standalone devices and machines acquire network interfaces? The somewhat obvious result will be an unprecedented number of "things" connected to the Internet. It is less obvious what this means for the governance of the Internet when this occurs. With the rising phenomenon known as the "Internet of Things"

(IoT) this scenario is unfolding right now, and the Internet's loosely coupled governance structures are already adapting to accommodate it. Billions of devices are increasingly online, beginning with smart phones and now ranging from our cars to video security systems to our thermostats at home [8]. One question is how the thousands of manufacturers who traditionally produced analog, not-connected physical "things" become stakeholders in the Internet and how that changes the way that we think about Internet Governance. The particular interest of this paper is how to address safety issues that become much more prominent with the spread of Internet-enabled physical environments. We also look at the integration of "old" industries and the transformation of their governance into the multistakeholder model.

The Internet has no single stakeholder, and no single category of actors who bear the primary responsibility for its governance. Instead, this is a shared responsibility among all stakeholders. We argue that challenges brought to the Internet by the IoT will be solved by the same multistakeholder mechanism that governs the Internet itself. Indeed, because as the "things" currently on the Internet do not clearly belong to any single stakeholder, governing the Net continues to be a matter of shared responsibility [3]. Additionally this existing governance framework must now take into account machines that become more "autonomous" (i.e. "smart" through machine learning etc.). In this context we believe that the concept of *digital safety* is particularly relevant as a concept for all stakeholder groups' shared responsibility.

1. Moving from digital security to digital safety

The freedom to innovate inherently includes the responsibility to protect the legitimate interests of users and of the integrity of the ecosystem that connects them. This premise holds true especially as the Internet proliferates into physical spaces and as the enabling of next-generation protocols and artificial intelligence continues to grow.

1.1. New players on the field

Billions of new interconnected things provide for many more opportunities to exploit vulnerabilities. These problems are popping up in platforms like Internet-enabled Closed Circuit Television (CCTV) devices. Because today's CCTV devices are digital cameras that process high volumes of video data, they have nontrivial computer processing and memory capacity. These de-

vices can and have been used increasingly as elements of botnets for large attacks. There are more than 245 million security cameras installed in the world, and their significant footprint throughout the Internet has led to various recent denial of service attacks. [9].

1.2. How responsibility is shared today

How do the Internet's denizens share responsibility today? The private sector plays its part. Companies like Cisco, Facebook, Google and Verizon all depend on the trust of users and therefore have strong incentives to maintain data securely, but also to keep children and loved ones safe from dangers of online fraud, identity theft, and predators. Governmental entities, as other stakeholders, hold an important responsibility because of the need to enact laws and regulations to deter and punish bad behavior and enforce norms with laws and regulations as a backstop. Additionally, civil society groups and non-governmental organizations played key roles in the development of the Internet and continue to analyze and evaluate private-sector practices and to establish norms, expectations and principles for behavior online. And crucially, the Internet's users must bear responsibility and learn to use tools that enable a safe and secure online experience.

1.3. Inconvenient tradeoffs and sharing responsibility

Some stakeholder groups may need to trade inconvenience for security and safety. An illustration of this is the use of two-factor authentication. This illustration applies to users and for private-sector entities that employ their use. On the Net, user authentication is usually some combination of three factors: (1) something that the user knows (e.g., a password), (2) something that the user has with them (e.g. a USB key, perhaps an app on a phone), and/or (3) something intimate to the user (e.g., something biometric like a fingerprint or iris scan). Most systems today use just one factor, but the use of two (or even more) factors provides much stronger security. With multiple-factor authentication, a nontrivial additional inconvenience is inserted in the user's experience to help prove that the users are who they say they are.

Because three-factor is less common today, we'll elaborate further on the case of two-factor authentication. Most logins are accomplished by a username and password, a form of single-factor authentication. In certain cases, a *second* factor can be used, and this delivers an additional code to the user by SMS, phone call, by

a USB device or an app. While more secure, this second factor can also bring some inconvenience to users. For example, if a user is in a rush and logs into her bank account and puts in her ID and password, it can be frustrating if she also has to find her phone in order to receive a text with the additional code. It can even be more inconvenient if a company-issued key is required to be inserted in the computer---if an employee does not have the key with her, then the work may not be accomplished. Still, even with these inconveniences, users are increasingly willing to tolerate the struggle (and companies are increasingly requiring it) because they realize that it is intended to help keep them and their data safer online. As such, the use of two-factor authentication is a good example of how a user contributes to the Internet security environment, even at the expense of their personal time and convenience. Said another way, users who are employing two-factor or three-factor authentication have implicitly acknowledged the *shared responsibility* that they have with the private sector in maintaining the safety of their data.

It is important to remember this shared responsibility when deconstructing any security incident. When Sony failed to protect users' data from exposure in a major breach in 2014, Sony was immediately assailed for having abrogated its duty to protect users [15]. To be sure, Sony bears its share of responsibility for the breach, but just as importantly, government, civil society and the technical community should accept some responsibility for having acquiesced to a system that failed in the dramatic way that it did. In the case of the Sony breach, the users may not have been empowered with a failsafe mechanism for recovery from Sony's failure. For example, if Sony's users could have set up 2-factor authentication on their PlayStation's e-payment system, perhaps the breach of Sony's data would have been reduced, if not rendered inconsequential. As a practical matter, it is unlikely that any breach would have been capable of simultaneously breaching the security of the core and the security at each of the millions of users' individual two-factor devices.

It may be the case that two-factor authentication was not an option for Sony's users. In any case, the onus for finding a way forward remains one that includes *all* stakeholders. We believe that the policy environment will develop to the point where technology like two-factor authentication will be a best practice that is so widely adopted that it becomes a *de facto,* compulsory standard, one that users are demanding just as much as civil society or the technical community. In the future, users will be demanding these additional features.

For a standard to become the *de facto* requirement, the cost of *not* implementing the standard will need to clearly be higher than implementing it. How is this calculus determined? It is done in various ways, sometimes by peer pressure, other times by judicial or legislative fiat. By analogy to another field, consider the case of glaucoma testing by optometrists during any regular eyeglass checkup. The chances of any individual getting glaucoma are very low, less than 1%. However, the consequences are very high (resulting in permanent blindness) and the test for glaucoma is extremely inexpensive: only a couple of dollars. This produces a cost/benefit analysis that is clear. For this reason, by a mix of lawsuits, lawmaking and pressure from hospitals, it first became *de facto* and then *de jure* standard that optometrists must include glaucoma tests in their regular checkups. That's why every time a patient gets a regular eyeglass exam there is often (if not always) a glaucoma test (e.g., the dilation of eyes and/or the short pulses of air). We believe that this phenomenon will apply to technologies like two-factor authentication, which are increasingly affordable, and it is increasingly clear that the consequences of breach are too high for society.

Of course, any company or organization clearly has a responsibility to protect their users and to protect their data. This responsibility should not be minimized. However, the many post-incident sirens that rang for more stringent laws and better enforcement mechanisms demonstrate a scenario that is better for sensationalism than it is for practicality. Of course, standards and laws are very important. However, we must not let this rigidity of our rules and systems make us lose focus of a more holistic view of Internet security, one that moves beyond the quality of the lock at the door but emphasizes a more complete view of the role that all stakeholders have in the Internet community. The press sensationalized the attack against Sony, but there was little exposure of the kinds of measures that either Sony or users could have taken to minimize the risk. As it turns out, Sony may have been able to purchase digital certificates with a two-factor option from DigiCert to validate the certificates by a hardware token [19]. If this is the case, even if the leaked certificate information had been made available, a two-factor option would have provided a "failsafe" to minimize the damage. This protective role is a *shared responsibility* to create and maintain a safe and secure environment online. It's not that all stakeholders share the same responsibility at all times, but without question, the responsibility is rarely, if ever, exclusively vested in a single stakeholder's scope of responsibility.

1.4. Digital safety distinguished from online security

We propose a clear distinction between the concepts of *digital safety* and *online security*. We think of digital safety as a principle that protects users from harm which may be caused by increasingly semi-autonomous systems that are amalgamations of services from various developers. An example of this could be the employment of "fail-over" (or perhaps "fail-safe") solutions in case part of the system is dysfunctional. In our model, fail-over systems are not designed to defend from any breach. Instead, they are designed to assume that there may be a breach, and the fail-over kicks in to provide a solution.

By contrast, online security is the work that is put into protecting the system against malicious actors, rather than the ways that the systems act to protect users when there is a problem. Because fail-over is not a defense in and of itself, good security to prevent the need for fail-over is also important. This is a crucial contrast to "traditional" security-development practices, which are more tolerant towards system errors given that the consequence is "only" virtual or informational (e.g., "Alert! There has been a security breach!") rather than protecting users from material and physical harm. Looking at this another way, we propose the following taxonomy to distinguish *safety* from *security:*

✦ *digital safety* is the protection of the user in his or her environment, with technical mechanisms and policies that protect the users from being harmed by improper operation of the device;

✦ *online security* is the protection of the *physical network, operating systems* and *content* from exposure, modification or functional damage, utilizing a combination of software and hardware mechanisms.

In the above taxonomy, both categories could include pre-emptive as well as fail-over action. A garage door opener offers an example in which safety is paramount. Normally, the garage door can be opened with a signal from a smartphone or a radio-based key device. If this fails for some reason, it is important − for safety − that the door can be opened manually by releasing it from the motorized control. This is usually accomplished with a physical release mechanism. Usually this mechanism is only accessible from inside the garage. The failure of the controls does not trap the user but, at the same time, the system is secure in that the door does not open from the outside. The user is safe, the garage is secure.

One can imagine other intelligent devices such as thermostats that can be controlled and monitored re-

motely. For safety, it should not be possible to allow the temperature to be set above some maximum level. The same standard might also apply for a water heater. Even if a command is sent to try to exceed limits, a fail-safe design would inhibit exceeding fixed limits. For security, it should not be possible—or at least, easy—for an unauthorized party to control or monitor the thermostat. One might use two-factor authentication to achieve the security objective.

In short, digital safety focuses on the end user and his or her interests and health, while online security focuses on protecting other aspects of the network or the device itself. Of course, for some device failures, there could be an effect that might cause a safety issue for the user, for example an unauthorized intruder operating a device in a way that causes it to catch fire. Various security challenges are taken to a new level in the context of IoT because of the increased possibility of penetrators to access the actual equipment used to sense, actuate and control it. While these devices are online, and might be penetrated through that path (e.g. local radio, local network), direct access is an additional hazard. Additionally, the software might be altered to do everything it is supposed to do for device operation and then generate spam, denial of service attacks or other harmful attacks. The user would not be aware of the problem if the device operation did not show any signs of tampering. With regard to safety, the real concern is that these highly complex IoT ensembles (consisting of hardware, software, and firmware) may have bugs. Alternatively, a bug-free system could still generate unforeseen constellations and outcomes in such a way as to interfere with safe operation.

1.5. Linking permissionless innovation and shared responsibilities

In an environment of permissionless innovation, any entity is able to utilize the network to bring novel products and services to countless users. This incredible opportunity brings with it, simultaneously, the prospect that permissionless innovation could inflict harm upon users. We suggest that a balance exists, then, beyond the perceived "good" or "bad" of permissionless innovation. Instead, there is a three-way intersection between permissionless innovation, freedom of action, and the accompanying responsibility that everyone shares to protect users. There is ample room at this intersection for the community to devise novel solutions.

One possibility to ensure that innovators can experiment and pioneer new solutions and products would be

to promote a stronger culture and acceptance of beta-testing access to cutting edge early versions of features and products. A "beta test" is a way to support experimentation by interested early adopters by offering a new set of functionalities in exchange for an understanding that things may not work perfectly [5]. However, it is important that this expectation is clearly and transparently communicated to the users (and done with the user's endorsement and ideally, their enthusiasm). A more open approach between the private sector and the users about the nature of their products and the possible risks to them might be one way to limit liability for innovators while promoting experimentation, and to ensure that users are aware and agree to handle the risks. Such a system may never be fully adopted in any law or any court decision, but industry might band together to promote an open standard for beta testing and a system for communicating it to users. This kind of thing exists to some degree in the context of Creative Commons, which publishes a voluntary system that publishers can use to contract with users for the use of their works.

How does the multistakeholder community respond to a call to action if there is a global Internet security incident? One excellent illustration is the Conficker Working Group. In 2009 the Conficker botnet was by far the most sophisticated botnet of its time. In a matter of weeks, the botnet had infected millions of computers, including government, business and home computers in all 193 countries. The Conficker Working Group (*http://www.confickerworkinggroup.org*) was rapidly (and informally) assembled, and made up of private sector actors such as Microsoft and Symantec, academic researchers, and NGOs such as ICANN and the ShadowServer Foundation. Various government representatives also engaged in the groThe Conficker Working Group carried out its work with no formal mandate, and operated under no formal entity structure. It was a loosely coupled group of representatives from all sectors that banded together with the sole purpose of quickly stopping the spread of Conficker. The fact that the group had no mandate, no budget and no formality allowed it to form and to act quickly and deliberately. To stop one of the strains, the collaborators "coordinated with over 100 countries and blocked over 50,000 domains per day" over the course of less than three weeks [7]. This kind of mobilization and responsiveness was made possible because of the loosely coupled and informal nature of the governance system that came together to address the problem.

2. Application to the Internet of Things

For the Internet's first twenty-five years of commercial operation, the network has largely consisted of user-operated end points and servers. This has also been described as the "end to end" model, which means that most "intelligence" occurred at the edge of the network, i.e., at the devices that connect to it. Today, the physical environment itself is coming "alive," and it is thriving and growing in a cloud of networked devices. This is the Internet of Things (IoT): a mental construct, or ordering principle we are using to describe effort to take the Internet to the next physical level. Indeed, the IoT is really just the Internet itself, but with one very different feature: online activities will not be only informational but include interactions with "things" in all contexts of life and often having physical consequences. Additionally these interactions will increasingly be determined by some form of machine intelligence that lets "things" coordinate semi-autonomously.

IoT has been defined as "a pervasive and self-organizing network of connected, identifiable, and addressable physical objects through the use of microprocessors" [18]. IoT's pervasiveness rings in Cisco's recent branding of the "Internet of Everything." This sounds more inclusive as it describes the networked connection of people, process, things, and data. Things are all inanimate physical objects and devices, like sensors, consumer devices, and enterprise assets, while process is how to manage the way that people, data, and things work together [6]. As David Kirkpatrick explained, "we are extending the connectivity of the Net to a much wider world-wide-web that includes the manifold physical objects of the world, letting them speak back to software that makes sense of what those objects tell us" [12]. Thus, as Neil Gross has opined, the Internet may continue to grow around the globe as an "electronic skin" that touches all aspects of life: "In the next century, planet earth will don an electronic skin. It will use the Internet as a scaffold to support and transmit its sensations. This skin is already being stitched together. It consists of millions of embedded electronic measuring devices: thermostats, pressure gauges, pollution detectors, cameras, microphones, glucose sensors, ERGs, electroencephalographs. These will probe and monitor cities and endangered species, the atmosphere, our ships, highways, and fleets of trucks, our conversations, our bodies — even our dreams" [10].

However we choose to describe the future Internet, it is big, and the things using it are no longer only human.

This may seem axiomatic now, and as IoT phenomenon takes off, the policy environment will need to adapt in order to accommodate the shift.

2.1. IoT policy challenges

A host of policy challenges exist by the rise of the Internet of Things. Luckily most of these fall into an existing law or policy process without the need for new, customized legislation. For example, the network standards developed for IoT, with attendant questions about interoperability and proprietariness, can have a home with standard-setting organizations like the Institute of Electrical and Electronics Engineers (IEEE), the Internet Engineering Task Force (IETF), and/or other nongovernmental organizations. Similarly, concerns about generating ubiquitous connectivity involves dealing with national and international bodies allocating radio spectrum.

We identify two new major policy challenges in the context of the emergent IoT: (1) the integration of "old" industries and the transformation of their governance into the multistakeholder model and (2) digital safety, which is not entirely new but becomes exponentially more widespread as connectivity continues to involve more physical spaces.

2.2. Integrating "old" industries into the multistakeholder governance mesh

On the one hand, traditional businesses are digitizing their means of production and they need to ensure safety and other good practices towards their employees (as well as towards the users of the products). On the other hand the products (or "things") produced are now Internet enabled and hence the companies become part of the existing Internet governance ecosystem. This second category is also known as the "industrial Internet" (or Machine to Machine networks or "industry 4.0" in Germany). The promise of the industrial Internet is in the potential to set up value networks in manufacturing, logistics, maintenance and business that have increasingly complete information capture, flow, sharing and productivity. In the industrial sphere of the IoT it is mainly the private sector that owns the facilities and is responsible for securing them. Hence manufacturers and businesses will be held responsible to ensure that their production, work and deployment environments are safe. In this context, the government's traditional role will be important for rule-setting. From there, interests of workers are represented and company practices are scrutinized by unions. Finally, the judicial system,

public scrutiny and peer pressure are responsible for enforcement.

Another challenge exists in the introduction and integration of traditional industries into the Internet Governance ecoystem. Tthis is significantly harder because of the sheer volume of new stakeholders that IoT brings with it, and the heterogeneity of policy challenges. Nevertheless we are faced with the task of introducing methods of and expertise from multistakeholderism gained in the Internet governance context to political and private sector spheres (manufacturing, city development, consumer electronics). These traditionally enjoyed more limited focus within the variety of local and global stakeholders into their decision making practices. Concepts like stratifying and delineating conflicts so that the "tussle" in this part of governance is productive and can be addressed and then solved by the right multistakeholder constellation (in the right layer) naturally become ever more important and more challenging.

2.3. Role of privacy in the discussion

From the user perspective, who controls what information about the user and her environment is a new question when applied to a world of connected things. And whether that information is adequately protected from misuse by government or private actors is a natural follow-on concern. As just one example, can one plausibly "opt out" of being involved with the devices and sensors that help comprise one's very environment? And is privacy only an issue where objects meet subjects?

Although the focus on our paper is on the distinction between safety and security, many of the consumer-protection agencies that handle consumer safety are also responsible for protecting privacy. We won't address this topic exhaustively here, although we believe that the IoT will likely see the entrance of many consumer-protection agencies into the field of governance on many topics, including privacy. In Europe, for example, there are several directives and rules that involve the digital market that are aimed at broadly protecting consumer rights, including privacy. Some of these include Directive 2011/83/EU on consumer rights; Directive 95/46/EC on the protection of individuals with regard to the processing of personal data; Directive 2000/31/EC on certain legal aspects of information society services; Directive 2002/58/EC concerning the processing of personal data and the protection of privacy in the electronic communications sector and many others. Although analysis of the relevance of these laws and their effect

on IoT is research for another day, we believe that it is relevant to identify some of the areas of exploration in the privacy space.

The move from a "small data" to a "big data" world necessitates several changes. For example, not all data will be personally identifiable, and instead, data growth will focus on the environment, infrastructure, and on the behavior of other devices and things on the Internet. Some data may fall into a gray area where predictions or revelations about individuals may be possible. The focus should be not on the collection of data, but whether it is harmful (e.g. use vs. collection). Because of this shift in use of the Internet, concepts like notice and choice will be less useful in the future, as will "data minimization" rules [2]. In this environment, distributed intelligence in the network can identify and protect against threats, even as it can also seed and spread threats throughout the user's environment.

2.4. Competing policy frameworks

Much like the Internet itself, the global cloud of things challenges national boundaries and barriers. The IoT also is not a single system, but comprises many overlapping networks of open, closed, and partially open systems. The multiplicity of different architectures and applications suggests no centralized governance structure; most is private company or user property, some is public on the Internet.

One way to look at the competing policy frameworks is through the lens of the Internet Governance Forum (IGF). The IGF has been in operation since 2006 and has met annually at the invitation of a host country. A multistakeholder advisory group (MAG), led by a chairman appointed by the UN Department for Economic and Social Affairs (UNDESA), organizes the annual meeting, deciding among proposed topics which will be on the agenda. A variety of formats are used to allow issues to be articulated and various perspectives to be shared.

The Internet Governance Forum (IGF) has looked at the IoT on various occasions. There is a Dynamic Coalition on the Internet of Things, but, to our knowledge, the experts so far concluded that no special IoT governance is warranted because it is "just" another Net application. If so, should responsibilities even be discussed in this early stage of IoT deployment? We think now is the right time exactly because it is still possible to adjust trajectories and inform choices. We argue that debate about how to share responsibility for IoT safety should happen at all levels and by all actors. That said we agree that because the benefits and harms of an IoT world are difficult to detect in the early stages, policymaking needs to be both flexible and future-proof.

3. Evolving Internet Governance

Because the Internet is a shared environment, its governance is a shared responsibility. Put more succinctly, the form of the Internet Governance Forum should follow its function. So what could a shared responsibility regime for IoT safety look like? We have previously argued that "forum follows function" in the context of the Internet Governance Forum [4]. This means that the IGF has the ability to perform the the functions to (1) identify emergent IoT safety challenges, (2) facilitate the creation of multistakeholder working groups that set out to develop solutions and to (3) monitor the effectiveness of these solutions allowing for open analysis and discussion, critiques and proposals for amendments or new recognition of problematic phenomena.

Traditional regulatory practices struggle to keep pace with the speed of innovation on the Internet. A more traditional "top-down" governance approach would stifle the very user benefits that stakeholders seek to enhance. This means that, far from any a single "one stop shop" for Internet governance, the informal practices that brought the Internet to the amazing state of constant growth and evolution today will, itself, continue to bring new and innovative governance systems along for the ride. Governments are important, but they do not play an exclusive, top-down, dominant role. Instead, governments participate on an equal footing as representatives of their respective constituents. These constituents — from the private sector to civil society to technical experts — are often in a tussle with governmental representatives and other relevant stakeholders on an issue by issue and sometimes case by case basis. Through an inclusive and transparent process, this helps create a truly global governance sphere.

What areas are each of these groups likely to advocate in the IoT context? A few of these roles include the following:

✦ **Government.** Set a high but implementable bar for protecting citizens, with two objectives: enforcement (e.g., consumer protection as well as health & safety concerns) and education (for example, phishing requires a sensitivity from the consumers that no system can absolutely protect).

✦ **Private sector.** Build the most secure systems possible, and share the education objective on how best to use tools available. Solicit ways to involve users as partners in providing better safety and security.

◆ **Civil society.** Represents the user and public-interest perspectives. Civil society plays a key role in verifying that checks and balances for governmental institutions are functioning and in keeping up with technological innovation. It also ensures that the line between government and the private sector is maintained. Additionally, there is also an educational objective to spread digital literacy, and this includes educating users about safety and security. Finally, civil society is on the front lines to protect human rights in dangerous parts of the world, and to educate citizens to be save and secure online.

◆ **Technical community.** The "technical community" makes sure that the standards are advancing properly (e.g., work of the IETF) and that protections and best practices (like DNSSEC) are implemented in the Internet's critical infrastructure. Although each group has its own educational mandate, the technical community also shares some similar educational objectives as civil society to educate all stakeholder groups (and in particular, users) on the proper design and functioning of various systems.

As we think further about the meaning of safety, it is useful to identify how these concepts can be found in different ways in the different layers of the Internet. In this vein, we believe that it is most sensible to look at the evolution of the Internet (and its governance) from the perspective of a "layered model" [4]. As we look at the different parts of each network layer, the resources (and hence the stakeholders) are a variable mix of private and public actors. Only in the logical layer is the resource largely a shared, public commons. As a result, in all the layers (except logical), the blend of private and public goods means that responsibility for safety issues remains with a mix of private and public actors. By contrast, the logical layer operates under a different governance regime than the others, one that accounts for this material difference. There, responsibility for safety matters lie with the pertinent multistakeholder governance community as a whole (rather than distributed among the stakeholders). We elaborate on the layered model as it applies to IoT below.

3.1. Placing IoT within the existing governance model

How do governance matters related to IoT relate to governance of the rest of the Internet? In Figure 1 we build on a model we previously proposed, but modified to include sample stakeholders who are specifically relevant to IoT governance. As we will see in our description below, the most important area of activity for governance of IoT will be in the content layer.

SOCIAL LAYER
- Identity theft and dangerous behavior
- Authentication problems (CCTV)

CONTENT LAYER

Software (Apps)   • Bad code (causing damage)
Cloud Platforms   • Ashley Madison (nafarious artifical intelligence
IoT Device OS   • Access to platforms & marketplaces (app store)
IoT Handware   • Traditional product safety (drones, 3D priting, issues with initial authorization and user modifications)

LOGICAL LAYER
- Naming and numbering
- Protocols and Standards

INFRASTRUCTURE LAYER
- streets, typically public goods
- or Internet, typically club goods

Fig. 1. Layered model of Internet Governance adapted to Internet of Things

The infrastructure layer at the base of Figure 1 is the physical foundation for the network. The rights-of-way, poles and areas shared with other providers are sometimes described as a "club good" provided by private ISPs, and it has converged with infrastructure supplied by public entities like power, water, streets and lighting. Governance practices of the physical layer are relatively well established, even if new innovations are challenging some of those traditions (e.g., regulation for things like High Altitude Platform Services). Above the infrastructure layer, the logical layer demonstrates a more robust history of multistakeholder governance models, with active groups like ICANN, IETF, IEEE and others. Adaptation of these groups to IoT-specific stakeholders should participate with the relevant institutions.

Most of the challenges related to IoT governance will occur in the content layer and above. Because of this tendency, the institutions that currently handle governance in this layer will need to be strengthened in order to address the arrival of IoT devices into the governance landscape. In the context of IoT, we prefer to refer to this layer as the "user and interaction layer" because the intent is to capture activities that include "traditional" web, virtual services, and services that have things attached to them or depend on connected sensors and devices. We expect that many governments will argue that the addition of sensing and actuating things to the network may require a heightened analysis when it comes to user safety.

With consumer safety issues becoming more important, we may see more active participation by a myriad of consumer-protection agencies in the heart of Internet governance. Here, no new special safety challenges arise with IoT. Obviously, consumers will continue to do silly things like using hair-dryers while sitting in a tub of water. On some level even this kind of problem will be less important with smarter devices (e.g., a hair dryer with sensors that turns off when there is too much moisture), but there will always be an opportunity to discuss where the lines of demarcation are in the "shared responsibility" of manufacturers to make devices that anticipate user errors, and the users themselves who must use devices responsibly.

3.2. Polycentric governance mechanisms

Recently, the logical layer has received a lot of attention because of the study of polycentric governance mechanisms as identified by Nobel Laureate Elinor Ostrom [14]. One of the many contributions that Ostrom brought with her work is to point out that complexity is not the same as chaos in polycentric governance systems. Polycentric governance is characterized by an organizational structure where multiple independent actors mutually order their relationships with one another under a general system of rules. There is no single stakeholder positioned to drive the Internet's future; instead, all stakeholders should act within their areas of expertise (in a "polycentric" way") and hold account for their area of responsibility. These stakeholders press forward in a spirit of interdependent cooperation and stewardship in the same way as the majority of the standards for the Internet, itself, have developed at the IETF.

3.2.1. The bazaar of Internet Governance

One of the best examples of how technology can be developed in a loosely coupled governance environment was penned by computer programmer Eric Raymond. Raymond's essay, "The cathedral and the bazaar" [16] outlines different approaches to software engineering, the "cathedral," which represents a top-down hierarchy, and the "bazaar" representing a flatter, more democratic (albeit chaotic) process for software development. This is a useful way to analyze the philosophy of the Internet's development as compared to older telecom industries (e.g., Bell Labs). In essence, the "bazaar" method for software writing is not unlike how Wikipedia works: the system is open, exposed, subject to comment by anyone

who has an opinion. Raymond's central claim is that "given enough eyeballs, all bugs are shallow," which, essentially, means that broad dissemination and discussion of coding provides better pThe equivalent of the bazaar in standard-setting organizations is the the Internet Engineering Task Force (IETF) — an open, volunteer-based standards-setting environment without any formal corporate "personality," where engineers have developed the core functionality that enables packets to transfer throughout the Internet. All IETF designs are freely accessible, and all IETF processes are published in their entirety on the Internet. If anything, reading the IETF website can be a bit onerous if only because it might feel like there's too much information available. Notably, the publications are all available and readable in any format, and it's expected that anyone, anywhere, can participate in the IETF process. As Harald Alvestrand describes, the IETF depends on an entirely open process, which means that "any interested person can participate in the work, know what is being decided, and make his or her voice heard on the issue. Part of this principle is our commitment to making our documents, our WG [working group] mailing lists, our attendance lists, and our meeting minutes publicly available on the Internet" [1].

Drawing from analogies throughout the open-standards space, the IETF is a true meritocracy: If members of the IETF community determine that an engineer's ideas have value, those ideas are adopted and incorporated into the Internet's suite of standards. Ideas that are dated or counterproductive, on the other hand, fester and fail. As famously stated by David Clark of the Massachusetts Institute of Technology: "We reject kings, presidents and voting. We believe in rough consensus and running code" [11]. By contrast, many top-down, government-centric systems rely on appointments by governments and formal committees, just as worldwide development of standards under the PTT model did.

3.2.2. Safety is already a shared responsibility

Importantly, it's not only the Internet where shared responsibility occurs. One example is in the way people organize themselves for home and neighborhood security. Here, the government plays a role via creating legal disincentives for criminals and enforcement through police, as well as the establishment of safety standards and codes. The private sector also participates via private security guards; alarm monitoring systems; theft-deterrent fencing, and the like. Civil society plays its own role through

mechanisms like neighborhood watch programs and collective actions by individual neighborhoods. Further, there is a question of education and user choice as well — e.g., some people choose to lock their door, some don't; some people choose to use deadbolts, others don't.

3.3. Mapping IoT
to the current governance system

In this last section we will take a look at the governance environment we want to work towards, especially as concerns IoT safety. Before we discuss the features of a good governance system, allow us to elaborate on what we see as ambition for future technologies and IoT in particular. We hold a user-centric conception of technology meaning that the Internet of Things should further human development and increase the well-being of users. All technology is a means to further scientific advancements and innovation not only for its own sake but to further a people-centred Information Society and richer, more just governance and economic conditions. Lawrence Lessig elaborated on this theme by distinguishing between technologies of access (to knowledge and services) and technologies of control (empowering institutions to better monitor and control users) [13].

If technology is user-driven, what kinds of experiences will users have when IoT goes mainstream? On an abstract level, most IoT benefits are generated through applications of systems. For example, networked sensors and traffic signals lead to more efficient traffic flow in a smart city, temperature sensors and smart distributed heating and cooling systems lead to energy savings in smart homes and similarly agricultural sensor networks allow for more targeted watering and fertilizing resulting in higher yields and higher sustainability.

We may find a significantly different and rapidly evolving constellation of responsibilities when analyzing the consumer oriented sector of the IoT, as this area will include notions of increasingly smart home products. As noted earlier, these roles may be increasingly filled by consumer-protection agencies and by consumer-focused civil society efforts. While responsibility for safety is not always easy to assign in bigger infrastructure, there is at least relatively clear ownership and hence control of the system in question (e.g, a road, factory, etc). In the consumer space, however, the multitude of connected devices and their interaction may create new safety threats. Additionally, the end uses are increasingly involved in the customization and in hacking the deployment of the product. Also, "prosumers" — consumers who produce and consume media, and who hack their devices — will continue to increase, and this trend complicates the task of communicating safety flaws of products.

In *Figure 2* we map the IoT governance world to the model that we have previously presented. Stakeholder groups such as automotive companies, consumer protection agencies and public safety will increasingly express themselves in the governance sphere. These and other groups will play an increasing role in (i) designing and defining safety standards, (ii) ensuring products and services are safe to use and (iii) dealing with safety hazards once they happen.
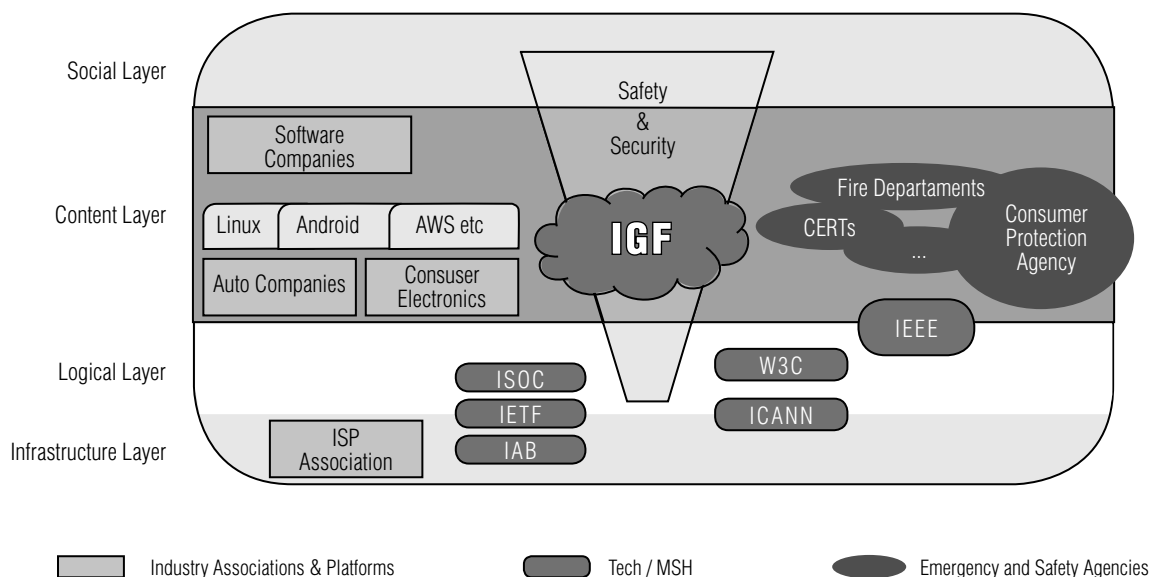
Fig. 2. IoT safety governance ecosystem

The approach described above provides additional groups within each of the layers to show on what layer agencies that deal with IoT issues would interface with colleagues. As in other aspects of Internet governance, this construct leads to a dynamic, interwoven and hyper-textual network of commitments between stakeholders that agree to develop and implement solutions together. This informal and decentralized method is, in our assessment, more effective and adequate than other (more traditional) approaches like trying to develop and ratify a hypothetical *Treaty on Cyber Safety or a hypothetical Treaty on Safety of the Internet of Things*. Among the many reasons, this approach allows for faster updates, more experimentation with new solutions and flexibility to address unique cultural, contextual and political contexts. These kinds of treaty proposals are not outlandish. We saw 193 nations struggle with how to implement (or not implement) Internet security measures by force of treaty at the World Conference on International Telecommunications (WCIT-12). The attempts seen at WCIT-12 and other measures are likely going to continue [4].

There are some stakeholder institutions that are worth singling out because of the role we expect them to play in IoT safety, and in particular, the role of Computer Emergency Response Teams (CERTs) and Computer Security Incident Response Teams (CSIRTs). The first CERTs were started almost 30 years ago and today there is a network of CERTs around the world, and one of their main activities is addressing internet security/safety incidents. CERTs provide technological activities in the cyberworld that are comparable to firefighters or emergency natural disaster response teams in physical world. The mandate of both CERTs, and other public-safety groups, may expand to include IoT emergencies, and hence their funding base should include all traditional private sector players that embrace connectivity in their products.

Conclusion

The advent of the Internet of Things gives us a tremendous opportunity to examine the very real issues at the heart of protecting the legitimate interests of users. An increased understanding of the distinction between the network, operating systems of things, application software and usages — especially when it comes to the security and safety of human users — will allow a crisper creation of solid multistakeholder practices in both virtual and physical worlds.

Many research questions of course remain open: What responsibilities does the general public inherit from the use of Internet enabled devices to protect others from dangerous behavior and against abuse (e.g. the fence around the swimming pool that is required by law in some jurisdictions)? Will the concept of negligence and causation change in the context of the Internet of Things? What are the consequences of a failure to update a safety critical software in a timely way? These and many other questions are certain to arise as the population of Internet-enabled devices increases and becomes a working part of our daily lives. ■

References

1. Alvestrand H. (2004) *A mission statement for the IETF. IETF*, Request for Comments 3935. Available at: https://www.ietf.org/rfc/rfc3935.txt (accessed 01 February 2016).
2. Andrade P.L., Hemerly J., Recalde G., Ryan P. (2014) From Big Data to big social and economic opportunities: Which policies will lead to leveraging data-driven innovation's potential? *The Global Information Technology Report 2014: Rewards and Risks of Big Data*. INSEAD, Cornell University, the World Economic Forum, pp. 81–86.
3. Cerf V.G., Ryan P.S., Senges M. (2014) Internet Governance is our shared responsibility. *I/S: A Journal of Law and Policy for the Information Society*, 10 ISJLP 1 (2014). Available at: http://ssrn.com/abstract=2309772 (accessed 01 February 2016).
4. Cerf V.G., Ryan P.S., Senges M., Whitt R.S. (2014) A perspective from the private sector: Ensuring that forum follows function. In: *Beyond Netmundial: The Roadmap for Institutional Improvements to the Global Internet Governance Ecosystem* (W.J.Drake, M.Price, eds). Center for Global Communication Studies, Annenberg School for Communication at the University of Pennsylvania. Available at: http://ssrn.com/abstract=2489348 (accessed 01 February 2016).
5. Chesbrough H., Van Alstyne M. (2015) Permissionless innovation. *Communications of the ACM*, vol. 58, no. 8, pp. 24–26.
6. Cisco (2012) *Tomorrow starts here*. Press release. 10 December 2012. Available at: http://goo.gl/vlKOfJ (accessed 01 February 2016).
7. Conficker Working Group (2010) *Conficker Working Group: Lessons Learned*. June 2010 (Published January 2011). Available at: http://www.confickerworkinggroup.org/wiki/uploads/Conficker_Working_Group_Lessons_Learned_17_June_2010_final.pdf (accessed 01 February 2016).
8. Faddell T. (2015) Nest CEO Tony Fadell on the future of the Internet. *Wall Street Journal*, 26 April 2015. Available at: http://www.wsj.com/articles/nest-ceo-tony-fadell-on-the-future-of-the-internet-1430104501 (accessed 01 February 2016).
9. Gayer O., Wilder O., Zeifman I. (2015) CCTV botnet in our own back yard. *The Incapsula Blog*, 21 October 2015. Available at: https://www.incapsula.com/blog/cctv-ddos-botnet-back-yard.html (accessed 01 February 2016).
10. Gross N. (1999) The earth will don an electronic skin. *Bloomberg Business Week*, 30 August 1999.
11. Hoffman P. (2012) The Tao of IETF: *A novice's guide to the Internet engineering task force*. Available at: http://www.ietf.org/tao (accessed 01 February 2016).
12. Kirkpatrick D. (2013) Why an Internet of everything event? 'It's the world waking up'. *Techonomy*, 3 May 2013.

13. Lessig L. (2002) The future of ideas: *The fate of the commons in a connected world.* Vintage Books.

14. Ostrom E. (2009) *Beyond markets and states: Polycentric governance of complex economic systems.* Nobel Prize Lecture, 8 December 2009. Available at: http://www.nobelprize.org/nobel_prizes/economic-sciences/laureates/2009/ostrom_lecture.pdf (accessed 01 February 2016).

15. Pettersson E. (2015) Sony to pay as much as $8 million to settle data-breach case. *Bloomberg Business*, 20 October 2015.

16. Raymond E.S. (2000) *The cathedral and the bazaar.* V. 3.0. Available at: http://www.catb.org/esr/writings/homesteading/cathedral-bazaar/ (accessed 01 February 2016).

17. Ryan P., Falvey S. (2012) Trust in the clouds. *Computer Law & Security Review*, vol. 28, no. 5, pp. 513−521.

18. Schindler H.R., Cave J. (2013) *Towards a dynamic and trustworthy Internet of things.* Rand Europe Research Brief, 88-9742-EC.

19. Westby J. (2014) *Instead of a real response, perennially hacked Sony is acting like a spoiled teenager.* Forbes, 17 December 2014. Available at: http://www.forbes.com/sites/jodywestby/2014/12/17/sony-earns-cyber-troglodyte-award/#2fc4f7f96942 (accessed 01 February 2016).

20. Whitt R.S. (2013) A deference to protocol: Fashioning a three-dimensional public policy framework for the Internet age. *Cardozo Arts & Entertainment Law Journal*, no. 689, pp. 754−756.

Интернет вещей:
безопасность и защита как коллективная ответственность

В.Г. Серф
вице-президент, Alphabet Inc.
Адрес: 1600 Amphitheatre Parkway, Mountain View, CA 94043, USA
E-mail: vint@google.com

П.С. Райан
директор по стратегии и операциям, Alphabet Inc.
Адрес: 1600 Amphitheatre Parkway, Mountain View, CA 94043, USA
E-mail: patrickryan@google.com

М. Сенгес
менеджер исследовательских программ, Alphabet Inc.
Адрес: 1600 Amphitheatre Parkway, Mountain View, CA 94043, USA
E-mail: maxsenges@google.com

Р.С. Витт
корпоративный директор по стратегическим инициативам, Alphabet Inc.
Адрес: 1600 Amphitheatre Parkway, Mountain View, CA 94043, USA
E-mail: whitt@google.com

Аннотация

Что произойдет, если отдельные технические устройства приобретут сетевой интерфейс? Очевидно, что результатом будет беспрецедентное количество «вещей», подключенных к сети Интернет. Менее очевиден ответ на вопрос, что это будет означать для управления Интернетом. В свете концепции Интернета вещей (Internet of Things, IoT) взаимодействующие управленческие структуры уже приспосабливаются к тому, чтобы соответствовать эволюции применений Интернета. Однако, поскольку структура управления продолжает развиваться, безопасность пользователей становится приоритетом для всех поставщиков технических и программных решений. В контексте Интернета вещей в данной статье предлагается определение цифровой безопасности, отличающееся от защиты данных, а также обсуждается то, как управление, затрагивающее различных заинтересованных лиц, может применяться для обеспечения такой безопасности. В работе также рассматриваются вопросы интеграции «старых» отраслей и трансформации управления ими в «мультистейкхолдерную» модель по мере того, как их продукты и услуги становятся онлайновыми. Обсуждается, как тысячи производителей, традиционно производящие не связанные между собой «вещи», адаптируются к роли стейкхолдеров Интернета и как это изменяет наше представление об управлении Интернетом. Особое внимание в статье уделено тому, как это связано с вопросами безопасности, которые становятся все более актуальными в связи с широким распространением Интернет-ориентированных физических устройств.

Авторы выполнили данную работу в академических целях, на основе своего видения рассматриваемых вопросов. Мнение авторов может не совпадать с официальной позицией их работодателя.

Литература

1. Alvestrand H. A mission statement for the IETF // IETF, Request for Comments 3935. 2004. [Электронный ресурс]: https://www.ietf.org/rfc/rfc3935.txt (дата обращения 01.02.2016).

2. Andrade P.L., Hemerly J., Recalde G., Ryan P. From Big Data to big social and economic opportunities: Which policies will lead to leveraging data-driven innovation's potential? The Global Information Technology Report 2014: Rewards and Risks of Big Data // INSEAD, Cornell University, the World Economic Forum. 2014. P. 81−86.

3. Cerf V.G., Ryan P.S., Senges M. Internet Governance is our shared responsibility // I/S: A Journal of Law and Policy for the Information Society, 10 ISJLP 1 (2014). 2014. [Электронный ресурс]: http://ssrn.com/abstract=2309772 (дата обращения 01.02.2016).

4. Cerf V.G., Ryan P.S., Senges M., Whitt R.S. A perspective from the private sector: Ensuring that forum follows function // Beyond Netmundial: The Roadmap for Institutional Improvements to the Global Internet Governance Ecosystem / W.J. Drake, M. Price, eds. Center for Global Communication Studies, Annenberg School for Communication at the University of Pennsylvania. 2014. [Электронный ресурс]: http://ssrn.com/abstract=2489348 (дата обращения 01.02.2016).

5. Chesbrough H., Van Alstyne M. Permissionless innovation // Communications of the ACM. 2015. Vol. 58. No. 8. P. 24−26.

6. Tomorrow starts here // Cisco press release. 10 December 2012. [Электронный ресурс]: http://goo.gl/vlKOfJ (дата обращения 01.02.2016).

7. Conficker Working Group: Lessons Learned. June 2010 (Published January 2011). [Электронный ресурс]: http://www.confickerworkinggroup.org/wiki/uploads/Conficker_Working_Group_Lessons_Learned_17_June_2010_final.pdf (дата обращения 01.02.2016).

8. Faddell T. Nest CEO Tony Fadell on the future of the Internet // Wall Street Journal, 26 April 2015. [Электронный ресурс]: http://www.wsj.com/articles/nest-ceo-tony-fadell-on-the-future-of-the-internet-1430104501 (дата обращения 01.02.2016).

9. Gayer O., Wilder O., Zeifman I. CCTV botnet in our own back yard // The Incapsula Blog, 21 October 2015. [Электронный ресурс]: https://www.incapsula.com/blog/cctv-ddos-botnet-back-yard.html (дата обращения 01.02.2016).

10. Gross N. The earth will don an electronic skin // Bloomberg Business Week, 30 August 1999.

11. Hoffman P. The Tao of IETF: A novice's guide to the Internet engineering task force. 2012. [Электронный ресурс]: http://www.ietf.org/tao (дата обращения 01.02.2016).

12. Kirkpatrick D. Why an Internet of everything event? 'It's the world waking up' // Techonomy, 3 May 2013.

13. Lessig L. The future of ideas: The fate of the commons in a connected world. Vintage Books, 2002. 384 p.

14. Ostrom E. Beyond markets and states: Polycentric governance of complex economic systems. Nobel Prize Lecture, 8 December 2009. [Электронный ресурс]: http://www.nobelprize.org/nobel_prizes/economic-sciences/laureates/2009/ostrom_lecture.pdf (дата обращения 01.02.2016).

15. Pettersson E. Sony to pay as much as $8 million to settle data-breach case // Bloomberg Business, 20 October 2015.

16. Raymond E.S. The cathedral and the bazaar. V. 3.0. 2000. [Электронный ресурс]: http://www.catb.org/esr/writings/homesteading/cathedral-bazaar/ (дата обращения 01.02.2016).

17. Ryan P., Falvey S. Trust in the clouds // Computer Law & Security Review. 2012. Vol. 28. No. 5. P. 513−521.

18. Schindler H.R., Cave J. Towards a dynamic and trustworthy Internet of things // Rand Europe Research Brief, 88-9742-EC. 2013.

19. Westby J. Instead of a real response, perennially hacked Sony is acting like a spoiled teenager // Forbes, 17 December 2014. [Электронный ресурс]: http://www.forbes.com/sites/jodywestby/2014/12/17/sony-earns-cyber-troglodyte-award/#2fc4f7f96942 (дата обращения 01.02.2016).

20. Whitt R.S. A deference to protocol: Fashioning a three-dimensional public policy framework for the Internet age // Cardozo Arts & Entertainment Law Journal. 2013. No. 689. P. 754−756.

Internet Governance: Trends and realities. Part 2

Peter Major

Vice-Chairman
United Nations Commission on Science and Technology for Development (CSTD)
Address: Palais des Nations, 8-14, Av. de la Paix, 1211 Geneva 10, Switzerland
E-mail: pmajor@bluewin.ch

Abstract

Internet Governance is one of the global governance issues that emerged at the end of the 1990s. As the Internet is taking on an important role in every aspect of our lives, it has been recognized that its governance needs to involve all stakeholders and institutions to allow global access, foster development and contribute to the global economy, education, information. Security and privacy concerns should also be addressed in the policy dialogues. This paper outlines the global nature of the Internet Governance. It argues that in addition to dialogues and negotiations formal approaches should be explored to handle global issues in a global way. In its conclusion a mechanism is proposed to address Internet related global public policy issues.

The views expressed in this paper are solely those of the author in his private capacity and do not in any way represent the views of the UN Commission on Science and Technology for Development.

5. Regional and national Internet Governance approaches

The Internet evolved as a project with the participation of the US government and the technical community/academia. Subsequently business took a great interest in the Internet and contributed to its success through massive investments to expand the infrastructure and to create an environment for applications. With the rapid evolution of the Internet, civil society and international organizations got involved on issues as human rights, consumer protection, intellectual property rights, as well as privacy and security issues. Involvement of governments has been quite visible since the second phase of WSIS.

The Tunis Agenda of WSIS left an ambiguity in the Internet Governance approaches: on one hand, it evoked the multi-stakeholder governance model, but the notion of enhanced cooperation implicitly encouraged the government lead approach. Instances of the first model can be found in many organizations.

Internet Governance on the international level has evolved as a bottom-up, business-led multi-stakeholder process with the involvement of governments, the technical community, academia, civil society and international organizations. On the national level, most countries advocate the multi-stakeholder model. In practice, Internet governance issues frequently reflect political system-specific characteristics. In the shaping of national IG policies, most governments try to have a coordinated strategy, usually elaborated by the ministry of economic development, the ministry of information and communication technologies and the ministry of foreign affairs. However, in the absence of clear leadership, there may be contradictory statements and initiatives in some countries. The main concerns for internet- related national policy: national sovereignty, security, human rights, surveillance, privacy, cyber espionage, censorship, etc.

Many countries consider the Internet to be borderless space where everyone should be able to exercise rights online that are granted by national legislation. It is thought

that this approach has fostered the evolution of the Internet and it should be adopted by all. Other countries view the Internet as an extension of the respective national physical space and consider that exercising national sovereignty over it is of primary importance. Yet another group of countries try to follow a policy in between the two approaches. Some issues such as managing infrastructure and critical resources have always been among the main concerns of countries not involved directly in the global management of these resources. Access, content, surveillance and combating cyber-crime are closely related to human rights and privacy, as well as to security. In this context, it is important to note that in July 2014 the UN Human Rights Council in its resolution cosponsored by Brazil, Nigeria, Sweden, Turkey, Tunisia and the United States reaffirmed its earlier resolution that 'the same rights that people have offline must also be protected online, in particular the freedom of expression" [17]. The resolution recognizes that the global and open nature of the Internet is a driving force in accelerating progress towards development, including the implementation of the right to education. It also calls upon States to address the digital divide and to promote digital literacy and access to information on the Internet. The resolution takes note of the need for human rights to underpin Internet governance, and it affirms the importance of the global, open and interoperable nature of the Internet. It calls on States to formulate, through transparent and inclusive processes, national Internet-related public policies. Universal access and enjoyment of human rights must be central to those policies. The resolution recognizes that respect for the rights to freedom of expression and privacy is key to building confidence and trust in the Internet, and that any attempt by States to address security concerns on the Internet must be in accordance with international human rights obligations. Critically, the resolution states this must be done through democratic, transparent institutions, based on the rule of law.

Internet Governance is deeply embedded in national politics. The multi-stakeholder approach to Internet Governance is the generally accepted principle. Surveillance, censorship, limiting access and curtailing human rights are justified by governments with the emerging security concerns, cyber threats and cyber crime. Significant differences in national approaches, also articulated in international discussions, may prevent consensus in open negotiations.

In international discussions, like-minded countries emerge as a group emphasizing particular points from the Tunis Agenda. The US, UK, France other members of the EU, Australia, Brazil, Japan, South Korea, etc. are unconditional advocates of the multi-stakeholder Internet governance and consider it to be the prerequisite of in-novation, competition and the freedoms on the Internet. Russia, China, Iran, Saudi Arabia and others recognize the importance of the multi-stakeholder approach on the international level, but their priorities are national sovereignty, security and social order.

Table 2 summarizes some initiatives for cooperation on cyber security issues on the international and regional levels.

Table 2.

**Instruments and scopes
of international cooperation on cyber security**

Instrument	Scope of international cooperation provisions
Draft African Union Convention [18]	Cybercrime
Commonwealth of Independent States (CIS) Agreement [19]	Offenses relating to computer information
Council of Europe (CoE) Cybercrime Convention [20]	Criminal offenses related to computer systems and data collection of evidence in electronic form of a criminal Offense
League of Arab States Convention [21]	Information and information technology offences Gathering of electronic evidence in offenses
Shanghai Cooperation Organization Agreement [22]	International information security

There are diverging views expressed in these documents, but commonalities found in the texts may offer a way forward.

In February 2015, the US Federal Communication Commission (FCC) accepted so-called net neutrality regulations to reclassify Internet Service Providers (ISPs) as common carriers, regulating them under Title II of the Communications Act, the same statute that governs telephone companies. Internet providers will be common carriers in their relationships with home Internet and mobile broadband customers; they will also be common carriers in their relationships with companies that deliver content to subscribers over the networks operated by ISPs. That includes online content providers. Broadband providers may not block access to legal content, applications, services, or non-harmful devices; they may not impair or degrade lawful Internet traffic on the basis of content, applications, services, or non-harmful devices and may not favor some lawful Internet traffic over other lawful traffic in exchange for a consideration. This rule also bans ISPs from prioritizing content and services of their affiliates. There is no ban on data caps (limiting the transfer of a specified amount of data over a period of time), but the proposal would let the FCC intervene when caps are used to harm consumers or competitors [23]. It is anticipated that many regulators will incorporate this approach in their national policies.

Several national and regional initiatives and dialogue forums help to shape the Internet Governance policy decisions. These initiatives in most cases follow the pattern of the IGF but concentrate on national and regional issues (*http://www.intgovforum.org/cms/meetings*).

6. Towards a global framework for global Internet Governance

The evolution of the Internet is very fast, but the evolution of the legal framework is much slower. In the discussions related to legal frameworks, it is assumed that the Internet is not fragmented. The Internet Governance is embedded in a multi-stakeholder approach and the legal framework needs to be harmonized. Many countries focus on human rights, data protection, privacy rights, consumer rights and intellectual property rights.

Most of the existing international treaties, conventions and agreements covering human rights were negotiated before the Internet (The Universal Declaration of Human Rights [24], UN International Covenant on Civil and Political Rights [25], UN International Covenant on Economic, Social and Cultural Rights [26], UN Convention on the Elimination of All Forms of Discrimination Against Women [27], UN Convention of the Rights of the Child [28]). They have to be analyzed with a view to extending their applicability to cyberspace.

More recent treaties and resolutions on human rights have already considered or made reference to Interne and ICTs (UN Convention on the Rights of Persons with Disabilities [29], UN Human Rights Council Resolution on the promotion, protection and enjoyment of human rights on the Internet).

Data protection and privacy rights, similarly to human rights, are considered to be very important. The primary purpose of data protection legislation is to protect individuals against possible misuse of personal data information about them held by others, to give people the right to know what information organizations hold about them, and to provide a framework for organizations handling personal data. There are international [30] and regional initiatives [31] about collecting, organizing or altering personal data; retrieving, consulting, using, storing or adapting the data; disclosing the data by transmitting, disseminating or otherwise making it available; or aligning, combining, blocking, erasing or destroying the data. During these activities, information must be fairly and lawfully processed, processed for specified purposes, appropriate, relevant and not excessive, accurate and up-to-date, not kept for longer than is necessary, processed in line with individuals' rights, secure and not transferred outside the country without adequate protection.

The UN Guidelines on Consumer Protection (UN-GCP) [32] provide a framework for governments to develop and strengthen consumer protection policies and legislation. The document adopted by the UN General Assembly in 1985 was amended in 1999. It contains the essential elements of comprehensive consumer protection, including the right to safety, the right to basic needs, the right to information, the right to consumer education, the right to redress, the right to be heard and the right to a healthy environment. The guidelines are being revised by UNCTAD.

Important, Internet related treaties: the WIPO Copyright Treaty [33] and the WIPO Performances and Phonograms Treaty [34].

The ITU International Telecommunication Regulations [35] establish general principles relating to the provision and operation of international telecoms; facilitate global interconnection and interoperability; underpin harmonious development and efficient operation of technical facilities; promote efficiency, usefulness, and availability of international telecommunication services, and treaty-level provisions are required with respect to international telecommunication networks and services. Net-neutrality regulations may have an impact on the interpretations of the International Telecommunication Regulations (ITRs) of the ITU modified in the controversial World Conference on International Telecommunications (WCIT) in Dubai, in 2012.

The ACTA − Anti-Counterfeiting Trade Agreement is an international treaty for the purpose of establishing international standards for intellectual property rights enforcement. The agreement aims to establish an international legal framework for targeting counterfeit goods, generic medicines and copyright infringement on the Internet, and would create a new governing body outside existing forums, such as the World Trade Organization, the World Intellectual Property Organization, and the United Nations. The treaty has not entered into force after worldwide protests arguing that that the 'intended benefits of this international agreement are far outweighed by the potential threats to civil liberties".

The General Agreement on Trade and Services (GATS), a treaty of the World Trade Organization (WTO), is one of the results of the 8th round of the multilateral trade negotiations (Uruguay Round), conducted from 1986 to 1994 and signed by all member states of the WTO [36]. The GATS provides an international framework for trade and services including e-commerce, e-banking and other e-services (except government services) using the Internet. The agreement came into force in 1995 after difficult nego-

tiations. Another result of the Uruguay Round is the Agreement of the Trade Related Aspect of Intellectual Property Rights (TRIPS) [37]. This agreement sets down minimum standards of intellectual property regulations as applied to nationals of other WTO members.

Ten years after the WSIS, stakeholders are assessing the achievements and considering the possible ways forward. There are significant results in bridging the digital divide. We have to recall that almost 60% of the world's population still does not have access to the Internet. We have to face a new challenge, the so-called broadband divide between developed and developing countries. There are significant results in Internet Governance: many of the proposals of the WGIG included in the Tunis Agenda have been or are being implemented. IGFs provide platforms for open discussions and contribute to better governance. It is clear, however, that global public policy issues have to be addressed using appropriate mechanisms in the UN system that are acceptable to all stakeholders. The approach may require reassessing, reviewing and if needed modifying available instruments with a view to taking into account the specific nature of the Internet. At the same time, the feasibility of using the multi-stakeholder approach should be explored and applied in Internet-related public policy issues dealt with within the UN system. In his report, the chair of the CSTD WGEC states that during the discussion on enhanced cooperation some participants proposed to create a new mechanism within the UN system mandated to deal with Internet related global public policy issues, while others argued that mechanisms already exist in the UN and in different agencies of the UN system.

To overcome the differences, the Secretariat of the UN CSTD may be tasked to help to explore the best way to deal with Internet related public policy issues within the UN system through performing clearing-house functions with the close cooperation of other UN agencies and other stakeholders. The Secretariat shall identify problems and issues, identify existing mechanism within the UN system where these problems are addressed, indicate where modification of the existing mechanism/mandate may be required and keep track of the outcome of the discussions/negotiations of the problems. In performing these functions, the Secretariat in its assessments will be helped by a voluntary, multi-stakeholder advisory group selected by each stakeholder group and nominated by the Chair of the CSTD. Global issues in general are complex and require multidisciplinary approaches.

Many methods and techniques have been developed to deal with the complexity of systems, including systems dynamics, fractals, chaos theory, the science of networks,

and complexity theory (*http://www.ipsonet.org/publications/open-access/policy-and-complex-systems*). They provide a powerful set of tools to model and/or simulate phenomena that are characterized by their scale-free and/or small-world network structure, sensitivity to initial conditions, power-law distributions, adaptability, self-organization, feedback loops, and emergent properties. However, applying such tools on any real-world problem will require the mastery of intricacies of both public policy and a wide variety of discipline-specific expertise, working together to uncover principles that both transcend and complement disciplinary contributions.

There are views that application of scale-free network theory to the study of international relations, with special attention to cyber venues, may contribute to better understanding of cyberpolitics [38]. Recently attempts have been made to analyze complex mechanisms using mathematical models such as complex systems or network theory [39]. Further research is needed to explore how formal methods can be applied to global governance issues in general and to Internet Governance in particular.

Conclusion

In this paper I tried to show the complexity of global governance in general and of Internet Governance in particular. The Internet is operating based on standards elaborated through rough consensus, the first of which was published in 1969. The governance of the Internet, infrastructure, names and numbers and related issues, is handled in a multi-stakeholder approach. Significant efforts are being made to find solution to the controversies related to the management of the root zone. Discussions are being conducted about the transition of the US Government's stewardship to a multi-stakeholder arrangement with a view to meeting the target date of 30 September 2015. Governments are getting more and more involved in Internet Governance: more than 140 governments are participating in the work of the Government Advisory Committee of ICANN. They are making significant contributions to public policy discussions and give advice to the ICANN Board. GAC delegates take an important part in the discussions on the transition and the accountability of ICANN. The governance on the Internet is embedded in the national interests, policy priorities and traditions. Application of existing multilateral arrangements on issues of human rights, security, censorship, etc. need to be analyzed and reconsidered with a view to adapting them for governance on the Internet. Considerable efforts have to be made to reconcile concerns for free access, human rights, development as well as national security and sovereignty. ■

References

17. United Nations Human Rights Council (2014) *Promotion and protection of all human rights, civil, political, economic, social and cultural rights, including the right to development.* Resolution A/HRC/RES/26/13. Geneva. Available at: http://daccess-dds-ny.un.org/doc/UNDOC/LTD/G14/059/67/PDF/G1405967.pdf?OpenElement (accessed 30 September 2015).

18. Economic Commission for Africa, African Union Convention (2012) *Draft African Union convention on the confidence and security in cyberspace. Version 01/09/2012.* Available at: http://au.int/en/sites/default/files/AU%20Convention%20EN.%20%283-9-2012%29%20clean_0.pdf (accessed 30 September 2015).

19. Commonwealth of Independent States (2001) *Agreement on cooperation in combating offences related to computer information* (Commonwealth of Independent States Agreement).

20. Council of Europe (2001) *Convention on cybercrime*, 23 November 2001, Budapest. Available at: http://conventions.coe.int/Treaty/Commun/QueVoulezVous.asp?CL=ENG&NT=185 (accessed 30 September 2015).

21. Arab League (2010) *Arab convention on combating information technology offences.* Arab League, 21 December 2010, Cairo.

22. Shanghai Cooperation Organization (2009) *The agreement between the governments of state members of the Shanghai Organization of Cooperation about cooperation in the field of ensuring the international information security.* Shanghai Cooperation Organization, 16 July 2009, Yekaterinburg. Available at: https://ccdcoe.org/sites/default/files/documents/SCO-090616-IISAgreementRussian.pdf (accessed 30 September 2015).

23. Brodkin J. (2015) Don't call them "utility" rules: The FCC's net neutrality regime, explained // *Arstechnica.com*, 05 February 2015. Available at: http://arstechnica.com/business/2015/02/dont-call-them-utility-rules-the-fccs-net-neutrality-regime-explained/ (accessed 30 September 2015).

24. United Nations (1948) *The universal declaration of human rights.* New York. Available at: http://www.un.org/en/documents/udhr/ (accessed 30 September 2015).

25. United Nations (1966) *International covenant on civil and political rights (1st generation rights).* New York. Available at: http://www.ohchr.org/en/professionalinterest/pages/ccpr.aspx (accessed 30 September 2015).

26. United Nations (1966) *International covenant on economic, social and cultural rights (2nd generation rights).* New York. Available at: http://www.ohchr.org/EN/ProfessionalInterest/Pages/cescr.aspx (accessed 30 September 2015).

27. United Nations (1981) *Convention on the elimination of all forms of discrimination against women.* New York. Available at: http://www1.umn.edu/humanrts/instree/women/engl-wmn.html (accessed 30 September 2015).

28. United Nations (1989) *Convention of the rights of the child.* New York. Available at: http://www.un.org/cyberschoolbus/humanrights/resources/plainchild.asp (accessed 30 September 2015).

29. United Nations (2006) *Convention on the rights of persons with disabilities.* New York. Available at: http://www.un.org/disabilities/convention/conventionfull.shtml (accessed 30 September 2015).

30. United Nations (2013) *The right to privacy in the digital age.* New York. Available at: http://www.un.org/en/ga/search/view_doc.asp?symbol=A/RES/68/167 (accessed 30 September 2015).

31. European Commission (2015) *Protection of personal data.* Available at: http://ec.europa.eu/justice/data-protection/ (accessed 30 September 2015).

32. United Nations (1999) *United Nations guidelines on consumer protection.* New York. Available at: http://www.un.org/esa/sustdev/publications/consumption_en.pdf (accessed 30 September 2015).

33. World Intellectual Property Organization (1996) *WIPO copyright treaty.* Geneva. Available at: http://www.wipo.int/treaties/en/text.jsp?file_id=295166 (accessed 30 September 2015).

34. World Intellectual Property Organization (1996) *WIPO performances and phonograms treaty.* Geneva. Available at: http://www.wipo.int/treaties/en/text.jsp?file_id=295578 (accessed 30 September 2015).

35. ITU (2012) *ITU international telecommunication regulations.* Dubai. Available at: http://www.itu.int/pub/S-CONF-WCIT-2012/en (accessed 30 September 2015).

36. World Trade Organization (1995) *General agreement on trade and services.* Available at: https://www.wto.org/English/docs_e/legal_e/26-gats.pdf (accessed 30 September 2015).

37. World Trade Organization (1994) *Trade-related aspects of intellectual property Rights. Annex 1C of the Marakesh agreement astablishing the World Trade Organization.* Marakesh, Morocco. Available at: https://www.wto.org/English/docs_e/legal_e/27-trips_01_e.htm (accessed 30 September 2015).

38. Choucri N. (2012) *Cyberpolitics in international relations.* Cambridge, MA, MIT Press.

39. Barabasi A.-L. (2015) *Network science.* Cambridge, UK, Cambridge University Press.

Управление Интернетом: Тенденции и реальность. Часть 2

П. Мейор

заместитель руководителя Комиссии ООН по науке и технологиям
Адрес: Palais des Nations, 8-14, Av. de la Paix, 1211 Geneva 10, Switzerland
E-mail: pmajor@bluewin.ch

Аннотация

Управление Интернетом представляет собой одно из направлений глобального управления, активно развивающееся с конца 1990-х годов. Поскольку Интернет играет важную роль в нашей жизни, представляется необходимым вовлечение в процессы управления всех заинтересованных лиц и институтов, а также

способствование развитию и вкладу в мировую экономику, образование и информационное обеспечение. Вопросы информационной безопасности и конфиденциальности также должны рассматриваться в рамках обсуждения политики в области управления Интернетом. В данной статье рассматривается глобальный аспект управления Интернетом. Показано, что в дополнение к обсуждениям и переговорам должны применяться формальные подходы, позволяющие решать глобальные вопросы. Предлагается подход к решению вопросов, связанных с глобальной политикой в области управления Интернетом.

Выводы и рекомендации, приведенные в статье, отражают исключительно личное мнение автора и не должны трактоваться в качестве официальной позиции Комиссии ООН по науке и технологиям.

Литература

17. Promotion and protection of all human rights, civil, political, economic, social and cultural rights, including the right to development. Resolution A/HRC/RES/26/13. Geneva: United Nations Human Rights Council, 2014. [Электронный ресурс]: http://daccess-dds-ny.un.org/doc/UNDOC/LTD/G14/059/67/PDF/G1405967.pdf?OpenElement (дата обращения 30.09.2015).

18. Draft African Union convention on the confidence and security in cyberspace. Version 01/09/2012. Economic Commission for Africa, African Union Convention, 2012. [Электронный ресурс]: http://au.int/en/sites/default/files/AU%20Convention%20EN.%20%283-9-2012%29%20clean_0.pdf (дата обращения 30.09.2015).

19. Agreement on cooperation in combating offences related to computer information (Commonwealth of Independent States Agreement). Commonwealth of Independent States, 2001.

20. Convention on cybercrime, 23 November 2001, Budapest. Council of Europe, 2001. [Электронный ресурс]: http://conventions.coe.int/Treaty/Commun/QueVoulezVous.asp?CL=ENG&NT=185 (дата обращения 30.09.2015).

21. Arab convention on combating information technology offences. Arab League, 21 December 2010, Cairo. Arab League, 2010.

22. The agreement between the governments of state members of the Shanghai Organization of Cooperation about cooperation in the field of ensuring the international information security. Shanghai Cooperation Organization, 16 July 2009, Yekaterinburg. Shanghai Cooperation Organization, 2009. [Электронный ресурс]: https://ccdcoe.org/sites/default/files/documents/SCO-090616-IISAgreementRussian.pdf (дата обращения 30.09.2015).

23. Brodkin J. Don't call them «utility» rules: The FCC's net neutrality regime, explained // Arstechnica.com, 05 February 2015. [Электронный ресурс]: http://arstechnica.com/business/2015/02/dont-call-them-utility-rules-the-fccs-net-neutrality-regime-explained/ (дата обращения 30.09.2015).

24. The universal declaration of human rights. New York: United Nations, 1948. [Электронный ресурс]: http://www.un.org/en/documents/udhr/ (дата обращения 30.09.2015).

25. International covenant on civil and political rights (1st generation rights). New York: United Nations, 1966. [Электронный ресурс]: http://www.ohchr.org/en/professionalinterest/pages/ccpr.aspx (дата обращения 30.09.2015).

26. International covenant on economic, social and cultural rights (2nd generation rights). New York: United Nations, 1966. [Электронный ресурс]: http://www.ohchr.org/EN/ProfessionalInterest/Pages/cescr.aspx (дата обращения 30.09.2015).

27. Convention on the elimination of all forms of discrimination against women. New York: United Nations, 1981. [Электронный ресурс]: http://www1.umn.edu/humanrts/instree/women/engl-wmn.html (дата обращения 30.09.2015).

28. Convention of the rights of the child. New York: United Nations, 1989. [Электронный ресурс]: http://www.un.org/cyberschoolbus/humanrights/resources/plainchild.asp (дата обращения 30.09.2015).

29. Convention on the rights of persons with disabilities. New York: United Nations, 2006. [Электронный ресурс]: http://www.un.org/disabilities/convention/conventionfull.shtml (дата обращения 30.09.2015).

30. The right to privacy in the digital age. New York: United Nations, 2013. [Электронный ресурс]: http://www.un.org/en/ga/search/view_doc.asp?symbol=A/RES/68/167 (дата обращения 30.09.2015).

31. Protection of personal data. European Commission, 2015. [Электронный ресурс]: http://ec.europa.eu/justice/data-protection/ (дата обращения 30.09.2015).

32. United Nations guidelines on consumer protection. New York: United Nations, 1999. [Электронный ресурс]: http://www.un.org/esa/sustdev/publications/consumption_en.pdf (дата обращения 30.09.2015).

33. WIPO copyright treaty. Geneva: World Intellectual Property Organization, 1996. [Электронный ресурс]: http://www.wipo.int/treaties/en/text.jsp?file_id=295166 (дата обращения 30.09.2015).

34. WIPO performances and phonograms treaty. Geneva: World Intellectual Property Organization, 1996. [Электронный ресурс]: http://ITU international telecommunication regulations. Dubai: ITU, 2012. [Электронный ресурс]: http://www.itu.int/pub/S-CONF-WCIT-2012/en (дата обращения 30.09.2015).

35. ITU international telecommunication regulations. Dubai: ITU, 2012. [Электронный ресурс]: http://www.itu.int/pub/S-CONF-WCIT-2012/en (дата обращения 30.09.2015).

36. General agreement on trade and services. World Trade Organization, 1995. [Электронный ресурс]: https://www.wto.org/English/docs_e/legal_e/26-gats.pdf (дата обращения 30.09.2015).

37. Trade-related aspects of intellectual property Rights. Annex 1C of the Marakesh agreement astablishing the World Trade Organization. Marakesh, Morocco: World Trade Organization, 1994. [Электронный ресурс]: https://www.wto.org/English/docs_e/legal_e/27-trips_01_e.htm (дата обращения 30.09.2015).

38. Choucri N. Cyberpolitics in international relations. Cambridge, MA, MIT Press, 2012.

39. Barabasi A.-L. Network science. Cambridge, UK, Cambridge University Press, 2015.

An overlay network for distributed exact and range search in one-dimensional space

Alexander A. Ponomarenko
Researcher, Laboratory of Algorithms and Technologies for Network Analysis
National Research University Higher School of Economics
Address: 136, Rodionova Street, Nizhny Novgorod, 603093, Russian Federation
E-mail: aponomarenko@hse.ru

Yury A. Malkov
Junior Research Fellow
The Institute of Applied Physics of the Russian Academy of Sciences
Address: 46, Ulyanova Street, Nizhny Novgorod, 603950, Russian Federation
E-mail: yurymalkov@mail.ru

Andrey A. Logvinov
Project Leader, MERA Labs LLC
Address: 192, Rodionova Street, Nizhny Novgorod, 603093, Russian Federation
E-mail: alogvinov@meralabs.com

Vladimir V. Krylov
Professor, Head of Big Data Laboratory
Nizhny Novgorod State Technical University
Address: 24 Minina Street, Nizhny Novgorod, 603950, Russian Federation
E-mail: vkrylov@heterarchica.com

Abstract

The ability to scale is a desirable business requirement for computer systems. Distributed systems clearly demonstrate this ability and might to process very large volumes of data. Many systems with distributed architecture are based on the distributed hash table (DHT), which manages a set of distributed network nodes connected not only by a physical channel, but also by an additional overlay network. This overlay network is used for searching nodes and for distributing tasks among them. The main feature of this approach is that there is no central element or node which knows the global topology of the network. Nodes in the network are searched by passing a query message from one node to another. Despite that, every node has knowledge only about a small number of other nodes, and the network is organized in such a way that search involves a logarithmical number of nodes.

There are several DHT implementations which specify how to construct and how to support the structure of the network. In this paper, we demonstrate the way in which such a network can be constructed much simpler by applying the sight modification of the recently published Metrized Small World algorithm to the case of one dimension. We provide a theoretical analysis for the case of uniform distribution and empirical analysis for other distributions. The main advantage of the proposed algorithm is that it is immutable to data distribution and does not need to support any particular distribution of the length.

In addition, we show how to separate completely the concept of network location of data from the search functionality. This separation is important, for instance, for building global storages where data is owned by multiple parties and each party is interested in keeping control over the aspects of physical storage and access to the data it owns. So in contrast to DHT, insertion of new data does not require relocation to an existing node.

Introduction

The scalability of a computer system is its ability to handle a growing amount of work [4]. It can also be referred to as the capability of a system to increase its performance under an increasing load. An analogous meaning is implied when the word "scalability" is used in an economic context, where scalability of a company implies that the underlying business model offers the potential for economic growth within the company. The concept of scalability is desirable in technology as well as in the business setting.

Data storage is a central part of any computer system. The scalability of any computer system strongly depends on the ability to scale its data storage. Traditionally, in most cases developers use relational database management systems (RDBMS) as a data storage. Unfortunately, scalability of RDBMS has inherent limitations [7]. The reason is that to support transactions and consistency, RDBMS need a central coordination point such as a transaction manager, which becomes a bottleneck with a growing number of servers on which RDBMS are deployed. Moreover, most RDBMS have an architecture in which servers have to share data with each other and therefore have to be synchronized, and this is a difficult task for a large number of servers.

To overcome the scalability problems, it was decided to sacrifice some part of functionality for a better performance [5]. Thus NoSQL databases appeared. The most important class of "NoSQL" databases is the key-value stores. Such systems mainly support two operations: retrieving and storing data by a given key. Many popular internet services are based on key-value storages. For example, the messaging system of Facebook and store of dashboard messages of SoundCloud both use Apache Cassandra; many products of Google such as Gmail, YouTube, Google Maps are based on BigTable, which is also a key-value store.

One of the possible ways to implement scalable key-value storage is to use a distributed hash-table (DHT). The distributed hash table is a class of systems which allows you to store and to search data by key in a set of network nodes [17]. A particular implementation of DHT is a protocol which specifies how nodes communicate and how they form an overlay network. It is important that every node doesn't know a full topology of the network. Instead, every node stores only a small amount of information about the network, which in turn is stored in a routing table. Every node has its own key − ID, which typically is a hash value, calculated from its IP address. The data with some key k is placed on the node which ID is closer to k than ids of other nodes in terms of some distance function d. The search is performed using a greedy algorithm by passing query message from one node to another, based on the list of nodes stored at every node in the routing table. Different DHT implementations use different distance functions. For instance, it can be a simple difference between two numbers x and y $d(x, y) = (x − y) \mod n$ [16] or it can be an XOR-metric defined as $d(x, y) = x \oplus y)$ [14].

The main advantage of all DHT implementations is that the expected number of nodes that a query message should pass before it reaches the destination node is $log(n)$. Here n is the total number of nodes in the network. Moreover the maximum size of a routing table at any node is also proportional to $log(n)$. These two properties together make DHT scalable. So the insertion of new nodes adds a very small overhead to the whole performance of the system.

The secret of these two properties lies in the structure of the routing table that is supported at every node. The main idea of all implementations of DHT is to maintain the routing tables in such a way that the i-th record in the routing table (link) of a particular node with id A points to the node with id B such that the distance from A to B belongs to the interval $2^i < d(A, B) \le 2^{i+1}$. This gives what in turns corresponds to the power law probability distribution of links length $P \sim d^{-1}$. When a new node is being inserted, DHT explicitly forms the routing table of the new node and updates the routing tables of the existing nodes according to this distribution. Together, all nodes with their routing tables form an overlay network which can be represented by graph $G(V, E)$, where the set of vertices V is a set of network nodes and E is a set of edges corresponding to the records in the routing tables. So any particular implementation specifies how to build graph G and how to support it.

As previously mentioned, all DHT implementations form i-th record in the routing table for node A explicitly by finding node B such that $2^i < d(A, B) \le 2^{i+1}$ in order to build a graph where the greedy walk algorithm can find a closest node to a given key with $log(n)$ steps.

In this paper we describe an algorithm which constructs a graph with similar navigation properties more naturally and more simply without an explicit support of a particular distribution of the links length. We apply our approximate algorithm suitable for a general metric space [12, 13] with a slight modification of the insertion algorithm to the exact search in the one-dimensional case. Unfortunately, the theoretical analysis of the search algorithm for a general case of Metrized Small

World is still absent. In this paper in section 4 we provide such analysis for the one-dimensional case. In the section 5, we give an empirical study of degree distribution, independence from data distribution, stability analysis and dependency of the search algorithm complexity on its parameters.

Note that previously the algorithm for the case of one dimension was mentioned and used as a basic algorithm for building a multi-attribute data structure in the paper [11]. This paper wrongly refers to the paper with the title "Single-attribute distributed Metrized Small World data structure" where this algorithm should be described in detail; however this paper by mistake was not published. In the presented paper we want to fill this gap. For consistency purposes. We provide this algorithm in section 2. We describe the insertion and search algorithms in detail with the complex empirical study and theoretical analysis.

We also suggest how to completely separate the concept of network location of the data from the search functionality. This separation is important, for instance, for building global distributed storages where data is owned by multiple parties and each party is interested in keeping control over the aspects of physical storage and access to the data it owns. So in contrast to DHT, insertion of new data does not require relocation of the existing node. In this way, the proposed overlay network with the search and insertion algorithm can be considered as a data structure built onto a data set.

The rest of the paper is structured as follows. Section 1 describes the Metrized Small World overlay data structure and its mapping to the network node collection. Section 2 specifies the data insertion algorithm. Section 3 describes the data search algorithms. In section 4 we prove that the search algorithm has $O(logn)$ complexity. Section 5 provides the experimental data and gives an analysis of the properties of the proposed structure. Finally, we summarize our contributions in the Conclusion.

1. One-dimensional Metrized Small World network

The distributed Metrized Small World (MSW) structure consists of a number of fully interconnected network nodes (so that each network node can communicate directly with another node). Each node contains a number of data units. Both nodes and data units have identifiers unique in the context of the whole structure. Node identifiers are resolved to network addresses (e.g. DN to IP) by an external system (such as DNS). A data unit identifier contains the identifier of a node on which the unit is allocated, thereby allowing instant identification of a net-

work node containing the data unit. This makes it possible to create an overlay structure on data unit level, which is inherently mapped to the physical network nodes. Each data unit A stores a mutable set $N(A)$ of identifiers of other data units, which are the links that constitute the overlay network. The overlay network can be considered as a graph $G(V, E)$, where the set V corresponds to the set of all data units and the set E contains an edge (A, B) if data unit A has a link to data unit B.

All links are bidirectional, i.e. each data unit is linked with data units linked with it. Nodes are not directly aware of each other; identifiers of other nodes can only be obtained from the links to data units allocated on those nodes.

An illustration of the structure is given on *Figure 1*. The structure is allocated on three network nodes identified by unique URL addresses. Data units are allocated on these nodes and also have unique URL addresses which are sub-URLs of the nodes. The overlay structure is built on the data units; the nodes do not have direct links to each other. The algorithms of data unit insertion and search operate solely on the overlay structure formed by data units, and links between them are completely agnostic to the physical node collection supporting the structure (data unit identifiers are treated as atomic by the algorithms). The choice of a node to allocate a new data unit or the decision to add a new node are completely independent of the data unit insertion algorithm which is executed after the new data unit is allocated. The introduction of a new node does not affect the existing overlay structure or the distribution of data units between nodes; it merely provides new allocation space and processing power for insertion and search algorithms.
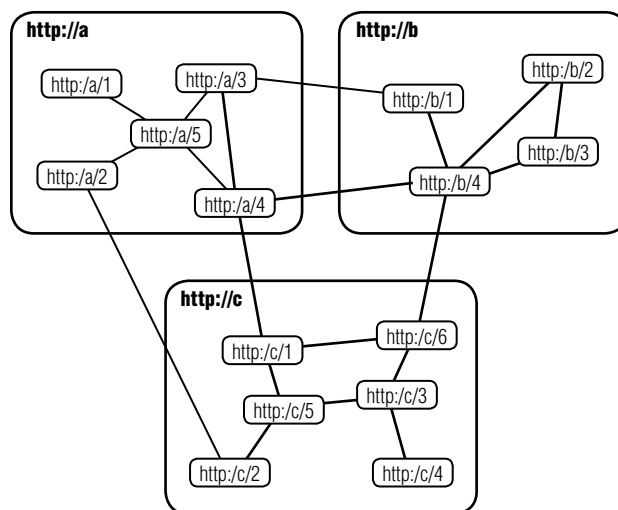
Fig. 1. An overlay network structure constructed on a set of data units

The overlay structure is accessed in a locally iterative manner, starting from an arbitrary known data unit (entry point) and following the links contained in the visited data units. To find a relevant data unit, one must choose an entry point and navigate through the structure, choosing a link at each step.

To make this process efficient, two requirements must be met:

1. There should be a relatively short path between the entry point and the target data unit. Since neither the entry point nor the target data unit are known beforehand, a short path must exist between any pair of data units in the structure;

2. A criterion must exist which would allow us to decide which link at each step will bring the search process closer to the result. It is also desirable that the gradient of the proximity criterion corresponds to the shortest path to the result in the structure, avoiding false local minimums when possible.

To address the first requirement, we have developed a data unit insertion algorithm that incrementally produces a structure with small-world properties. The small-world graphs [1, 2, 19] have the essential property that they have a short path between any two vertices while most vertices are not neighbors of each other.

To address the second requirement, a data unit contains searchable information and non-searchable information. As a model of searchable information in this paper, we consider the set of objects U that can be bijectionally mapped to the unit interval of real numbers. For example, U can be a set of all possible strings $S = (s_1, ..., s_n)$, where s_i is the index of a symbol in the alphabet B. If we define a mapping,

$$C(S) = \sum_{i=1}^{n} \frac{s_i}{|B|^i}.$$

For the purposes of the description of algorithms, we will denote as $C(D_i)$ a mapped value of the searchable information of the data unit D_i. We used a simple metric M between data units which is calculated as $M(D_i, D_j) = |C(D_i) - C(D_j)|$. This metric is used both by the insertion and search algorithms. Such a simple model of the searchable information allows us to make the search exact, since there is a natural linear order on the set of real numbers. This order is related to the metric M. Based on the linear order, the insertion algorithm has an ability to find the exact Voronoi's neighbors for every data unit. That is the main advantage of this model in contrast with an approximate search in a general metric space [10, 12, 13], where it is hard to find the Voronoi's neighbors exactly.

However, this approach is useful for the key-value storages and systems like CAS (Content Addressable Storage) [18] where a content-derived location-agnostic identifier of the data object, e.g. hash-code, is transformed into a content-agnostic location identifier, e.g. network or disk address, by the use of which the data object is retrieved. In terms of the MSW structure, the content-derived identifier would constitute the searchable information of the data unit and the content would be the non-searchable information. As a result of the search process, the content-derived identifier will be transformed into the identifier of the data unit which stores the content.

Other comparable solutions are the distributed hash table (DHT) [3] systems such as Chord [16]. These systems use a consistent hashing approach [8] to decide which network node will store the data unit. Insertion of a new network node requires relocation of K/n number of data units on average where K is the total number of data units and n is the number of nodes. In the MSW system, by contrast, insertion of a new node does not affect existing nodes because allocation of data units is completely irrelevant to the search and insertion algorithms. The new node becomes involved in the structure when a data unit allocated on the new node is added to the MSW structure using an entry point on one of the existing nodes. Another difference from the DHT systems is that MSW creates an overlay network on the data unit level while the overlay network of DHT systems like Chord is built on the node level.

A detailed description of the technological aspects of a scalable database system based on the MSW structure is presented in [9].

2. Data insertion algorithm

The purpose of the data unit insertion algorithm is to connect a newly allocated data unit to the structure by populating its list of links and correspondingly extending link lists of the data units chosen to be connected with the new data unit so that the greedy walk algorithm can find any data unit starting from any data unit. When there is an ordering relationship R related to the distance metric (for example the natural order of rational numbers), we can ensure the greedy search is guaranteed by connecting each data unit with its direct predecessor D_{pre} and direct successor D_{succ} — in contrast to an approximate search for a general metric space. In order to make the search algorithm faster, we additionally connect a new data unit with other m closest data units and do not remove these links during evolution of the structure.

The pseudo code of the algorithm is presented below.

Insertion_with_Order

Input: D_{new} − new data unit; D_{ep} − entry point data unit; m − parameter (small natural number).

1. Let $D_{cur} = D_{ep}$.

2. For each $D \in N(D_{cur})$ calculate $M_i = M(D_i, D_{new})$.

3. If $min(M_i < M(D_{cur}, D_{new})$ let $D_{cur} = D_i$ for which $M_i = min(M_i)$ and go to step 2.

4. If $C(D_{cur}) < C(D_{new})$ let $D_{pre} = D_{cur}$ and let D_{succ} be the direct successor of D_{new} chosen from the neighbors of D_{cur}.

5. If $C(D_{cur}) > C(D_{new})$ let $D_{succ} = D_{cur}$ and let D_{pre} be the direct predecessor of D_{new} chosen from the neighbors of D_{cur}.

6. Connect D_{new} with D_{pre} and D_{succ} if they exist.

7. Repeat m times:

    7.1. If D_{pre} exists, let D'_{pre} be the direct predecessor of D_{pre} chosen from its neighbors.

    7.2. If D_{succ} exists, let D'_{succ} be the direct successor of D_{succ} chosen from its neighbors.

    7.3. If none of D'_{pre} and D'_{succ} exist then break.

    7.4. If only D'_{pre} exists or $M(D'_{pre}, D_{new}) < M(D'_{succ}, D_{new})$ connect D_{new} and D'_{pre}, and let $D_{pre} = D'_{pre}$.

    7.5. If only D'_{succ} exists or $M(D'_{succ}, D_{new}) < M(D'_{pre}, D_{new})$ connect D_{new} and D'_{succ}, and let $D_{succ} = D'_{succ}$.

At each iteration, the algorithm obtains the neighbors of the current data unit D_{cur} closest to D_{new} and calculates the distance M_i to each of them. After that, the algorithm updates D_{cur} and makes the next iteration. So, we use a kind of greedy search algorithm to find the data unit closest to the new data unit (after all iterations the closest to D_{new} data unit will be in D_{cur}). We find out whether D_{cur} precedes or follows D_{new} and connect the new data unit with its direct predecessor D_{pre} and direct successor D_{succ}. Finally, we connect the new data unit with m closest existing data units from the right (succeeding data unit) and left (preceding data unit) side of it.

3. Search algorithm

The greedy search algorithm finds a data unit with a string closest to a given string S_q starting from the entry point data unit D_{ep}. If there is a data unit with a string identical to S_q present in the structure, the greedy algorithm will return this data unit. For brevity, we will treat each data unit as if it were the same entity as the searchable string which it contains.

Greedy_Search

Input: S_q − key; D_{ep} − entry point data unit.

Return: data unit which is closest to S_q.

1. Let $D_{cur} = D_{ep}$.

2. For each $D \in N(D_{cur})$ calculate $M_i = M(D_i, S_q)$.

3. If $min(M_i > M(D_{cur}, S_q)$, then return D_{cur}.

4. Let $D_{cur} = D_i$ for which $M_i = min(M_i)$ and go to step 2.

The greedy search algorithm described above relies on the fact that each data unit is always connected by the insertion algorithm with its direct predecessor and successor in the order defined by natural order of values $C(D_i)$. This ensures absence of false local minimums in the structure. In section 4, we will show that the complexity of this search algorithm grows logarithmically with the number of data units in the structure.

Since all data units are ordered, it is possible to define an operation that retrieves all data units from the given interval. We will denote the direct predecessor of element x as $D_{pre}(x)$ and the direct successor of x as $D_{suc}(x)$. The pseudo code of range search is presented below.

Range_ Search

Input: L_q − left bound; R_q − right bound; D_{ep} − entry point data unit.

Return: the set of data units from interval $[L_q, R_q]$.

1. Let $D_{cur} = Greedy\,Search(\dfrac{L_q + R_q}{2}, D_{ep})$

2. $T = \varnothing$; $D'_{cur} = D_{cur}$

3. While $D_{cur} \geq L_q$ do $T = T \cup D_{cur}$ and let $D_{cur} = D_{pre}(D_{cur})$;

4. Let $D_{cur} = D_{suc}(D'_{cur})$.

3. While $D_{cur} \leq R_q$ do $T = T \cup D_{cur}$ and let $D_{cur} = D_{suc}(D_{cur})$;

4. Return T.

The idea of the range search algorithm is very simple. We start from the center of the interval, go left until we reach the left bound and collect all data units to the set T. After that, we do the same in the opposite direction until we reach the right bound and return the set T as a result.

4. Theoretical analysis of the search algorithm

Theorem. The algorithm Greedy_Search has $O(log\,n)$ complexity in the structure formed by the construction algorithm Insertion_with_Order.

Proof. Let us assume that the images $C(D_i)$ are uniformly distributed in the segment $[0, 1]$. The goal is to demonstrate that each time the size of the structure doubles, the average number of steps required to reach

any data unit from any other data unit increases by an amount bounded by a constant, thereby producing logarithmic search complexity growth. First, we consider the situation when we have the set of N first data units in the structure. We call this set the first generation. Let us assume that we can reach any data unit from any other data unit in at most P steps. The probability density of the metric distance L between adjacent (in terms of metric) data units falls exponentially, so if we ignore the intervals between adjacent data units R times larger than average (i.e. larger than R/N) we omit only exponentially small number of cases. We thereby ignore the rare cases of very large gaps between adjacent data units in the first generation.

Next, we add N of the second generation data units, i.e. double the number of data units. Consider the largest interval between adjacent data units in first generation. The number of second generation data units k falling into the interval conforms to the Poisson distribution

$$p(k) = \frac{\lambda^k}{k!} e^{-\lambda} \text{ for large } N.$$

Since the length of the interval is R/N we have the expected value of the number n_{II} of second generation data units falling into the interval $\lambda = R$ and hence

$$p(k) = \frac{R^k}{k!} e^{-R}.$$

If we assume that $n_{II} < M$ we will be wrong only in a small number of cases if M is large enough. Now we will determine the upper bound U on the number of steps required to reach any data unit from any other data unit after the insertion of the second generation data units. Consider the worst case when both initial and target data units belong to the second generation. We are taking into account only the links between adjacent first generation data units, adjacent second generation data units and adjacent first and second generation data units, which

are a subset of all links in the structure, which further increases U. The links between adjacent first generation data units are longer in terms of metric than the links between adjacent data units in the second generation.

Hence, the former links are preferable to the latter in reaching the target data unit in the smallest number of steps. At most, M steps are required to reach the closest first generation data unit from the initial data unit, at most, P steps are required to reach the first generation data unit closest to the target data unit, and again, at most, M steps are required to reach the target data unit from the closest first generation data unit.

Therefore, at most $P + 2M$ steps are required to reach any data unit from any other data unit after the insertion of second generation data units. Thereby, we have shown that when the structure size is doubled, the number of search steps is only increased by the constant $2M$, which means that we have $O(\log n)$ search complexity where is the number of data units in the structure.

5. Simulations

We have implemented a single-attribute MSW structure. The simulations have been performed for different numbers of data units in the structure and for different values of parameter m of the insertion algorithm.

Figure 2 shows the average path length (number of steps) which the search algorithm needs to reach the data unit closest to the query. Networks with sizes in range from 1000 to 1000000 were generated by the Insertion_with_Order algorithm on the set of real numbers which has five different random distributions: uniform on the segment [0, 1], power law with power 0.5; normal distribution with mean 0.0 and standard deviation 1.0; log-normal with mean 4.0 and standard deviation 0.5; and points distributed normally with deviation 1.0 around 10 centers {-10.0, -7.0, -5.0, 0.0, 1.0, 2.0, 10.0}.
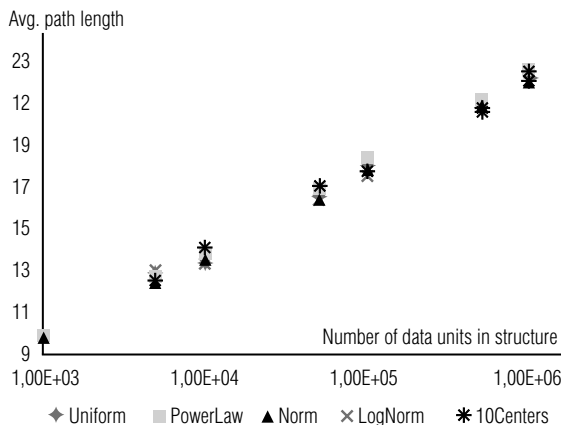
Fig. 2. Average path length for different distributions
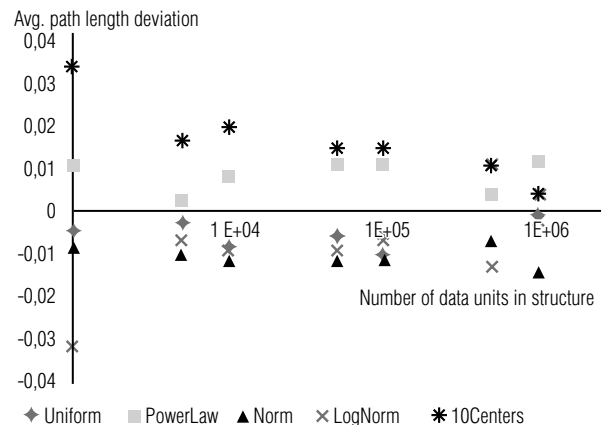
Fig. 3. Deviation of average path for different distributions
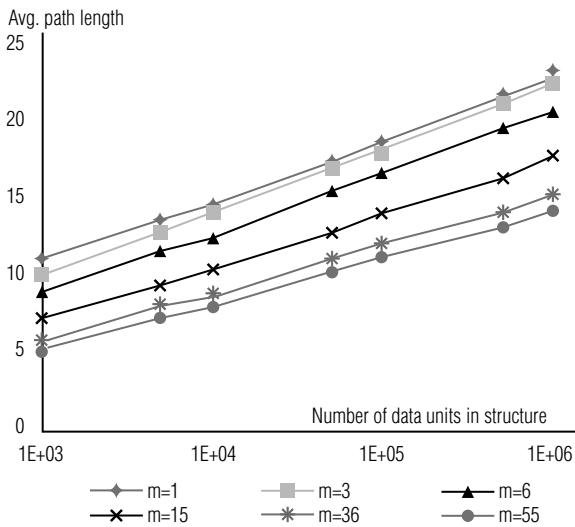
Avg. path length

Fig. 4. Average path length of search algorithm

The deviations of the path length from the average value for the different distributions are shown in *Figure 3*.

As can been seen from *Figure 2* and *3*, the structure demonstrates clear logarithmic dependency from the structure size with very small deviation for all distribution types. Independence of different distributions was expected, since all algorithms of the MSW structure do not directly use the value of distance between data units and are based only on the relative order. Therefore, all remaining simulations have been performed on the sets of data units whose images have been uniformly distributed on the interval (0, 1).
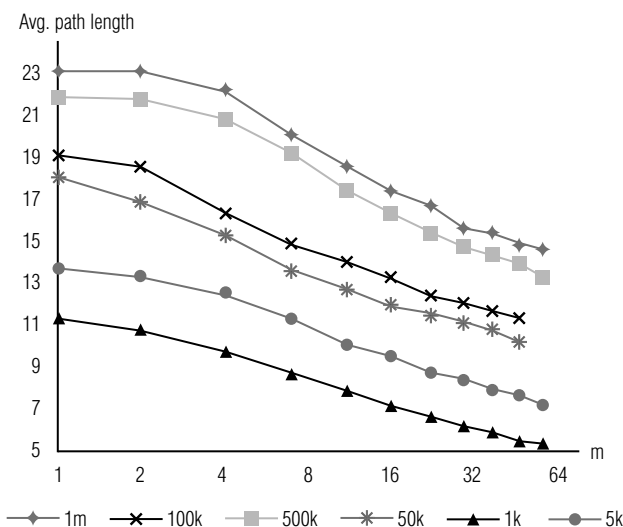
Avg. path length

Fig. 5. Dependency of path length for parameter m for fixed structure sizes 1K-1M

Figure 4 and *5* show dependency of the path length on the parameter *m*. F*igure 4* presents results for various

structure sizes and *Figure 5* for six different fixed sizes: 1 million, 500 thousand, 100 thousand, 50 thousand, 5 thousand and 1 thousand data units. As can be seen from *Figure 5*, the dependence of the average path length on the parameter *m* is close to inverse logarithmic law.

In order to study how the network nodes, which are presented as data units, may be loaded during the search, we have measured how often the search algorithm visits the particular data unit. The value of loading was calculated as a fraction of the number of searches in which the particular data unit has been involved to the total number of searches. As can be seen from *Figure 10*

Number of vertexs (data units)
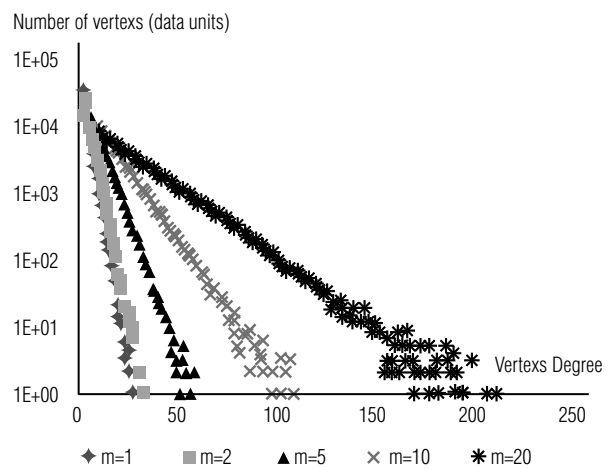
Fig. 6. Vertex degree distribution

and *11*, the maximum value of the data unit load is independent of the structure size and the fraction of load for a particular data unit has exponential dependence on the vertex degree (number of links to other data units). This indicates that additional links increase the diversity of search paths, thereby lowering the load on the data units with larger number of links, all of which helps to avoid bottleneck problems.

Max vertexs degree
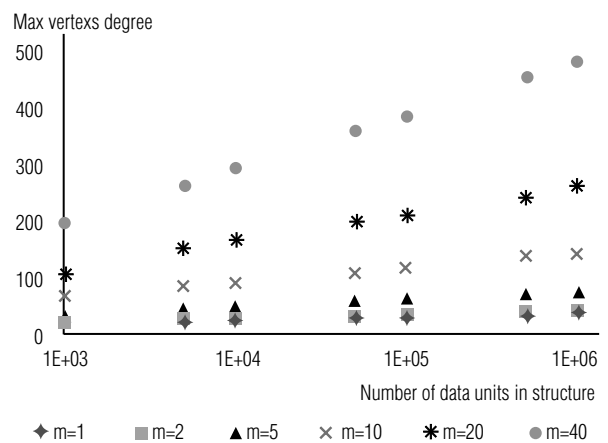
Fig. 7. Maximum vertex degree
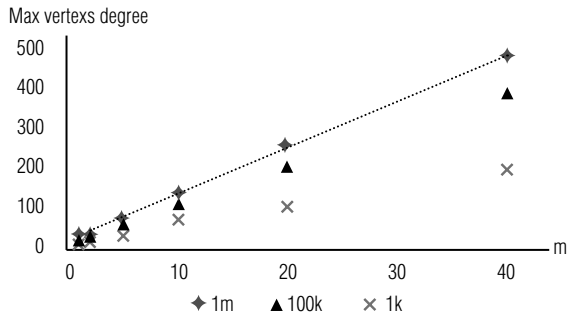
Fig. 8. Maximum vertex degree from parameter m
for 1 million, 100K and 1K data units in structure

The dependence of the maximum fraction load on the parameter m is plotted in *Figure 9* . The study of vertex degrees is presented in *Figure 6*, 7 and *8*. It is easily seen that the maximum vertex degree has a logarithmic dependence on the structure size (*Figure 8*) and has a linear dependence on the parameter m. *Figure 6* shows that the distribution of vertex degrees follows exponential law.
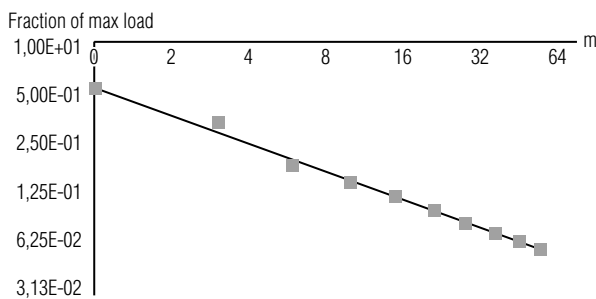
Fig. 9. Fraction of Max Load from parameter m
for 100K data units in structure
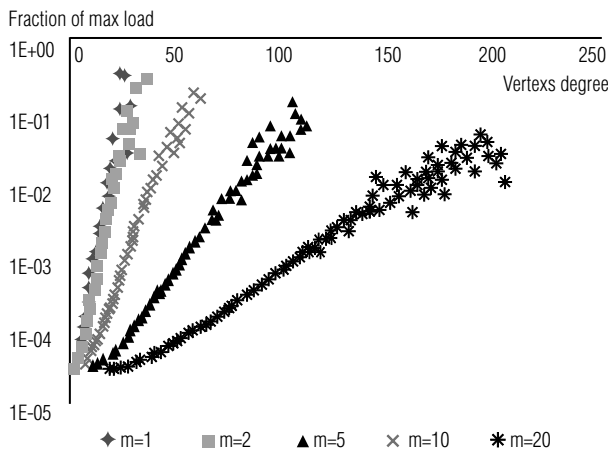
Fig. 10. Data unit load vs the number
of links for 100K data units in the structure

We have also investigated the stability property of the structure, i.e. its ability to work in case of node removal. *Figure 12* shows the average percentage of the number of nodes which should be removed (randomly with uni-

form distribution) to split the network into two components of connectivity. The average percent of removed data units follows the power law from the total number of data units in the structure. The value of the exponent in this power law also obeys the power law with the exponent equal to -1.24 empirically (*see Figure 13*). Therefore, we observe a relation $\frac{r}{n} \sim n^{c \cdot m^{-1.24}}$, where c is some constant ~ 1.7 and r is the expected number of removed data units for which the network is split.
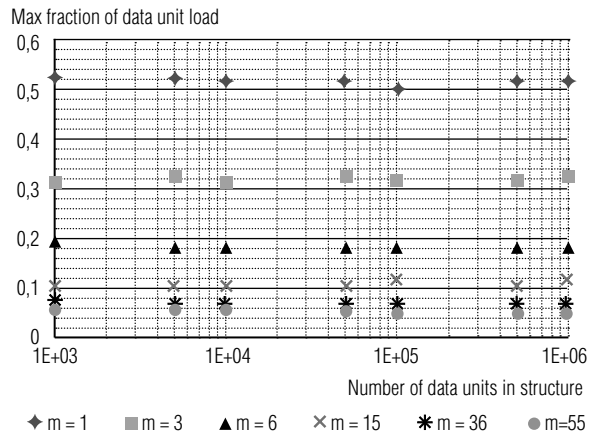
Fig. 11. Maximum data unit load
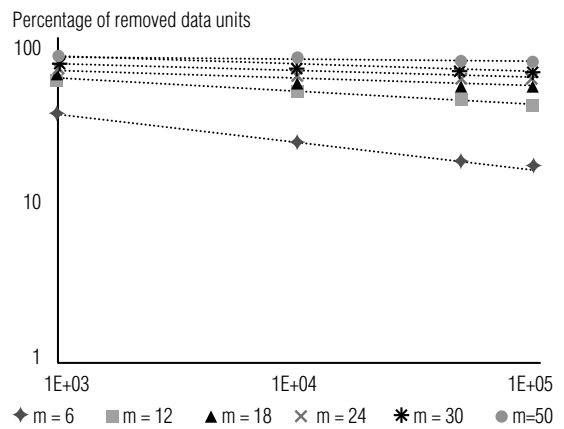
Fig. 12. The average percentage of how many nodes should
be removed randomly with uniform distribution to split a network
into two components of connectivity
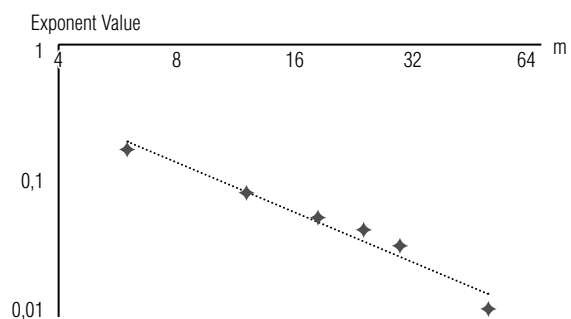
Fig. 13. The value of the exponent in the law of dependency
on the number of removed data units from the structure size

Conclusion

In this paper, we have introduced a variant of Metrized Small World data structure which allows exact searching for data units by a single unique string attribute. We have described data insertion and data search algorithms and have demonstrated that the structure has $O\left(\frac{\log n}{m^{\alpha}}\right)$ search complexity, where empirically $\alpha \sim 0.11$.

We have provided experimental data that demonstrates the scalability and decentralization properties of the structure. Our contributions can be summarized as follows:

✦ overlay structure on the data unit level with any number of entry points providing $O\left(\frac{\log n}{m^{\alpha}}\right)$ search complexity using a simple greedy search algorithm;

✦ the structure is independent of data distribution;

✦ an incremental data unit insertion algorithm modifies only a small amount of data units;

✦ the short paths for the greedy search algorithm are formed without explicitly maintaining any particular distribution of the length of links;

✦ physical allocation of data units is independent of data content and search functionality;

✦ dynamic insertion of new network nodes does not need a relocation of data units on existing nodes;

✦ search paths are distributed in a way that avoids overloading single data units.

Acknowledgments

References

1. Albert R., Barabasi A.L. (2002) Statistical mechanics of complex networks. *Reviews of Modern Physics*, vol. 74, no. 1, pp. 47−97.
2. Amaral L.A.N., Scala A., Barthelemy M., Stanley H.E. (2000) Classes of small-world networks. *Proceedings of the National Academy of Sciences (USA)*, vol. 97 (21), pp. 11149−11152.
3. Balakrishnan H., Kaashoek M.F., Karger D., Morris R., Stoica I. (2003) Looking up data in P2P systems. *Communications of the ACM*, vol. 46, no. 2, pp. 43−48.
4. Bondi A.B. (2000) Characteristics of scalability and their impact on performance. Proceedings of the *2nd International Workshop on Software and Performance (WOSP 2000), 17-20 September 2000, Ottawa, Canada*, pp. 195−203.
5. Cattell R. (2011) Scalable SQL and NoSQL data stores. *ACM SIGMOD Record*, vol. 39 (4), pp. 12−27.
6. Dean J., Ghemawat S. (2008) Mapreduce: Simplified data processing on large clusters. *Communications of the ACM*, vol. 51, no. 1, pp. 107−113.
7. Harizopoulos S., Abadi D.J., Madden S., Stonebraker M. (2008) OLTP through the looking glass, and what we found there. Proceedings of the *2008 ACM SIGMOD International Conference on Management of Data (SIGMOD 2008), 9-12 June 2008, Vancouver, Canada*, pp. 981−992.
8. Karger D., Lehman E., Leighton T., Panigrahy R., Levine M., Lewin D. (1997) Consistent hashing and random trees: Distributed caching protocols for relieving hot spots on the World Wide Web. Proceedings of the *29th Annual ACM Symposium on Theory of Computing, ACM, 4-6 May 1997, El Paso, Texas, USA*, pp. 654−663.
9. Krylov V., Logvinov A., Ponomarenko A., Ponomarev D. (2008) Active database architecture for XML documents. Proceedings of the *21st International Conference on Computer Applications in Industry and Engineering (CAINE 2008), 12-14 November 2008, Honolulu, Hawaii, USA*, pp. 244−249.
10. Krylov V., Logvinov A., Ponomarenko A., Ponomarev D. (2008) Metrized small world properties data structure. Proceedings of the *17th International Conference on Software Engineering and Data Engineering (SEDE 2008), 30 June - 2 July 2008, Los Angeles, California, USA*, pp. 203−208.
11. Logvinov A., Ponomarenko A., Krylov V., Malkov Y. (2010) Metrized small world approach for nearest neighbor search. Proceedings of the *2010 Spring/Summer Young Researchers' Colloquium on Software Engineering (SYRCoSE 2010), 1-2 June 2010, Nizhny Novgorod, Russia*, vol. 4, pp. 151−156.
12. Malkov Y., Ponomarenko A., Logvinov A., Krylov V. (2014) Approximate nearest neighbor algorithm based on navigable small world graphs. *Information Systems*, no. 45, pp. 61−68.
13. Malkov Y., Ponomarenko A., Logvinov A., Krylov V. (2012) Scalable distributed algorithm for approximate nearest neighbor search problem in high dimensional general metric spaces. *Lecture Notes in Computer Sciene − Similarity Search and Applications*, vol. 7404, pp. 132−147.
14. Maymounkov P., Mazieres D. (2002) Kademlia: A peer-to-peer information system based on the XOR metric. *Peer-to-Peer Systems*, vol. 2429, pp. 53−65.
15. Rabl T., Gomez-Villamor S., Sadoghi M., Muntes-Mulero V., Jacobsen H.A., Mankovskii S. (2012) Solving big data challenges for enterprise application performance management. *Proceedings of the VLDB Endowment*, vol. 5, no. 12, pp. 1724−1735.
16. Stoica I., Morris R., Karger D., Kaashoek M.F., Balakrishnan H. (2001) Chord: A scalable peer-to-peer lookup service for internet applications. *ACM SIGCOMM Computer Communication Review*, vol. 31, no. 4, pp. 149−160.
17. Tarkoma S. (2010) *Overlay networks: Toward information networking*. Boca Raton, FL: CRC Press.
18. Tolia N., Kozuch M., Satyanarayanan M., Karp B., Bressoud T.C., Perrig A. (2003) Opportunistic use of content addressable storage for distributed file systems. Proceedings of the *2003 USENIX Annual Technical Conference (USENIX 2003), General Track, 9-14 June 2003, San Antonio, Texas, USA*, vol. 3, pp. 127−140.
19. Watts D.J. (1999) *Small worlds: The dynamic of networks between ordered and randomness*. New Jersey: Princeton University Press.

Оверлейная сеть для распределенного точного и интервального поиска в одномерном пространстве

А.А. Пономаренко

научный сотрудник, Лаборатория алгоритмов и технологий анализа сетевых структур
Национальный исследовательский университет «Высшая школа экономики»
Адрес: Российская Федерация, 603093, г. Нижний Новгород, ул. Родионова, 136
E-mail: aponomarenko@hse.ru

Ю.А. Мальков

младший научный сотрудник
Институт прикладной физики, Российская академия наук
Адрес: Российская Федерация, 603950, г. Нижний Новгород, ул. Ульянова, 46
E-mail: yurymalkov@mail.ru

А.А. Логвинов

руководитель проектов, ООО «МераЛабс»
Адрес: Российская Федерация, 603093, г. Нижний Новгород, ул. Родионова, 192
E-mail: alogvinov@meralabs.com

В.В. Крылов

доктор технических наук, профессор, заведующий лабораторией больших данных
Нижегородский государственный технический университет им. Р.Е.Алексеева
Адрес: Российская Федерация, 603950, г. Нижний Новгород, ул. Минина, 24
E-mail: vkrylov@heterarchica.com

Аннотация

Для современной компьютерной системы возможность масштабироваться является необходимым бизнес-требованием. Распределенные системы явно демонстрируют это свойство, как и способность легко обрабатывать сверхбольшие объемы данных. Многие системы с распределенной архитектурой имеют в основе распределенную хэш-таблицу (distributed hash table, DHT), которая используется в качестве дополнительной логической сети, объединяющей множество распределенных серверов. Эта логическая сеть используется для поиска узлов и распределения задач между ними. Главное преимущество такого подхода – отсутствие какого-либо центрального элемента или узла, который обладал бы информацией о структуре всей сети. Поиск узлов в сети происходит путем передачи поискового сообщения от одного узла к другому. Несмотря на то, что каждый узел «знает» только небольшое число других узлов, сеть организованна таким образом, что процедура поиска затрагивает логарифмическое число узлов.

Существует несколько реализаций концепции DHT, которые регламентируют, каким образом логическая сеть строится и поддерживается. В этой работе мы демонстрируем, как такая сеть может быть сконструирована очень простым способом, с применением (с небольшими изменениями) недавно опубликованного алгоритма метризованного тесного мира (Metrized Small World) для случая одномерного метрического пространства. В работе мы приводим теоретический анализ для случая равномерного распределения данных и эмпирический анализ для других распределений. Главным преимуществом предложенного алгоритма перед аналогами является его устойчивость к различным распределениям данных и отсутствие необходимости поддержания распределения длин ссылок, определенного явным образом.

Также в работе описывается, каким образом можно полностью отделить физическое размещение данных от поисковой функциональности. Это разделение важно, например, для построения глобальных хранилищ данных, данными которых владеют несколько сторон, причем каждая сторона заинтересована в том, чтобы иметь полный контроль над физическим размещением и доступу к своим данным. В отличие от DHT, добавление новых данных не требует их перемещение на другие узлы.

Литература

1. Albert R., Barabasi A.L. Statistical mechanics of complex networks // Reviews of Modern Physics. 2002. Vol. 74. No. 1. P. 47−97.
2. Amaral L.A.N., Scala A., Barthelemy M., Stanley H.E. Classes of small-world networks // Proceedings of the National Academy of Sciences (USA). 2000. Vol. 97 (21). P. 11149−11152.
3. Looking up data in P2P systems / H.Balakrishnan [et al.] // Communications of the ACM. 2003. Vol. 46. No. 2. P. 43−48.
4. Bondi A.B. Characteristics of scalability and their impact on performance // Proceedings of the 2nd International Workshop on Software and Performance (WOSP 2000), 17-20 September 2000, Ottawa, Canada. 2000. P. 195−203.
5. Cattell R. Scalable SQL and NoSQL data stores // ACM SIGMOD Record. 2011. Vol. 39 (4). P. 12−27.
6. Dean J., Ghemawat S. Mapreduce: Simplified data processing on large clusters // Communications of the ACM. 2008. Vol. 51. No. 1. P. 107−113.
7. Harizopoulos S., Abadi D.J., Madden S., Stonebraker M. OLTP through the looking glass, and what we found there // Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data (SIGMOD 2008), 9-12 June 2008, Vancouver, Canada. 2008. P. 981−992.
8. Consistent hashing and random trees: Distributed caching protocols for relieving hot spots on the World Wide Web / D.Karger [et al.] // Proceedings of the 29th Annual ACM Symposium on Theory of Computing, ACM, 4-6 May 1997, El Paso, Texas, USA. 1997. P. 654−663.
9. Krylov V., Logvinov A., Ponomarenko A., Ponomarev D. Active database architecture for XML documents // Proceedings of the 21st International Conference on Computer Applications in Industry and Engineering (CAINE 2008), 12-14 November 2008, Honolulu, Hawaii, USA. 2008. P. 244−249.
10. Krylov V., Logvinov A., Ponomarenko A., Ponomarev D. Metrized small world properties data structure // Proceedings of the 17th International Conference on Software Engineering and Data Engineering (SEDE 2008), 30 June - 2 July 2008, Los Angeles, California, USA. 2008. P. 203−208.
11. Logvinov A., Ponomarenko A., Krylov V., Malkov Y. Metrized small world approach for nearest neighbor search // Proceedings of the 2010 Spring/Summer Young Researchers' Colloquium on Software Engineering (SYRCoSE 2010), 1-2 June 2010, Nizhny Novgorod, Russia. 2020. Vol. 4. P. 151−156.
12. Malkov Y., Ponomarenko A., Logvinov A., Krylov V. Approximate nearest neighbor algorithm based on navigable small world graphs // Information Systems. 2014. No. 45. P. 61−68.
13. Malkov Y., Ponomarenko A., Logvinov A., Krylov V. Scalable distributed algorithm for approximate nearest neighbor search problem in high dimensional general metric spaces // Lecture Notes in Computer Sciene − Similarity Search and Applications. 2012. Vol. 7404. P. 132−147.
14. Maymounkov P., Mazieres D. Kademlia: A peer-to-peer information system based on the XOR metric // Peer-to-Peer Systems. 2002. Vol. 2429. P. 53−65.
15. Solving big data challenges for enterprise application performance management / T.Rabl [et al.] // Proceedings of the VLDB Endowment. 2012. Vol. 5. No. 12. P. 1724−1735.
16. Chord: A scalable peer-to-peer lookup service for internet applications / I.Stoica [et al.] // ACM SIGCOMM Computer Communication Review. 2001. Vol. 31. No. 4. P. 149−160.
17. Tarkoma S. Overlay networks: Toward information networking. Boca Raton, FL: CRC Press, 2010. 237 p.
18. Opportunistic use of content addressable storage for distributed file systems / N.Tolia [et al.] // Proceedings of the 2003 USENIX Annual Technical Conference (USENIX 2003), General Track, 9-14 June 2003, San Antonio, Texas, USA. 2003. Vol. 3. P. 127−140.
19. Watts D.J. Small worlds: The dynamic of networks between ordered and randomness. New Jersey: Princeton University Press, 1999. 168 p.

Solution of integration tasks in enterprise management information systems

Valerij Yu. Vahrushev
Dean, Faculty of Automation and Computing Machines
Vyatka State University
Address: 36, Moskovskaya Street, Kirov, 610000, Russian Federation
E-mail: vahrushev@vyatsu.ru

Elena V. Bogomolova
Leading Programming Engineer, LEPSE Company
Post-graduate, Department of Automatics and Telemechanics, Vyatka State University
Address: 36, Moskovskaya Street, Kirov, 610000, Russian Federation
E-mail: elenaoops@mail.ru

Anna M. Lanskih
Associate Professor, Department of Automatics and Telemechanics
Vyatka State University
Address: 36, Moskovskaya Street, Kirov, 610000, Russian Federation
E-mail: usr00221@vyatsu.ru

Yurij V. Lanskih
Associate Professor, Department of Automatics and Telemechanics
Vyatka State University
Address: 36, Moskovskaya Street, Kirov, 610000, Russian Federation
E-mail: lyuv@inbox.ru

Andrej V. Luppov
Associate Professor, Department of Automatics and Telemechanics
Vyatka State University
Address: 36, Moskovskaya Street, Kirov, 610000, Russian Federation
E-mail: dei123@mail.ru

Anna V. Malysheva
Lecturer, Department of Automatics and Telemechanics
Vyatka State University
Address: 36, Moskovskaya Street, Kirov, 610000, Russian Federation
E-mail: annyml@mail.ru

Natalya A. Shmakova
Senior Lecturer, Department of Automatics and Telemechanics
Vyatka State University
Address: 36, Moskovskaya Street, Kirov, 610000, Russian Federation
E-mail: shmakova.natalya@mail.ru

Abstract

The authors consider the problems of constructing systems of accounting, management and analysis of a company. One of the most cost-effective ways to build such systems is the integration of legacy, acquired and developed components on a middleware base, which can be characterized as a technological platform for integration.

Consideration is given to the formalization of principles for creation of such platform software. These principles are formed similarly to the principles of forming a microprogram processor control unit. We are shown how to develop a software platform that makes it possible to organize the launch of procedures for processing, visualization, data exchange by a schedule or by user request, using the microprogram processor control unit as a formal model. The totality of the information and software that acts as a platform is decomposed into instructions memory, metacommand memory, the instruction memory controller and metacommand memory controller.

A document management system is considered as one of the options of platform software. The electronic document management system for the exchange of information is protected by digital signature. It facilitates integration between information subsystems of reference, accounting and other purposes. This practice is called EDMS-bus.

Within the theoretical formulation of the problem the authors considered the informational domain model as a set of business processes performed in it. The automation of activity in the subject field amounts to forming a set of automated business processes, whose power is determined by the functionality of the implemented information system, and which cover to the maximum extent a set of business processes of the subject area. Examination is made of cases of overlapping sets of legacy information systems business processes and of implemented information systems business processes.

Introduction

Formation of enterprise information management systems is a necessary condition for improving the performance efficiency of enterprises in various areas. A specific feature for this is to drive cost savings to the maximum during introduction of new and updated old information systems and to use existing developments to the extent possible.

Corporate enterprise applications cannot exist separately. In integration of applications, IT developers need to ensure reliability of the data transmission network, provide an acceptable data rate, take into account the application differences, and make it possible to adapt to alteration of merged applications. To ensure the above requirements, the following basic approaches are identified: file transfer, shared database, remote procedure invocation, messaging [1].

The approach addressed by the authors is an alternative approach which rejects introduction of a single information system which can fully satisfy all the information needs of the data domain given that for large enterprise information systems it is often impossible to find software that would cover the solution of all existing business tasks. The integration problems of heterogeneous information system modules, data porting between subsystems of the information systems, configuring information subsystems to develop additional information system components are solved using the aforementioned approaches. In practice, the solution of such problems at large enterprises in many cases is unique and specific, despite of a large selection of ready-made solutions for application integration using middleware proposed by developers of large and expensive enterprise systems (IBM, MS Biztalk, SAP, Oracle and others.). Replacement of properly functioning enterprise subsystems of various platforms with a multipurpose ERP system is not always justified due to possible overpayment for excessive functionality. In addition, special offers can successfully solve the problem of narrow-profile business tasks. Experience in solving problems of accounting, analytics, information exchange and integration of enterprise information subsystems makes it possible to put forward formal information and structural platform-supported models for effective solution of the enterprise subsystem integration tasks.

1. Theoretical statement of the problem

The authors have already addressed a formal characterization of the synchronization problem of reference information from several enterprise information subsystems [2].

Let us form an information model of the domain area as a set of business processes which will be conditionally defined as "idealized":

$$OF = \{BP_i\}.$$

Consequently, the information model of the information system available at the time of statement of the automation problem ('as is' model) can be presented as a set of real system-supported business processes:

$$IS' = \{BP_i'\}.$$

The solution of the automation problem should result in obtaining an information system whose information model will be also presented as a new set of business processes (the 'to be' model), which, in general, is not consistent with idealized ones:

$$IS'' = \{BP_i''\}.$$

With regard to each business process BP_i'' it is worthy of note that it is identical to the initial one BP_i', if not subject to changes in the course of solving the automation problem, and, otherwise, is not identical to the initial one.

Each subsystem of the information system computerizes some subset of the set of computerized business processes:

$$ISS_j'' = \{BP_i^{j''}\}.$$

In the optimal case, the subsystem functions disjoint, making it possible to entirely cover all functional requirements for the introduced system:

$$\bigcap_j BP_i^{j''} = \varnothing$$

$$\bigcup_j BP_i^{j''} \equiv \{BP_i\}.$$

In reality, the formation of the information system through a sequential selection of components results in inability to accomplish the last two expressions.

Among other things, there can be situations of mutual redundancy of the implemented components, and their partial redundancy and failure. The most common (and commonly occurring) case is the following:

$$\begin{cases} \bigcup_j BP_i^{j''} \setminus \{BP_i\} \neq \varnothing \\ \{BP_i\} \setminus \bigcup_j BP_i^{j''} \neq \varnothing, \end{cases}$$

when the implemented set of subsystems does not completely computerize all enterprise business processes, but along with this there are functions of the implemented subsystems which are not necessary in terms of information needs of the enterprise.

Functional insufficiency of the implemented subsystems is compensated by in-house developments, so we will consider the target information system as a set of purchased subsystems

$$ISSb_j'' = \{BPb_i^{j''}\}$$

and developed subsystems

$$ISSd_k'' = \{BPd_i^{k''}\}.$$

The automation activity should result in establishing an information system for which the following condition is to be met:

$$\bigcup_j BPb_i^{j''} \cup \bigcup_k BPd_i^{k''} \supseteq \{BP_i\}.$$

The tasks of selecting purchased components and designing components being developed are actually limited to the problem of multi-criteria optimization in which specifically such hard-to-formalize variables as the degree of risk of introducing unpopular software systems, usability of a particular subsystem, etc. shall be considered.

2. Structural model of information system management

A structural model of the technological information system platform focused on solution of integration problems of legacy, developed and purchased components is proposed. Decomposition of functions performed by the system and function modules for storage, processing, data transmission and management functions [3], as well as the methodology of synthesis of firmware processor control devices lay the foundation for the proposed model tested by the authors in solution of a number of problems [4, 5].

The platform control module (CM) is formed per a known structure of the firmware processor control device [6] (*Figure 1*).

Let us call an analogue of the firmware program of the processor control unit (CU) a CM metaprogram. Just as the firmware program is a set of microcommands, and microcommands in turn are a set of microoperations, let us represent the metaprogram as a set of metacommands consisting of sub-operations. A command sent to CM contains an operation code (OC). A metacommand
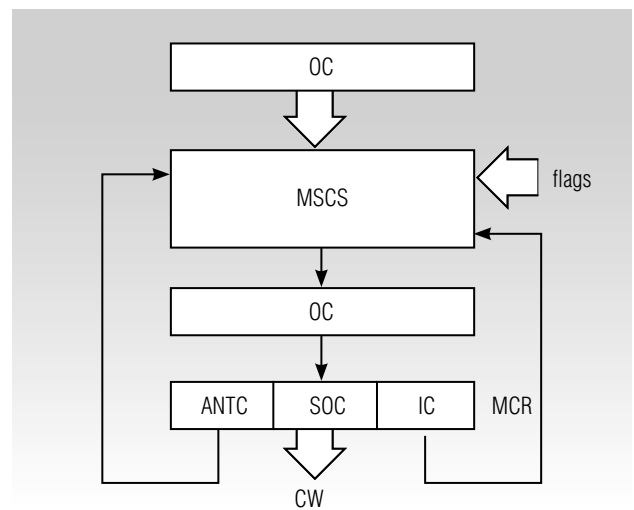
Fig. 1. Structure of CM for information system platform

sequence controller (MCSC) selects a metacommand from the metacommand memory (MCM) and places it in the metacommand register (MCR). The metacommand consists of an address of the next metacommand (ANTC), sub-operation code (SOC), which generates a control word (CW) supplied to the processing module, as well as the identifier code (IC).

MCM is a metadata base. MCM is characterized by associativity (i.e. a possible memory call by values of various gadgets), and structural hierarchical pattern. MCM recording format:

$$MetaCmd = <MCCode,\ NextMCCode,\ SOCode,\\ \{<ParName,\ ParDef>\}>,$$

where *MCCode* is a metacommand code;

NextMCCode is a code of the next metacommand;

SOCode is a sub-operation code;

ParName is a name of the sub-operation parameter;

ParDef is a description of the sub-operation parameter.

The sub-operation code in the existing implementation is represented by names of the executable results processing and visualization modules. The description of the sub-operation parameters includes a type of parameter, its minimum and maximum values, change step, etc. The integration of purchased and developed modules on the proposed platform is based on consistency of modules (1) and (2) start format with this description.

The instruction memory (IM), which similar to MCM has associativity and hierarchical structure, stores the operation code and parameters. The IM recording format is as follows:

$$Op = <OpCode,\ \{<MCCode,\ \{<ParName,\\ ParValue>\}>\}>,$$

where *OpCode* is an operation code;

MCCode is a metacommand code;

ParName is a name of sub-operation parameter;

ParValue is a value of a sub-operation parameter.

Figure 2 illustrates the operation of the instruction and metacommand memory controllers in the course of sampling instructions from the instruction memory and their processing.

For data processing automation, the instruction memory should be generated automatically by the processing planning module (scheduler). The scheduler work is based on the existing set of operations, accessible ranges of parameters for each sub-operation, processed data (specifically, metacommands and sub-operation parameters, execution of which resulted in obtaining available information units), and planned (present in the instruction memory) processing.

These data indicate that the data memory (DM) storing the data analysis results should also possess the properties of associativity and hierarchy. The data memory recording format:

$$Data = <MCCode,\ \{<\{<ParName,\ ParValue>\},\\ DataBlock>\}>,$$

where *MCCode* is a metacommand code;

ParName is the name of a sub-operation parameter;

ParValue is a value of the sub-operation parameter;

DataBlock is a data unit.

The operation of the combined instruction/metacommand memory controller, data memory controller and scheduler in the process of planning instructions and their entry into the instruction memory is provided in *Figure 3*.

3. Use of an electronic document management system as a platform for applications integration in the enterprise information system

One example of implementation of the structural model considered is the use of an electronic document management system (EDMS) as an integration platform [7,
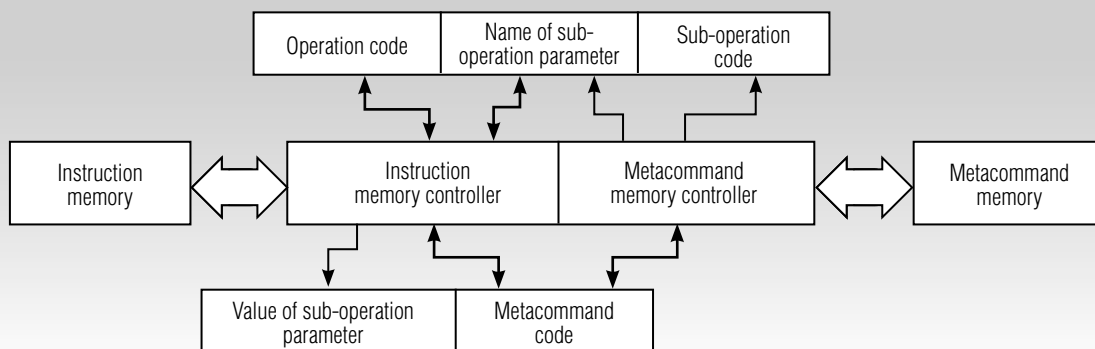
Fig. 2. Functionality of the instruction and metacommand memory controllers
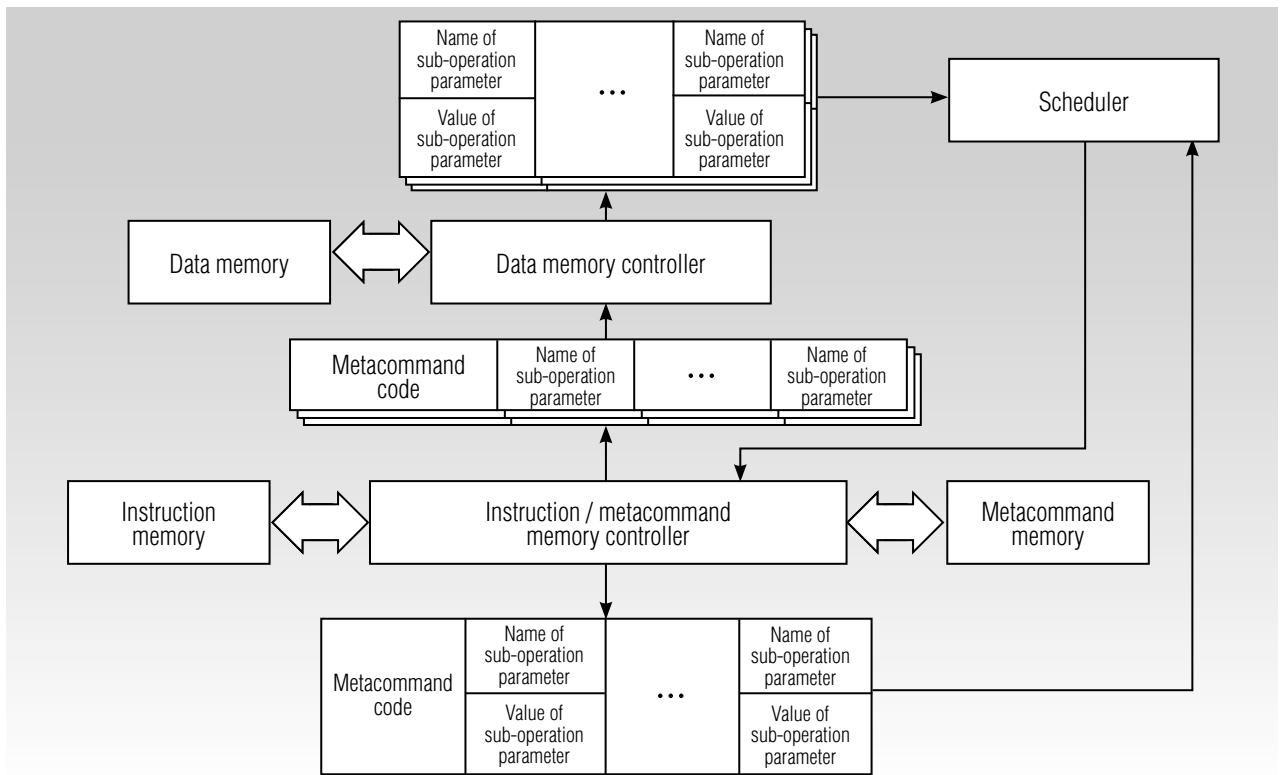
Fig. 3. Operation of the instruction/metacommand memory controller, data memory controller and scheduler

8], called in this case "SED-bus" providing an interaction of various systems (accounting, ERP, CAD, and others) via the document flow.

The experience of using EDMS at a large enterprise in accordance with the concept [7, 8] demonstrates a wide range of tasks for execution, including tasks of integration between the enterprise sub-systems:

1. Organization of fast network data exchange in electronic form within the enterprise, approved by an electronic signature. Use of information on observing the dates of document processing while generating reports which further makes it possible to improve performance discipline in taking management decisions and raises the overall efficiency of the enterprise.

2. Organization of exchange of all types of documents (including legally valid documents) with contracting parties and, therefore, acceleration of agreement of contracts, business correspondence, conclusion of transactions.

3. Extended handling of the help system used in various departments and software modules, limited by filtration and access rights appointed to roles (positions) or groups (departments). EDMS makes it possible to handle this reference information of the enterprise provided that EDMS manuals are synchronized with other systems in real time. These problems are solved within subsystems (2), which was addressed by the authors in a number of examples [2, 9]. The problems of integration are solved in a similar way using EDMS between the enterprise IS subsystems and an external IS.

4. The integration of EDMS with the accounting system, which resulted in ability to automatically start the document flow routes for approval or information distribution (documents or reports — accounting, legal, financial, relating to processing orders, works performed, payments made and shipped products).

5. Integration with budgeting, financial management, customer relationship management systems.

The use of standard EAI (Enterprise Application Integration) templates makes it possible to solve an unlimited number of problems within the problem area. The most famous EAI-solutions for integration of enterprise applications are IBM WebSphere MQ, Microsoft BizTalk Server, TIBCO, WebMethods, SeeBeyond, Vitria, Cross Worlds and others [1].

According to the EAI principle, while using the EDMS-bus as an intermediate integration layer (middleware) of the enterprise it is preferable to use IS replicated patterns. For example, for EDMS operating as a platform for enterprise application integration organized in the form of guaranteed delivery of tasks (independent of the senders and receivers) according to business routes, a standard approach, namely messaging, has been used.
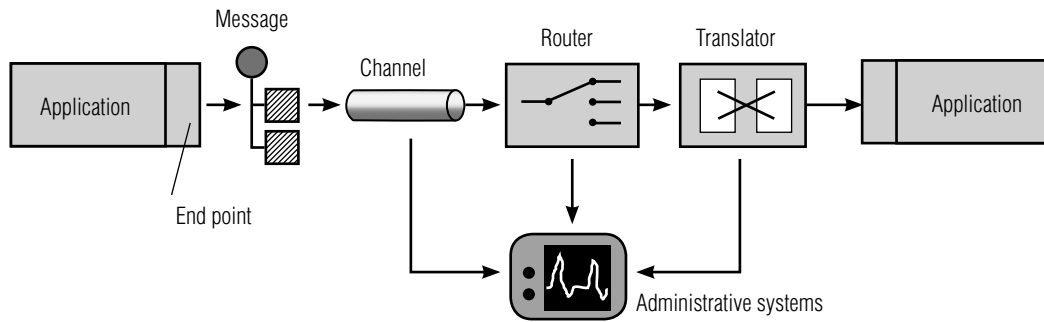
Fig. 4. Key elements of an integrated solution based on messaging [1]

Messaging in accordance with EAI concept is provided upon availability of a number of elements required for integrated solution (*Figure 4*).

When using the electronic document management system as a platform for application integration, the enterprise information system also uses a number of standard elements listed in *Table 1*.

Table 1.

Key elements of the integrated solution based on EDMS-bus

Elements	Purpose of element according to EAI provision	ERP – EDMS – counting system integration
Application (sender)	Platform	Informix ERP MAX+
Sender endpoint	Component which publishes a message in a message channel	IBM Informix ODBC Driver
Message (header, body)	Smallest data unit transmitted via a message channel	Data is transmitted as TCP/IP packets
Message channel	Channel of specific data transmission	Channel "point-to-point" with asynchronous message delivery to EDMS based on MS SQL Server
Message router	Special filter which retrieves a message from one message channel and places into another one (to control a load in several queues)	None
Translator	A special filter for data format conversion	Conversion of data format using SQL-type adapter
Receiver end point	Component enabling the application to receive a message	Microsoft FoxPro VFP ODBC Driver
Application (sender)	Platform	DBMS dBase VFP 9

The use of messaging technology ensures the following advantages [1]:

✦ platform and language integration (template is based on a message bus) - messaging system is responsible for data transmission;

✦ asynchronous interaction ("sent and forgot", no simultaneous availability of the sender and receiver is required);

✦ time mismatch (as opposed to synchronous, asynchronous interaction makes it possible to place and handle calls at different rates and without waiting for completion of their processing by the receiver);

✦ load control of the transmission channel (flow control);

✦ reliable interaction between applications (through approach with an intermediate storage);

✦ interposition between interacting applications with various language environments.

The following disadvantages of the messaging model technology can be pointed out:

✧ complicated programming model, which is split into numerous processing procedures;

✧ asynchronous interaction model is not suitable for all cases (including a need for implementing synchronous models);

✧ time and procedure of message delivery are not stipulated, so transmission delays are not excluded;

✧ significant time and resources consumed organizing corporate wide integration.

Conclusion

As can be seen from the above, the information model concerning the problem of forming the enterprise information system through integration of legacy, purchased and developed information subsystems was considered, the formal and theoretical basis of constructing integration platforms of the administrative-analytical systems was analyzed, and the structural model of the platform software was proposed. The implementation of the above principles of the middleware layer is possible by using modern electronic document management systems and standard templates of replicated applications for integration.

The prospects of further research are to solve the problems of integration of enterprise information system taking into account a possible reconfiguration of the system and replacement of the existing modules and subsystems in the future driven by new requirements imposed by a changing economic environment. ∎

References

1. Hohpe G., Wolf B. (2007) *Shablony integratsii korporativnykh prilozhenii* [Enterprise integration patterns]. Moscow: Williams (in Russian).
2. Lanskikh Yu.V., Bogomolova E.V., Tatarinova E.O. (2012) Reshenie zadach sistemnoj integracii pri avtomatizacii transportnogo ucheta v regionalnom predpriyatii svyazi [Meeting the challenges of system integration in the automation of vehicle accounting in the regional communications enterprise]. Proceedings of the *International Scientific Conference "Regions in the Conditions of Unsustainable Development"*, 1–3 November 2012, Sharya – Kostroma. Vol. 2. P. 188–193 (in Russian).
3. Balashov E.P., Puzankov D.V. (1987) *Proektirovanie informatsionno-upravlyayushchikh sistem* [Design of management information systems]. Moscow: Radio and Communications (in Russian).
4. Lanskikh Yu.V. (2012) Principy organizacii texnologicheskoj platformy i procedur obrabotki dannyx v informacionno-analiticheskoj sisteme [Principles of organizing the technological platform and data processing procedures in the information-analytical system]. *Herald of ITARK*, no. 1, pp. 24–28 (in Russian).
5. Lanskikh Yu.V., Nechaev D.S. (2004) Organizaciya upravleniya v sistemax intellektualnogo analiza dannyx [Organization of data mining system control]. *Herald of Vyatka Scientific Center. Series: Problems of Information Processing*, no. 1 (5), pp. 144–149 (in Russian).
6. Balashov E.P., Grigoriev V.L., Petrov G.A. (1984) *Mikro- i mini-EVM* [Micro- and mini-computers]. Leningrad: Energoatomizdat (in Russian).
7. Bogomolova E.V., Lanskikh Yu.V. (2014) Prakticheskie resheniya po ispol'zovaniyu sistemy elektronnogo dokumentooborota v informatsionnoi sisteme predpriyatiya [Practical solutions for using of an electronic document management system in the enterprise information system]. Proceedings of *Annual Scientific and Practical Conference "Society, Science and Innovations". Section "Design of Data Processing and Data Presentation Systems"*. 15–26 April 2014, Kirov. Available on CD-ROM, article no. 8 (in Russian).
8. Arif M., Kulonda D., Egbu C., Goulding J.S., Toma T. (2011) Enterprise-wide information system for construction: A document based approach. *KSCE Journal of Civil Engineering*, vol. 15, no. 2, pp. 271–280.
9. Bogomolova E.V., Lanskikh Yu.V. (2013) Prakticheskie resheniya po sinkhronizatsii spravochnoi informatsii v korporativnoi sisteme predpriyatiya [Practical solutions for synchronizing of reference information in an enterprise corporate system]. Proceedings of *Annual Scientific and Practical Conference "Society, Science and Innovations". University-wide section*. 15–26 April 2013, Kirov. Available on CD-ROM (in Russian).

Решение задач интеграции в информационных системах управления предприятием

В.Ю. Вахрушев
кандидат технических наук, декан факультета автоматики и вычислительной техники
Вятский государственный университет
Адрес: 610000, г. Киров, ул. Московская, д. 36
E-mail: vahrushev@vyatsu.ru

Е.В. Богомолова
ведущий инженер-программист ОАО ЭМСЗ «ЛЕПСЕ»
аспирант кафедры автоматики и телемеханики, Вятский государственный университет
Адрес: 610000, г. Киров, ул. Московская, д. 36
E-mail: elenaoops@mail.ru

А.М. Ланских
кандидат технических наук, доцент кафедры автоматики и телемеханики
Вятский государственный университет
Адрес: 610000, г. Киров, ул. Московская, д. 36
E-mail: usr00221@vyatsu.ru

Ю.В. Ланских
кандидат технических наук, доцент кафедры автоматики и телемеханики
Вятский государственный университет
Адрес: 610000, г. Киров, ул. Московская, д. 36
E-mail: lyuv@inbox.ru

А.В. Луппов
кандидат технических наук, доцент кафедры автоматики и телемеханики
Вятский государственный университет
Адрес: 610000, г. Киров, ул. Московская, д. 36
E-mail: dei123@mail.ru

А.В. Малышева

преподаватель кафедры автоматики и телемеханики
Вятский государственный университет
Адрес: 610000, г. Киров, ул. Московская, д. 36
E-mail: annyml@mail.ru

Н.А. Шмакова

старший преподаватель кафедры автоматики и телемеханики
Вятский государственный университет
Адрес: 610000, г. Киров, ул. Московская, д. 36
E-mail: shmakova.natalya@mail.ru

Аннотация

Авторы рассматривают проблематику построения информационных систем учета, управления и анализа деятельности предприятия. Один из наиболее экономически целесообразных путей построения таких систем — интеграция унаследованных, приобретаемых и разрабатываемых компонентов на базе программного обеспечения промежуточного уровня, которое может быть квалифицировано как технологическая платформа интеграции.

Рассматривается формализация принципов построения такого платформенного программного обеспечения. Эти принципы формируются по аналогии с принципом формирования микропрограммного управляющего устройства процессора. В работе показано, как, используя микропрограммное устройство управления процессором в качестве формальной модели, разработать программную платформу, позволяющую организовать запуск процедур обработки, визуализации и обмена данными по расписанию или запросу пользователя. Совокупность информационно-программного обеспечения, играющего роль платформы, декомпозируется на память инструкций, память метакоманд, контроллер памяти инструкций и контроллер памяти метакоманд.

В качестве одного из вариантов платформенного программного обеспечения рассматривается система управления документооборотом. Указано на возможность использования системы электронного документооборота для обмена информацией, защищенной электронной цифровой подписью, а также интеграции информационных подсистемам справочного, учетного и прочих назначений. Такая практика получила название СЭД-шины.

В рамках теоретической постановки задачи рассматривается информационная модель предметной области как совокупность выполняемых в ней бизнес-процессов. Автоматизация деятельности в предметной области заключается в формировании множества автоматизированных бизнес-процессов (мощность которого определяется функциональностью внедряемой информационной системы), максимально покрывающего множество бизнес-процессов предметной области. Рассматриваются случаи совмещения множеств бизнес-процессов унаследованных информационных систем и бизнес-процессов внедряемых информационных систем.

Литература

1. Хоп Г., Вульф Б. Шаблоны интеграции корпоративных приложений. М.: Вильямс, 2007. 672 с.
2. Ланских Ю.В., Богомолова Е.В., Татаринова Е.О. Решение задач системной интеграции при автоматизации транспортного учета в региональном предприятии связи // Материалы международной научной конференции «Регионы в условиях неустойчивого развития» (1—3 ноября 2012 г., Шарья — Кострома): в 2 т. — Т. 2 / Сост. А.М. Базанков, И.Г. Криницын, А.П. Липаев. — Кострома: КГУ им. Н.А. Некрасова, 2012. С. 188—193.
3. Балашов Е.П., Пузанков Д.В. Проектирование информационно-управляющих систем. М.: Радио и связь, 1987. 256 с.
4. Ланских Ю.В. Принципы организации технологической платформы и процедур обработки данных в информационно-аналитической системе // Вестник ИТАРК. 2012. №1. С. 24-28.
5. Ланских Ю.В., Нечаев Д.С. Организация управления в системах интеллектуального анализа данных // Вестник Вятского научного центра Верхне-Волжского отделения АТН РФ. Серия: Проблемы обработки информации. 2004. № 1 (5). С. 144—149.
6. Балашов Е.П., Григорьев В.Л., Петров Г.А. Микро- и мини-ЭВМ. — Л.: Энергоатомиздат. Ленинградское отделение, 1984. 376 с.
7. Богомолова Е.В., Ланских Ю.В. Практические решения по использованию системы электронного документооборота в информационной системе предприятия // Всероссийская ежегодная научно-практическая конференция «Общество, наука, инновации» (НПК-2014). Секция «Вопросы проектирования систем обработки и представления данных». 15—26 апреля 2014 г., Киров: Сб. материалов. Киров, 2014. 1 электрон. опт. диск (CD-ROM).
8. Enterprise-wide information system for construction: A document based approach / M. Arif [et al.] // KSCE Journal of Civil Engineering. 2011. Vol. 15. No. 2. P. 271—280.
9. Богомолова Е.В., Ланских Ю.В. Практические решения по синхронизации справочной информации в корпоративной системе предприятия // Всероссийская ежегодная научно-практическая конференция «Общество, наука, инновации» (НПК-2013). Общеуниверситетская секция, БФ, ГФ, ФЭМ, ФАВТ, ФАМ, ФПМТ, ФСА, ХФ, ЭТФ. 15—26 апреля 2013 г., Киров: Сб. материалов / отв. ред. С.Г. Литвинец. — Киров, 2013. 1 электрон. опт. диск (CD-ROM).

Designing executable business processes as a programming paradigm

Andrey G. Mikheev
Associate Professor, Department of Business Informatics and Industrial Management Systems
National University of Science and Technology MISiS
Address: 4, Leninsky Prospect, Moscow, 119991, Russian Federation
E-mail: andrmikheev@gmail.com

Valeriy E. Pyatetskiy
Professor, Head of Department of Business Informatics and Industrial Management Systems
National University of Science and Technology MISiS
Address: 4, Leninsky Prospect, Moscow, 119991, Russian Federation
E-mail: 7621496@gmail.com

Abstract

This article discusses techniques used to design business processes that are directly executable on the computer system of an enterprise (executable business processes). It also describes the experience of teaching the elements of this technology. This experience was accumulated within two years of teaching process disciplines to bachelors and masters in National University of Science and Technology MISiS and Moscow State University of Economics, Statistics and Informatics (MESI).

One of the reasons to choose the process way of enterprise automation is reducing the cost of automation. In traditional automation, at first the business analyst describes the functionality of the designed system in the form of text, then the programmer translates it into code. The use of executable business processes would make it possible to avoid duplication of work in many ways. In this case the business analyst with the customer uses visual graphic software to develop the business processes of automated functionality which will then be executed directly in the computer environment. Schemes of executable business processes are the human-readable graphical description of the corresponding functionality and it's not necessary to translate them into code. Therefore, the cost of analytical work in this case is the same while the cost of programming is significantly lower. If the business environment changes, the business analyst can quickly change the schemes of business processes accordingly without involving the programmer. In addition, in many cases, the business analyst can independently (without programmer) develop new business processes. Therefore, the cost of development, maintenance and support of such IT-solutions is significantly lower than the cost of traditional solutions, while the speed of development, implementation and subsequent changes is significantly higher.

These advantages (faster, cheaper, easier to maintain and support) are the same advantages the paradigm of object-oriented programming has over the procedural programming paradigm. By analogy, we can call the development of software solutions based on executable business processes a new programming paradigm with respect to the traditional approach.

Process automation based on executable business processes requires process thinking from business analysts that differs from the thinking of IT specialists in the traditional enterprise automation. In addition to knowledge of business process notations, business analysts should be able to implement the typical situations in enterprise business in the form of executable business processes. This article presents the methodology that was used to teach students the process thinking.

Introduction

Notwithstanding the fact that the first analogues of modern business process management systems (at that time they were called workflow-systems [1]) appeared more than fifteen years ago, until recently, most of the process management works (for example, [2 − 4]) were limited by the study of production activity of enterprises, identification of repetitive chains of actions, formalization and integration of these chains into completed business processes, analysis of selected business processes, and development of recommendations for changes in business processes so that at the same time the operating efficiency of the enterprise increased. This did not imply automated execution of business processes. The use of computer systems was limited only to business process modeling. That is, after the development or modification of a business process it was introduced to organizations by administrative methods, which are long, clumsy and expensive.

In recent years, quality changes occurred in this field [5]. Currently, enterprises have been actively implementing computer systems directly executing business processes in the computer environment which are called the business process management system (BPMS). These systems distribute tasks to the executors and monitor their implementation. The sequence of tasks is determined by the business process diagram. Control points move across the diagram; in the design nodes control points generate tasks for the executors.

Thus, at an "office"-type enterprise, an analogue of the production line appears: this mechanism makes it possible to exclude routine operations from the employees' actions, inefficient procedures related to information search and transmission, and significantly to increase the rate of employee interaction. This is due to the fact that by using BPMS, employees accomplish received tasks sticking to receiving data required for task execution from other employees; they transmit the results of their work to other employees; and they study the job descriptions. All information needed to perform the task appears on the employee's computer screen.

At enterprises with stable recurring chains of operations, the use of BPMS provides other advantages:

✦ significantly simplifies the control activities for works in progress and increases the transparency of business operations;

✦ improves the enterprise product quality: through automatic regulatory activity and monitoring tools to observe all rules provided

✦ makes it possible to promptly change business proc-esses in response to the changing business environment of the enterprise;

✦ makes it possible to solve the problem of enterprise-scale integration;

✦ reduces the cost of enterprise automation, and improves the rate of software development and reliability.

Let us explain the last item in the list. Reducing costs is one of the reasons for selecting a process automation option. In the traditional approach, at the beginning a detailed technical project (in the form of a plain text) is drawn up. This is approved by the customer, and subsequently it is converted to program code by software programmers. Automation based executable business processes makes it possible to eliminate duplication: in this case, the analyst immediately develops executable business processes, which are approved by the customer and do not have to be translated into a program code. Therefore, the development time and costs of the work of the executors are significantly reduced.

Automation based on executable business processes makes it possible to quickly adapt the development to changing problems and new ideas started up in the course of development, to reduce development costs, to reduce the technical support costs, and significantly to reduce the cost of modifications and maintenance.

Thus, system implementation, customization and management based on executable business processes prove to be faster and cheaper as compared to traditional automation in which separate application components are developed for various problems and functions. These advantages (faster, cheaper, easier to support and maintain) coincide with the paradigm advantages of object-oriented programming as compared to procedural programming. By analogy, the activity of designing executable business processes can be called a new programming paradigm.

In this case, the concept of the paradigm is considered in terms of R.Floyd's programming paradigm concept [6], which is an extension of T.Kuhn's paradigm concept proposed in the paper [7].

In recent years, executable business processes have been actively implemented both in business and government organizations. However, automation based on executable business processes requires process thinking of specialists different from thinking of IT specialists using the traditional enterprise automation. Apart from the knowledge of notations describing business processes and interfaces, BPMS, business analysts should be able to implement some business specific situations in the form of executable business processes. In this regard, a task appears relating to training specialists for both eco-

nomic and information technology-related specialties in the process approach and BPMS handling.

This article offers rules of outlining business processes, and it provides solutions for some typical situations. These rules can be considered as an extension and refinement of a set of rules set forth in the section 'BPMN Best Practices' [8].

1. Training for executable business processes development

The paradigm of object-oriented programming has led to the emergence of specialists whose way of thinking deeply varies from the traditional thinking of procedure-oriented software programmers. Making comparisons with process automation, it is fair to say that after progressing to a certain stage of the business, rapidly growing business processes-based automation will require a large number of specialists, i.e. business analysts possessing process thinking.

Even today these specialists have to be trained in higher education institutions. By analogy with programming training, teaching students business processes development can be divided into two approaches:

❖ study of business process description notations and training to work with specific BPMS (similar to learning the syntax of programming languages and specific compilers)

❖ study of various possible implementation options in the form of executable business processes of various typical situations in the enterprise business (similar to learning programming techniques)

There is a large number of training courses dedicated to the first approach. For example, papers [9-10] summarize the experience of teaching students to develop executable business processes in MESI, NITU MISiS and UGATU. In the lessons students learn the theory of executable business processes, graphical business process description notations (the most popular notation is BPMN, but sometimes UML AD is used [11]), the main components of typical BPMS, and they acquire practical experience in development and execution of the simplest business processes.

In the course of training, the issues of handling transient business processes, rules of selecting traffic of control points and capabilities of assigning terms of job execution are studied and consolidated in practice. The developed business processes are executed by students under different roles in the software environment.

Training courses dedicated to construction of various process automation solutions based on executable business processes are currently still being established. Within the framework of such courses, on the basis of agreed sets of specially selected business situations and options of process solutions for them, techniques of outlining effective solutions can be taught, and the process thinking of students can be developed. The following section provides examples of situations, and it formulates rules of constructing business process diagrams developed in BPMS operating practice at enterprises.

2. Engineering practice of executable business processes

2.1. Formulations used in the names of business process diagram nodes appropriate to executor actions

The name of a node, which carries a job for the executor, in a majority of BPMS is identical to the name of the job that is displayed to the user. The jobs should be formulated so that they are clear to the extent possible to the executor. From the authors' experience, the clearest are wordings including an infinitive and a noun, such as "issue an order", "review the application". In the lessons conducted by the authors, this kind of naming business process nodes is mandatory.

2.2. Size of business process diagram

It is extremely difficult to analyze business process diagrams having a size larger than one and a half times the size of the computer screen. If the diagram does not fit on the screen, it is necessary to try to move its parts into internal or external sub-processes.

2.3. Motion direction of control points across the business process diagram

It is comfortable to analyze motion of control points across the business process diagram when a general motion of control points corresponds to motion of the area which a person looks at while reading texts. Therefore, it is desirable to place the business process diagram nodes so that the general motion of the control points in them go from left to right or from top to bottom. In case of long sections of control point motion, the diagram nodes connected by junction lines can be arranged from left to right and from top to bottom much as a person reads words on a sheet of paper document (*Figure 1*).

In case of complicated behavior logic of a business process, when a large number of loops of directed transitions appear in the diagram, this cannot be achieved. However, the majority of business processes used in practice have a simple logic, and in their development we need to pay attention to correspondence of the overall motion of control points in the selected direction (from left to right or from top to bottom).
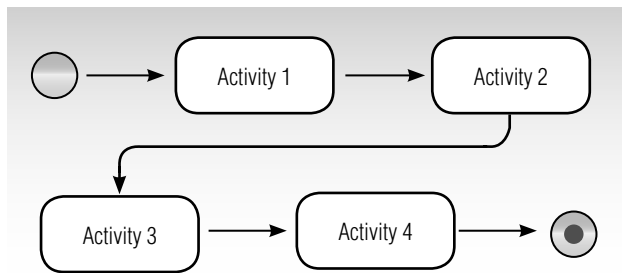
Fig. 1. Diagram of control point motion from left to right and from top to bottom

2.4. Do not use roles in the form of swimlines in business process diagrams

The roles are intended to link the business process diagram nodes with the task executors. The majority of modern graphical notations makes it possible to assign roles in the diagram in the form of horizontal or vertical stripes, called swimlines. In this case, all diagram nodes located in the swimline are associated with a role corresponding to the track.

The practical experience of the authors shows that the use of roles in the form of swimlines is inconvenient in the industrial business processes of the enterprise, inasmuch as the need for placing the business process diagram nodes on a certain strip prevents your developing diagrams that are easy-to-understand in terms of motion of the control points, and it also significantly increases the area occupied by the business process diagram.

Information on a node-related role is important to analyze the business process diagram. Therefore, it is proposed to put the role name in parentheses at the top of the graphic element of the action node and consider it as a prefix of the node name. This technique will be used in the diagrams provided in this article.

2.5. Implementation of actions to be simultaneously performed by two executors

In some cases, the action should be simultaneously performed by two executors (for example, one employee should sign a document which is in the possession of the other employee). As a rule, the intuitive realization of such a scenario corresponds to a sequential arrangement of two nodes in the business process diagram; in so doing, the executor in the first node is an employee who should sign the document, and the second executor is an employee who is in possession the document is (*Figure 2*).
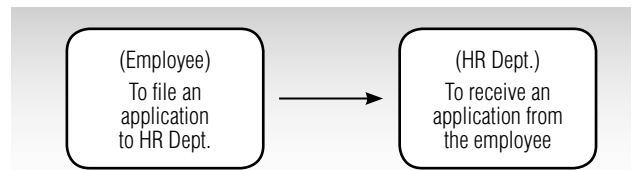
Fig. 2. "Intuitive" implementation of an action performed simultaneously by two persons

However, the practice of business process maintenance shows that such a solution fails, since such node arrangement does not allow for coordinating the interaction. In this case, it is proposed that the nodes be placed in parallel (*Figure 3*).
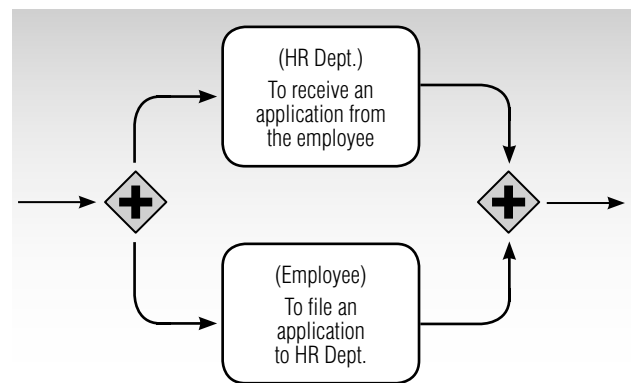
Fig. 3. Proper implementation of an action being performed simultaneously by two persons

2.6. Taking secondary actions to a parallel branch

Let us consider the case when several consecutive actions should be performed simultaneously by two executors. The practice of executable business processes shows that the roles of officials, such as the "accountant" or "cashier" correspond to "responsible" employees, and the role of "employee" or "applicant" corresponds to much lesser "responsible" employees who can forget to mark the job processing for weeks.

Figure 4 provides an example of a business process diagram in which the tasks, for which execution is introduced by an "employee" into BPMS, can disable the normal business process flow. These tasks are marked in the Figure by oval curves. For example, if an employee does not mark a job execution in BPMS "become familiar with the approval", the business process will not proceed to execution of the order and disbursement of money. In the commercial operation mode, such diagrams of business processes can lead to serious disruptions in enterprise performance.
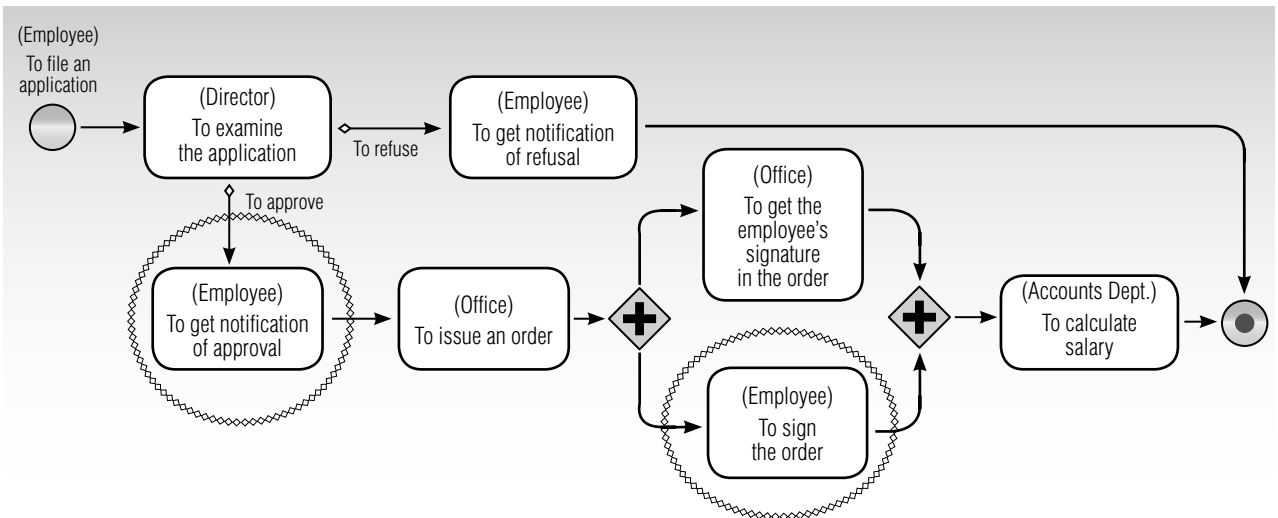
Fig. 4. Improper implementation of the business process with secondary actions
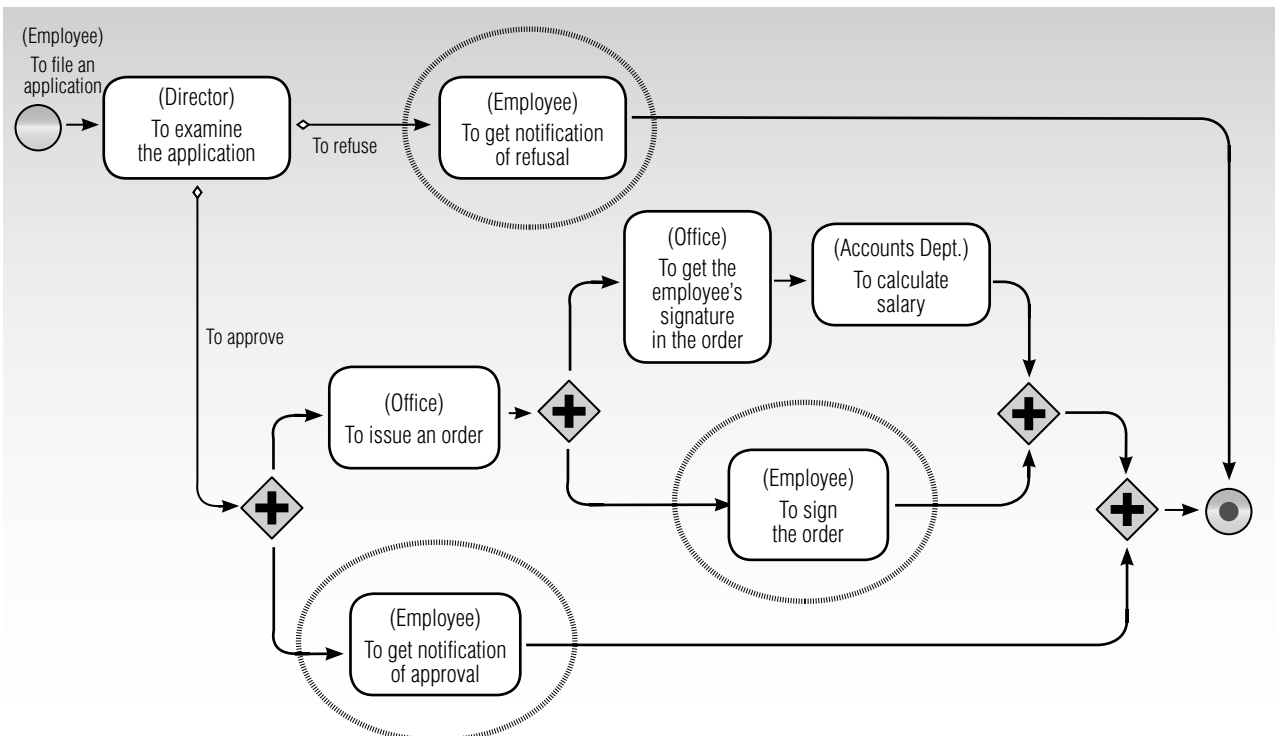
Fig. 5. Correct implementation of a business process with secondary actions

Therefore, the business process diagram is to be drawn out so that the secondary jobs executed by the employee do not suspend further implementation of the business process. Each such job should be performed in a parallel branch, and after it no essential tasks of the business process should be performed. An example of a correct outlining of the business process diagram corresponding to the diagram actions in *Figure 4* is provided in *Figure 5*.

2.7. Using paired splits and merges: Implementation of the possibility of decomposing a diagram section

In complicated cases, split elements without their paired elements – merges (split is what we call a parallel gateway with one incoming and many outgoing transitions, and merge is what we call a parallel gateway with multiple incoming and one outgoing transitions) are
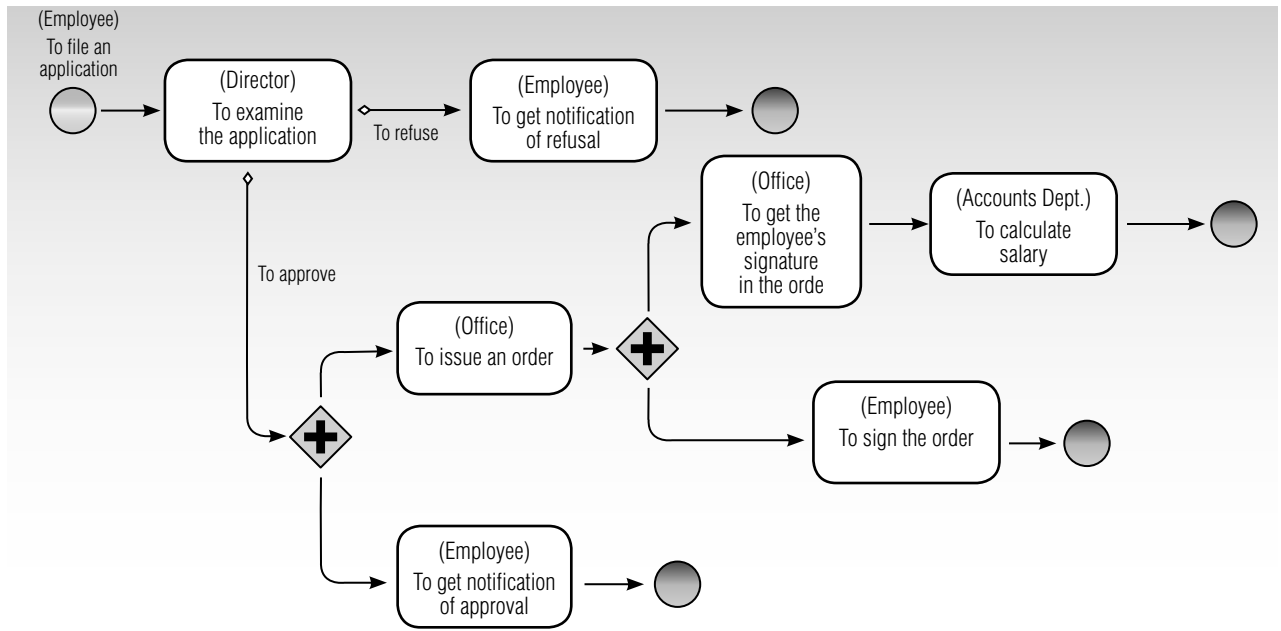
Fig. 6. Alternative implementation of a business process with secondary actions
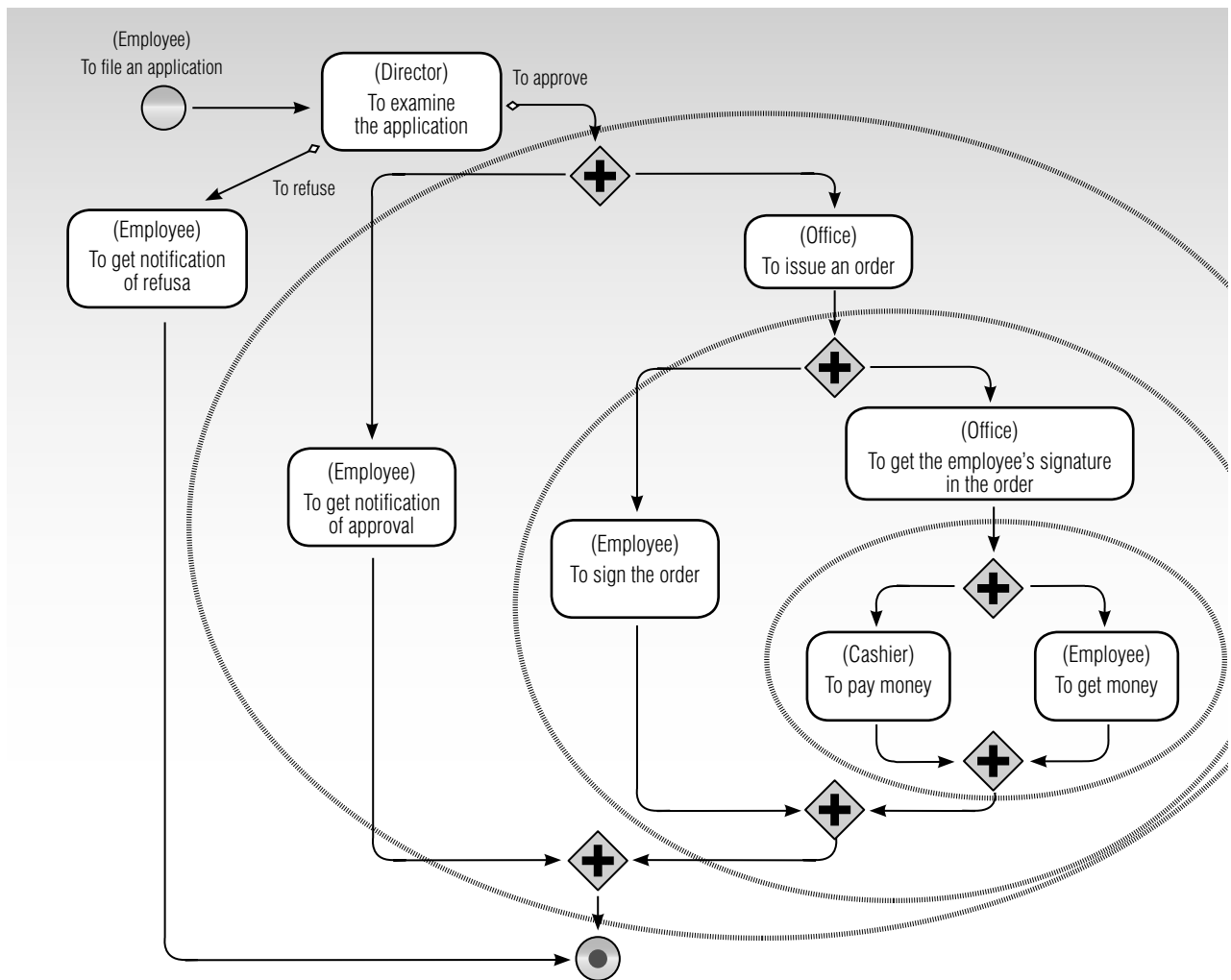
Fig. 7. Example of a business process diagram with three embedded split and merge pairs

sometimes used in the business process diagrams. An element "end of control flow" is used in such diagrams to remove the control points which completed their job (*Figure 6* shows an example of diagram with unpaired elements equivalent to the diagram in *Figure 5*). However, as noted in paper [8], a preferred diagram is a diagram with paired splits and merges, since such diagrams are easier to understand.

This happens because the diagram section between the split and its paired merge can be mentally decomposed and, thus, the business process diagram is split into two simpler diagrams. Having experience, the business analyst can quickly "read" such diagrams. In case of large diagrams with unpaired elements, the business analyst has to "decode" these diagrams; that requires much more time and effort. *Figure 7* shows a diagram with three embedded split and merge pairs. It is evident that the diagram drawn out in this way enables us to mentally decompose it successively three times and, thus, to simplify the complexity of its visual perception.

2.8. Location of paired splits — merges and connecting transitions

It is convenient to locate splits and their paired merges on the same horizontal or vertical line so that a paired element for one element can be easily found in the business process diagram. It is desirable that the transition lines corresponding to simultaneously running action flows be parallel. This makes it easier to understand the diagram, as it is easier for the business analyst to arrange

sequences of actions in the diagram in parallel as "running in parallel." Examples of such arrangements are shown in *Figures 5* and *7*.

2.9. Use of the "end of business process" element

It is preferable to use "end of the business process" elements rather than "end of control flow" elements, because in this case, a business analyst can more easily analyze the chart of a workflow instance process being performed with control points posted thereon. Once a control point arrives at an "end of business process" element, the workflow instance is immediately completed, whereas when an "end of control flow" element is used, a business analyst has to expend more effort to ensure that all control points have come to "end of control flow" elements.

Based on the "end of the business process" element, one can build process solutions for certain situations. Let us consider the case of document concurrence: three departments should agree on the document. Each department may approve or reject the document. If any of the departments rejects the document, the document receives the status "not agreed" and the concurrence should stop immediately, since there is no need for other departments to review the documents.

Concurrence by all departments should be done in parallel. The approval procedure in the business process is not important. *Figure 8* provides a diagram which when used within the sub-process solves the task set.
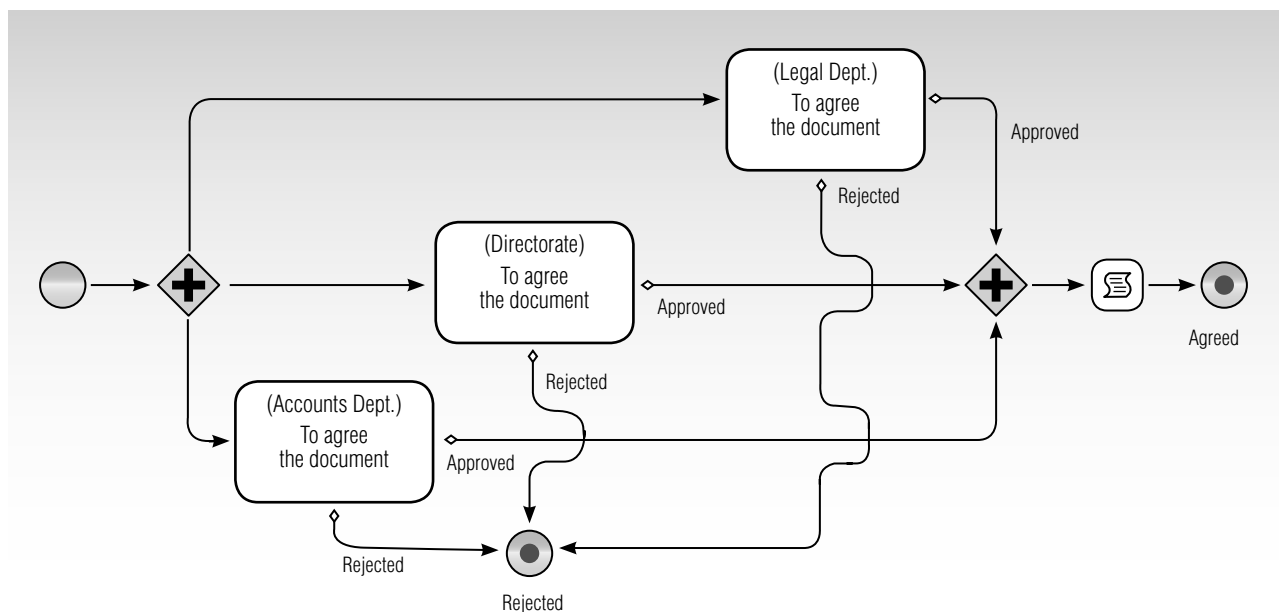
Fig. 8. The business process diagram implementing document concurrence

According to this diagram, any control point which comes to an end-node "Rejected" due to a rejection by any department will stop implementation of the entire sub-process and, in particular, remove the control points from the nodes in which document was concurred with two other departments. In case of a positive decision made by any department, the control point gets into a merge element, in which "it is waiting" for positive solutions from other departments.

2.10. An example of a compromise solution on splits — merges and use of the "end of control flow" element

The development of executable business processes is an art like an art of traditional programming. There are no ready recipes for all possible situations. In many cases, the solution proves to be compromise schemes combining both recommendations proposed in this article, and some exceptions.

As an example, *Figure 9* depicts a simplified business process diagram of a retail lending bank. It contains both paired splits and merges or unpaired splits, and an end node of business process, and two end nodes of the control flow. It goes like this, because if the credit manager

rejects it, there will be another executer who will also have to be informed about the rejection.

2.11. Use of algorithms in the business process diagram

Due to the fact that business process diagrams are very similar to control flow charts, a solution algorithm for a certain problem can be included directly in the business process diagram. This approach can be applied in development of both industrial and educational business processes.

As an example, let us consider a process implementation of the classical M.Gardner's problem of a discerning bride as formulated by academic E.B.Dynkin [12]. Grooms one by one come to one bride; the total number of grooms is known in advance. She keeps company with each of them no more than once and can compare a groom with any of the previous ones. If she selects someone, the selection process is terminated. If the bride rejects someone, she cannot meet him again. The bride's intention is to select the best groom with a maximum probability.

Figures 10 and *11* provide a business process and its sub-processes implementing this task.
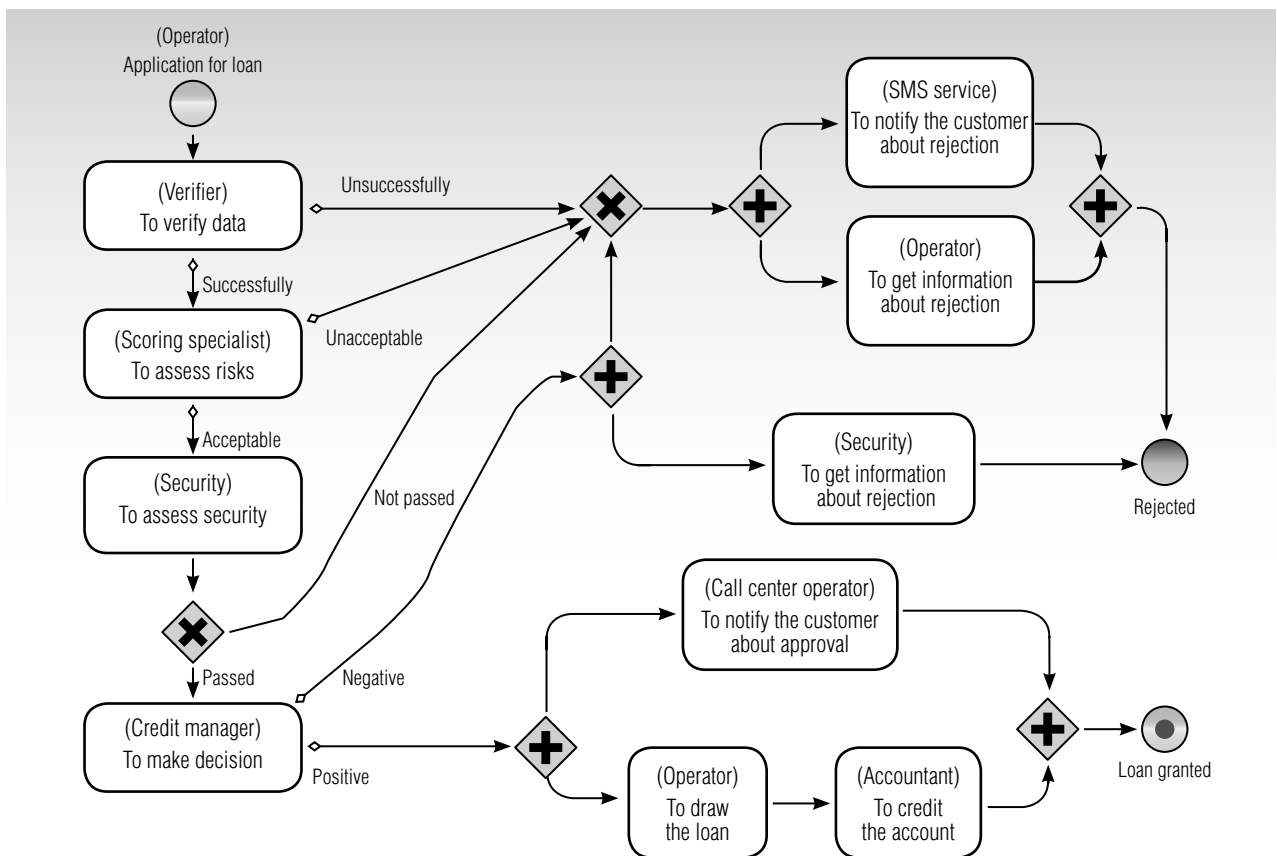
Fig. 9. Business process diagram of bank retail lending
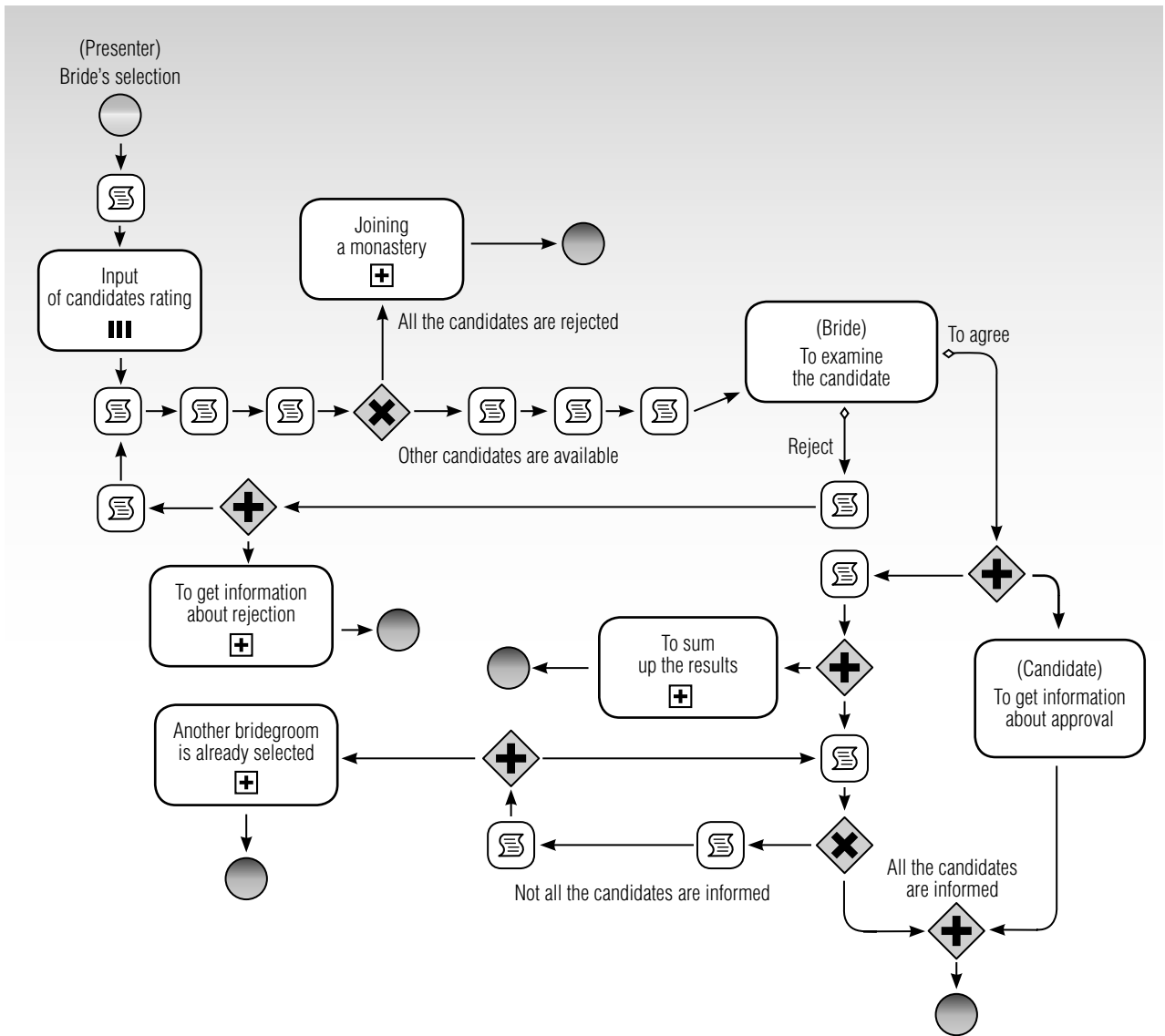
Fig. 10. Business process diagram of the discerning bride problem

In the business process diagram, tasks-scripts (small-size elements) correspond to support operations such as determining the number of marriage candidates in the list of candidates, extracting the current marriage candidate from the list of candidates, replacing the current marriage candidate with the next candidate in the list, etc.

3. Application of free software with an open code to train specialists in process automation

To teach students process automation, courses [9, 10] apply free software − BPMS RunaWFE [13]. The application of the free software for education makes it easy to introduce a training course in the educational process of any Russian university: it is free, available online at the project website RunaWFE (*http://runawfe.org/rus*). No keys or license files are required to install the system. The number of installations is not limited.

Conclusion

This article describes business situations for which solutions are proposed in the form of executable business processes. In addition we offer rules of constructing business process diagrams and an approach for training specialists − business analysts. The cited examples show that the development of executable business processes is a new area of activity for which new methods of building IT solutions and new training procedures should be created. ∎
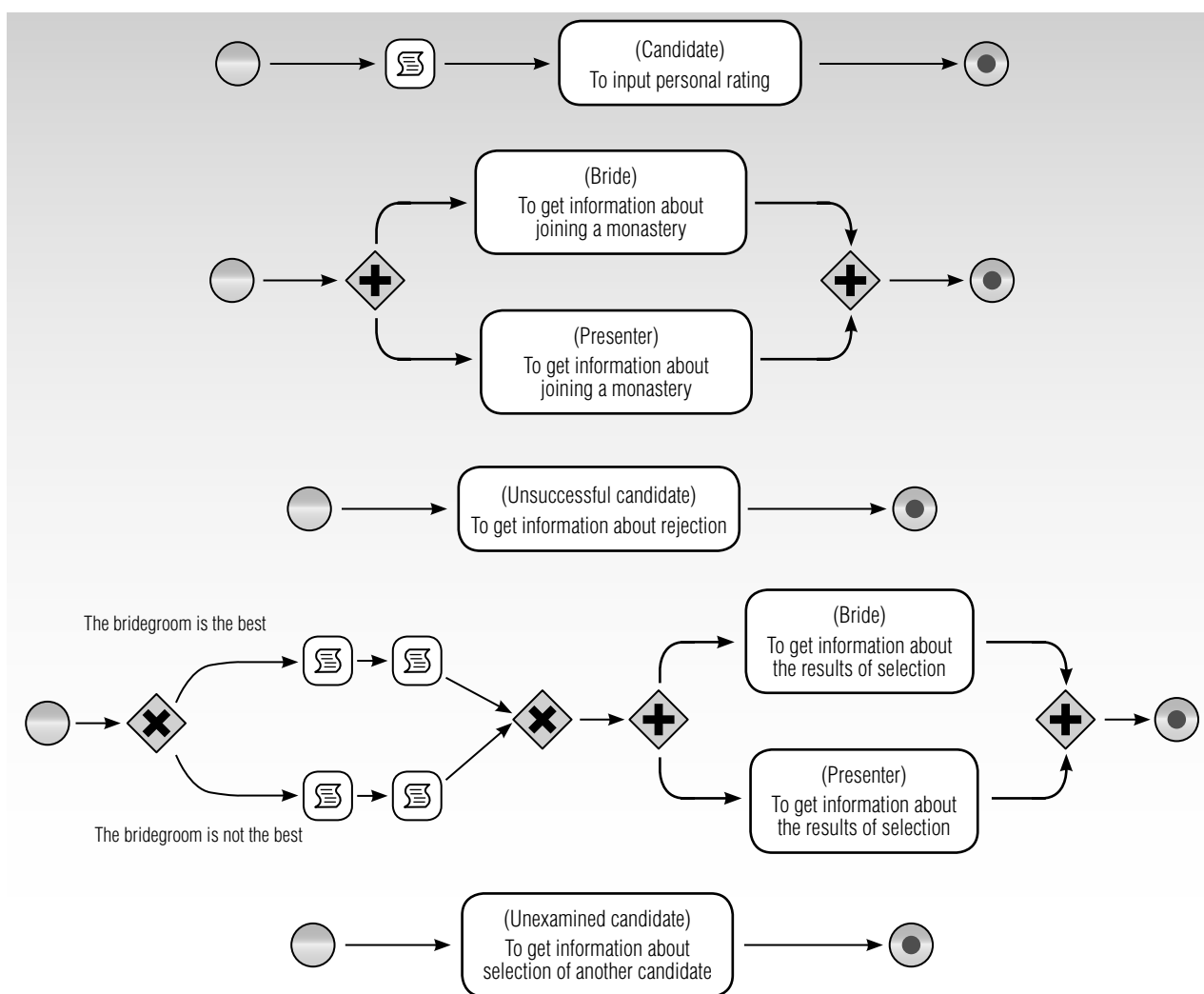
Fig. 11. Sub-process diagrams of the business process of the discerning bride problem

References

1. Van der Aalst W., Van Hee K. (2004) *Workflow management: Models, methods and systems.* Cambridge, MA: MIT Press, 2004.
2. Abdikeev N.M., Danko T.P., Ildemenov S.V., Kiselev A.D. (2005) *Reinzhiniring biznes-protsessov* [Business process reengineering]. Moscow: Eksmo (in Russian).
3. Telnov Yu.F. (2004) *Reinzhiniring biznes-protsessov. Komponentnaya metodologiya* [Business process reengineering. Component methodology]. Moscow: Finance and Statistics (in Russian).
4. Kalyanov G.N. (2006) *Modelirovanie, analiz, reorganizatsiya i avtomatizatsiya biznes-protsessov* [Business process modeling, analysis, reorganization and automation]. Moscow: Finance and Statistics (in Russian).
5. Vagner Yu.B. (2009) BPMS-effect [Effect ot BPMS]. *Automation in Industry*, no. 7, pp. 42−47 (in Russian).
6. Floyd R.W. (1979) The paradigms of programming. *Communications of the ACM*, no. 22 (8), pp. 455−460.
7. Kuhn T.S. (1962) *The structure of scientific revolutions.* Chicago: University of Chicago Press.
8. White S.A., Miers D. (2008) *BPMN modeling and reference guide: Understanding and using BPMN.* Lighthouse Point, FL: Future Strategies, 2008.
9. Kulikov G.G., Mikheev A.G. (2011) Osobennosti realizatsii protsessnogo podhoda I obuchenija upravleniu biznes-protsessami pri pomoshi svobodnogo PO s otkrytym kodom [Particularities of implementation of process approach and business processes management using free software]. *Open Education*, no. 4, pp. 47−57 (in Russian).
10. Pyatetskiy V.E., Mikheev A.G., Novichihin V.V (2013) *Sistema upravlenija biznes-protsessami: osnovy razrabotki biznes-protsessov pri pomoshi svobodnogo programmnogo obespechenija* [Business process management system: The basics of business process design using free software]. Moscow: MISiS (in Russian).
11. Fiodorov I.G. (2011) Sravnitelnyj analiz notatsij modelirovanija biznes-protsesov [Comparative analysis of business process modeling notations]. *Open Systems*, no. 8, pp. 28−32 (in Russian).
12. Gusejn-Zade S.M. (2003) *Razborchivaja nevesta* [The smart bride]. Moscow: MCCME (in Russian).
13. Mikheev A.G. Orlov M.V. (2011) Sistema upravlenija bisnes-protsessami i administrativnymi reglamentami [Business process and administrative regulation system]. *Programmnye Produkty i Sistemy*, no. 3, pp. 126−130 (in Russian).

Проектирование исполняемых бизнес-процессов как парадигма программирования

А.Г. Михеев

кандидат физико-математических наук,
доцент кафедры бизнес-информатики и систем управления производством
Национальный исследовательский технологический университет «МИСиС»
Адрес: 119991, г. Москва, Ленинский проспект, д. 4
E-mail: andrmikheev@gmail.com

В.Е. Пятецкий

доктор технических наук, профессор,
заведующий кафедрой бизнес-информатики и систем управления производством
Национальный исследовательский технологический университет «МИСиС»
Адрес: 119991, г. Москва, Ленинский проспект, д. 4
E-mail: 7621496@gmail.com

Аннотация

В статье рассматриваются приемы, применяемые при разработке бизнес-процессов, непосредственно исполняемых в компьютерной системе предприятия (исполняемых бизнес-процессов). Представлен опыт обучения элементам этой технологии, полученный в Национальном исследовательском технологическом университете МИСиС и Московском государственном университете экономики, статистики и информатики (МЭСИ) в течение двух лет преподавания процессных дисциплин в бакалавриате и магистратуре.

Одной из причин выбора процессного варианта автоматизации предприятия является уменьшение затрат на автоматизацию. При традиционной автоматизации сначала бизнес-аналитик описывает функциональность проектируемой системы в виде текста, а затем программист переводит это описание в программный код. Использование исполняемых бизнес-процессов позволяет в значительной степени избежать дублирования работы: в этом случае бизнес-аналитик совместно с заказчиком при помощи визуальных графических программных средств разрабатывает бизнес-процессы автоматизируемой функциональности, которые будут потом непосредственно исполняться в компьютерной среде. Схемы исполняемых бизнес-процессов представляют собой понимаемое человеком графическое описание соответствующей функциональности, при этом их не требуется переводить в программный код. Поэтому затраты на аналитическую деятельность в этом случае будут примерно такими же, а затраты на программирование — существенно ниже. При изменении условий бизнеса бизнес-аналитик может быстро изменить соответствующим образом схемы бизнес-процессов без участия программиста. Также во многих случаях бизнес-аналитик может самостоятельно (без участия программиста) разрабатывать новые бизнес-процессы. Поэтому стоимость разработки, сопровождения и поддержки такого ИТ-решения оказывается существенно меньше стоимости традиционного решения, а скорость разработки, внедрения, а также последующих изменений — существенно выше.

Эти преимущества (быстрее, дешевле, легче в поддержке и сопровождении) совпадают с преимуществами парадигмы объектно-ориентированного программирования по сравнению с парадигмой процедурного программирования. По аналогии мы можем назвать разработку программного решения на основе исполняемых бизнес-процессов новой парадигмой программирования относительно традиционного подхода.

Процессная автоматизация на основе исполняемых бизнес-процессов требует от специалистов — бизнес-аналитиков «процессного» мышления, отличающегося от мышления ИТ-специалистов, занимающихся традиционной автоматизацией предприятий. Кроме знания процессных нотаций, бизнес-аналитики должны уметь реализовать в виде исполняемых бизнес-процессов типичные ситуации, возникающие в бизнесе предприятия. В статье приведены методики, используемые при обучении студентов «процессному» мышлению.

Литература

1. Van der Aalst W., Van Hee K. Workflow management: Models, methods and systems. Cambridge, MA: MIT Press, 2004. 384 p.
2. Абдикеев Н.М., Данько Т.П., Ильдеменов С.В., Киселев А.Д. Реинжиниринг бизнес-процессов. М.: Эксмо, 2005. 592 с.
3. Тельнов Ю.Ф. Реинжиниринг бизнес-процессов: Компонентная методология. М.: Финансы и статистика, 2004. 320 с.
4. Калянов Г.Н. Моделирование, анализ, реорганизация и автоматизация бизнес-процессов. М.: Финансы и статистика, 2006. 240 с.
5. Вагнер Ю.Б. BPMS-эффект // Автоматизация в промышленности. 2009. № 7. С. 42−47.
6. Floyd R.W. The paradigms of programming // Communications of the ACM. 1979. No. 22 (8). P. 455−460.
7. Kuhn T.S. The structure of scientific revolutions. Chicago: University of Chicago Press, 1962. 172 p.
8. White S.A., Miers D. BPMN modeling and reference guide: Understanding and using BPMN. Lighthouse Point, FL: Future Strategies, 2008. 231 p.
9. Куликов Г.Г., Михеев А.Г. Особенности реализации процессного подхода и обучения управлению бизнес-процессами при помощи свободного ПО с открытым кодом // Открытое образование. 2011. № 4. С. 47−57.
10. Пятецкий В.Е, Михеев А.Г., Новичихин В.В. Система управления бизнес-процессами: основы разработки бизнес-процессов с помощью свободного программного обеспечения: Практикум. М.: МИСиС, 2013. 207 с.
11. Федоров И.Г. Сравнительный анализ нотаций моделирования бизнес-процессов // Открытые системы. 2011. № 8. С. 28−32.
12. Гусейн-Заде С.М. Разборчивая невеста. М.: МЦНМО, 2003. 24 с.
13. Михеев А.Г., Орлов М.В. Система управления бизнес-процессами и административными регламентами // Программные продукты и системы. 2011. № 3. С. 126−130.

Development of the information and communication sector as a factor in the evolution of fair vehicle insurance: From concept to determination of rates

Vasiliy N. Shcherbakov
MSc Program Student, Department of Data Processing Automation
Tomsk State University of Control Systems and Radioelectronics
Address: 40, Lenina Prospect, Tomsk, 634050, Russian Federation
E-mail: sherbakovasiliy@gmail.com

Anatoly A. Sidorov
Associate Professor, Department of Data Processing Automation
Tomsk State University of Control Systems and Radioelectronics
Address: 40, Lenina Prospect, Tomsk, 634050, Russian Federation
E-mail: astroasregion@gmail.com

Abstract

The paper describes the features of auto insurance services in Russia, including the factors affecting the cost of insurance products. A review of existing legislation in the field of car insurance is presented: according to the Law "On the State Information System ERA-GLONASS, manufacturers of categories B and C vehicles will be required to implement this tracking system for all cars produced in Russia. This fact was a prerequisite for research aimed at exploring the possibility of adapting the foreign system of "smart insurance" to the Russian market. Modern processes of car insurance and related technologies for determining the insurance policy value used in such countries as the US, UK and Australia rely on telematics sensors that determine individual performance and driving style. Some of components of the ERA-GLONASS system may be applied as such sensors.

In this paper, we present different models of participant interaction during implementation of a new technology for the process of delivering CASCO insurance products in Russia. Such models include the public-private model, independent-private model, as well as industry and outsourcing models. We identify the main mechanisms for improving the competitiveness of insurance companies through general introduction of a "smart insurance" approach: reducing the policy rates in case of safe driving, and expanding the functionality of IT services for drivers. A survey of vehicle owners and interpretation of the results are presented: the purpose of the survey was to identify readiness of drivers for transition to the proposed model of car insurance, and identification of functionality of appropriate mobile services.

Relying on an analysis of the methodology for calculating policy rates for risky types of insurance approved by order of Rosstrakhnadzor, a model of insurance rate calculation is proposed using special discounts applied in recognition of careful driving.

Introduction

The development of the vehicle insurance market in foreign countries demonstrates that the factors affecting the tariff formation of the CASCO insurance policies in Russia are not in correlation with the frequency of occurrence of insured events. The USA, United Kingdom and Australia are rapidly moving to an improved system of determining the insurance value which takes into account individual data on motor vehicle driving quality of the given driver through installation of telematic sensors on the motor vehicle.

The technology under consideration has been applicable in some foreign countries since 2010. It makes it possible to quantify the client's driving habits (speed, braking power, observance of traffic rules, etc.). As a result, due to the potentially low hazard he presents by virtue of his conduct, a careful driver pays much less for insurance coverage [1]. Such a method is called "smart insurance" or "pay-as-you-drive". Therefore, the merit index of driving lays the foundation for determining the value of insurance services.

It seems obvious that such an approach is of interest both for insurance companies and drivers [2]. "Smart insurance" enables the insurance companies to select their customers, efficiently manage insured events and improve road safety. For the customer, "smart insurance" is, on the one hand, a discount for proper driving, and on the other hand, gives real-time ability to monitor the vehicle, as well as prevent critical driving errors through the use of mobile services [3].

The USA and Great Britain have been using this approach for nearly five years. For example, Snapshot developments have been introduced in 43 states (which points to large-scale technology penetration). The maximum discount for the customers who have used the services under consideration was 30%, and additional corporate revenue was more than 1 million dollars. In Illinois, Ohio and Arizona the principle is being used by Allstate Drive Wise. In Great Britain approximately 100,000 devices are activated, and at present there is a monthly influx of 6,000 new customers. Planned total sales revenue of "smart insurance" services in the USA and Great Britain in 2015 is estimated at 2 billion 600 million dollars [4, 5].

However, the introduction of "smart insurance" in Russia is constrained by a number of factors of organizational, legal, methodological and psychological nature. They are the subject matter of this paper.

Table 1.

Factors having an impact on the policy value

s.n.	Factor	Motor insurance product			Remark
		OSAGO	DSAGO	KASKO	
1.	Type of vehicle	+	+	-	Category of vehicle is taken into account
2.	Year of manufacture of vehicle	-	-	+	
3.	Cost of vehicle	-	-	+	
4.	Make and model of vehicle	-	-	+	
5.	Method of buying the vehicle	-	-	+	Credit or own funds
6.	Presence of anti-theft alarm system	-	-	+	
7.	Type of amount of coverage	-	-	+	Aggregate or non-aggregate insurance coverage
8.	Intended use of vehicle	+	+	-	Private use, taxi-driving and others
9.	Insurance holder	+	+	-	Legal entity or private individual
10.	Place of vehicle registration	+	+	-	
11.	Record of persons approved to drive the vehicle	+	+	-	Presence / absence of a fixed record of persons approved to drive the vehicle
12.	Age and length of driving experience of persons approved to drive the vehicle	+	+	+	
13.	Power of vehicle's engine	+	+	-	For trucks a Gross Vehicle Weight Rating is accounted, for buses a number of seats is accounted.
14.	Seasonality of vehicle use	+	+	-	
15.	Insured history	+	+	-	Number of previous years of insurance and occurrence of vehicle accidents due to the fault of the insured person.
16.	Driving habits	-	-	-	

1. Specifics of the Russian motor insurance market

The rapid development of the modern motor insurance market in the Russian Federation started in 2002 with adoption of Federal Law No.40-FZ dated 25.04.2002 "On compulsory public indemnity insurance of vehicle owners" [6]. The current regulations oblige drivers to take third party liability insurance. In addition, issues relating to providing additional services – DSAGO and CASCO are predetermined.

Policy value depends on various factors. Table 1 provides a comparison of three major vehicle insurance products from the perspective of conditions taken into account when setting rates.

As the table shows, when setting the insurance rate for OSAGO and DSAGO the type of vehicle, its intended use, the age and driving experience of all its drivers are mainly taken into account, and when setting the CASCO insurance rate, the characteristics and equipping of the vehicle, age and driving experience of a driver are mainly taken into consideration. It should also be noted that driving habits are not taken into account for any of the products under consideration in the cost calculation.

2. Prerequisites for including driving habits as a factor determining the policy value and model for managing interaction of participants

The State Duma of the Russian Federation passed the law No. 395-FZ of 12.28.2013 "About the National Automated Information System ERA-GLONASS" [7], which obliges manufacturers of vehicles in categories B and C to equip vehicles produced in Russia with the tracking system ERA-GLONASS. In the first phase, the equipment will only act as an emergency button. Beginning in 2017, the system shall start to transmit information on accidents in automatic mode. In this case, if the system detects a road traffic accident, an accident signal is automatically sent to the dispatcher, and he, in his turn, communicates with the driver via a speaker phone.

If no response from people in the car ensues, a medical emergency team will be sent to the scene of the accident. Authors of the ERA-GLONASS system maintain that after the system is fully operational the rate of mortality on roads should be reduced by 30—40%. In addition, according to the document, as from 2017 installation of the emergency response system will be mandatory on passenger and freight transport [8].

The leading insurance companies have put forward an initiative to get access to data coming from the system sensors via a single server to determine the rate of vehicle insurance based on individual telematic indicators of driving style. In this case, in each region data centers will be established, which will form, store and process huge data arrays from all sensors. The establishment of such organizations pursues such objectives as improving road safety, increasing traffic capacity of highways, managing traffic signs and traffic lights and setting coordinated traffic lights.

It is expected that these data centers will be government institutions. However, it is unknown how relevant information is going to be used, and who will have access to it. This does not make it possible to determine the final list of participants of relations under consideration (the state, insurance companies, customers), the scheme of their interaction and, as a consequence, the institutional design of this system and the cost of providing motor insurance services in a new format. In this regard, there can be several models of participants' interaction when implementing the new technology of providing services.

The public-private model is based on the fact that stand-alone subdivisions of the federal data center ERA-GLONASS will be organized in the territorial subdivisions (subjects) of the Russian Federation. They will collect and process the vehicle sensor information which can then be accessed by insurance companies (*Figure 1*).

The autonomous-private model is based on the assumption that access to ERA-GLONASS information data-centers will be closed, and the intentions of the insurance companies to transfer the process of provid-
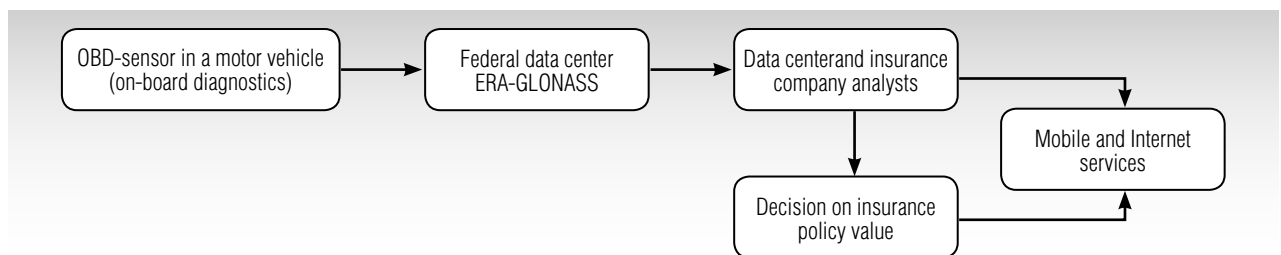
Fig. 1. Public-private interaction model
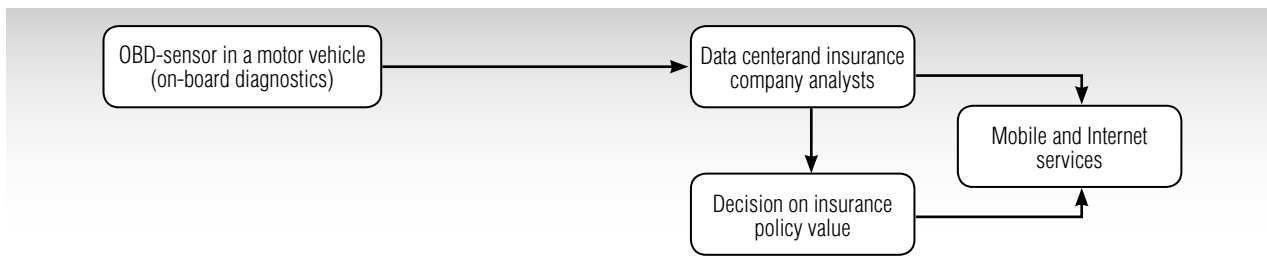
Fig. 2. Autonomous private interaction model
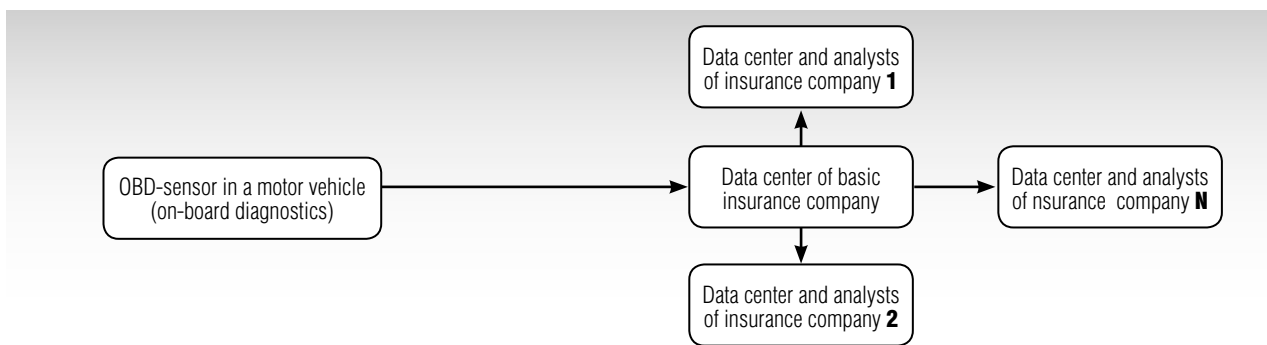
Fig. 3. Industrial interaction model

ing services to a new format will remain unchanged. The state shall be excluded from the proposed model. Each insurance company will need to revise its approaches to providing services: develop or acquire their own sensors (similar to those used in the ERA-GLONASS project), install them in customers' motor vehicles, and then collect and process data by the efforts of their engineering, information and analytical services (*Figure 2*).

The industry model is based on the idea that the insurance company, which first will be able to develop and implement a new approach to the delivery of motor insurance services without the state participation can make their data center common for several (or most) insurance companies, and act as a provider of data from each vehicle (*Figure 3*).

In the outsourcing model, the "intelligent" approach can be implemented by an organization outside the in-

surance industry that offers OBD-sensors, installation services, information acquisition and transmission to the analytic insurance centers of insurance companies for further use (*Figure 4*).

The public-private model will be possible only if the insurance companies get access to information of federal centers of ERA-GLONASS. Otherwise the companies will have to independently develop their own telematic system, install their own sensors in the customers' motor vehicles and establish their data center (in the case of the autonomous private model) or use data centers of other insurers (industry model). Another option that is being considered is outsourcing services of an organization external to the insurance industry.

It should be noted that currently there is no agreed position of the Government of the Russian Federation and the Russian Association of Motor Insurers concerning
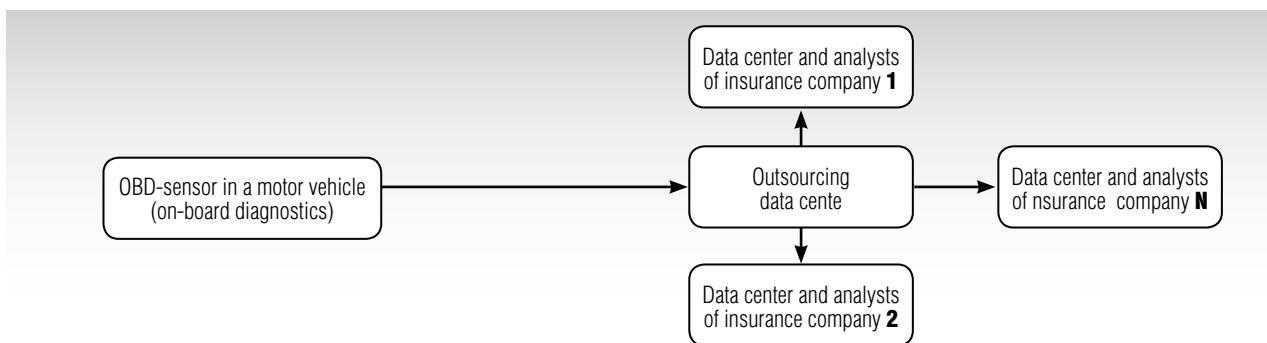
Fig. 4. Outsourcing interaction model

the applicability of data received from the ERA-GLO-NASS system to the insurance companies or the level of their access. This imposes restrictions on a clear choice of institutional model. Only after the interaction format is determined will it be possible to study the practical implementation and costs for implementing and maintaining the "smart insurance".

3. Consumer expectations from "smart insurance"

As already noted, "smart insurance" allows the customer to receive a discount for safe driving, to monitor the vehicle by means of mobile services in real time and to prevent driving errors. During a widespread introduction of the insurance model by Russian insurance companies using telematic sensors, a new strategy should be developed for competitive recovery to attract more customers. Two actuators become apparent: a more attractive discount rate (decrease of CASCO rate value) for careful and safe driving, or an expansion of IT-services functionality for drivers.

The OBD-connectors (telematic sensors) are designed to operate not only on unidirectional basis, transferring data to the data center of the insurance company. Information being transmitted can be used by the drivers, thereby correcting their behavioral driving habits to improve carefulness and safety all of which then influences the amount of future discounts. The information can be tracked and driving habits can be corrected via the mobile applications interface. In so doing, the functionality of a mobile service, as well as the content of the insurance product will not contradict consumer expectations. To identify these expectation, a survey of potential users was conducted. They were selected through a target sampling (400 respondents). The study results were as follows:

1. More than 60% of respondents know that the cost of full comprehensive insurance service in Russia is very high.

2. About 40% of respondents have for some time used or are currently using CASCO services.

3. The vast majority of respondents (about 90%) believe that CASCO service is expensive and inefficient.

4. The functionality of the mobile application synchronized with OBD-connectors to increase driver demand should include the following services (in declining order of priority):

✦ navigation;

✦ autologger;

✦ online ransom payments;

✦ monitoring fuel use;

✦ car service center navigation;

✦ handbook of driving regulations;

✦ on-line recording for vehicle maintenance and washing;

✦ vehicle structure and handbook to calculate repair costs;

✦ search for the vehicle in a parking lot;

5. Over 60% of respondents prefer a single multi-function application.

In addition, the respondents were asked to make a five-grade-scale assessment of the effect of individual factors on the probability of an insured event occurring, and consequently, the insurance value (*Table 2*).

Table 2.

Assessment of the level of factors in the probability of an insured event occurring, as a percentage of the number of respondents

Factor	Assessment				
	1	2	3	4	5
Age of driver	18	14	29	18	19
Driving experience	5	6	15	20	50
Year of manufacture of the vehicle	38	23	18	6	11
Driving habits	10	15	16	21	34
Carefulness	5	14	23	25	31

As the table shows, for calculation of insurance rates the majority of drivers put emphasis on such factors as driving experience, skill and carefulness. At the same time the driver's age and year of the vehicle production are classified as minor items in terms of the probability of an insured event occurring.

4. Method of calculating a discounted insurance rate

Today every Russian insurance company calculates its rates on a stand-alone basis, but as mentioned above, the product value for a consumer depends on the overall indexes defined by the values of parameters and characteristics of the vehicle and the driver. Their functional impact on the final cost and weighting is determined by the insurance company on a stand-alone basis. For a client or a casual observer such a model seems nothing more than a "black box" model (*Figure 5*).
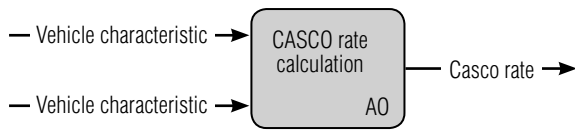
Fig. 5. The "black box" model: a traditional approach
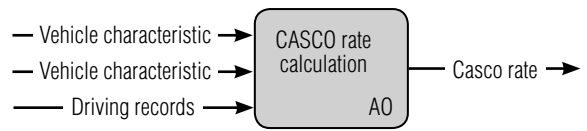
Fig. 6. "Black box" model: "intelligent" approach

Insurance companies which have implemented the "pay-as-you-drive" approach add another set of input characteristics to the final rate calculation − driving records (*Figure 6*).

The core issue in CASCO value is the rate, which essentially represents the insured risk price and other costs, and serves as an appropriate expression of the monetary value of the insurer's liabilities under the insurance contract that has been concluded. Insurance rates are determined by actuarial expectations − a system of mathematical and statistical regularities governing the relationship between the insurer and the insured. They reflect a mechanism for forming and spending from the insurance fund in long-term insurance operations related to the length of life of the population in the form of mathematical formulas. In a broad interpretation, the actuarial expectations include tariffs on any type of insurance. With their use, the participation share of each insurer in creation of an insurance fund is determined [9]. The rate at which an insurance contract is concluded bears the name of a gross premium. It consists of two parts: a net premium, expressing the price of the insured risk, and administrative overheads, covering the insurer expenses for organizing and carrying an insurance case, including contributions to reserve funds and profit elements.

Among the risky types of insurance are those which do not provide for insurer liabilities to pay the insured amount upon completion of the contract term (for example, life insurance) and which are not related to accumulation of the amount of coverage during the contract term. The method of calculating rates on risk-related types of insurance may be applied when statistics or other information is available. That makes it possible to calculate the probability of an insured event occurring, the amount of coverage and insurance benefit (amount of indemnity):

$$T_n = T_o + T_p, \qquad (1)$$

where T_n is a tarrif net premium;

T_o is a key rate;

T_p is a risk premium.

A lot of papers are dedicated to calculation procedures of each element of the net premium, for example, de-

tailed algorithms are presented in paper [10]. It is obvious that a discount provided to the driver for careful driving should impact only the key rate T_o. In this case, the risk premium T_p remains unchanged (see calculation procedures for rates on risk-related types of insurance approved by the Insurance Supervision Service decree of 08.07.1993 No. 02-03-36 [11]), i.e., its value for the purpose of this paper can be taken as a constant.

$$T_o = \frac{q \cdot S_b \cdot 100}{S}$$

where q is an estimate of probability of the insured event defined as a ratio of a number of insured events to the number of contracts for a certain period of time;

S_b is an amount of insurance indemnification in insured events;

S is the amount of coverage in the contracts.

In determining the discount rate, the following model is proposed to be used. Let us assume that it is affected by such telematic measured parameters as driving speed, hard braking, and frequency of lane changes. In calculations, it is reasonable to consider not the values of these parameters, but a ratio of the critical measurements (beyond the "red border") m_i to a total number of measurements n_i: $q_i = m_i / n_i$. Then the careful driving factor q can be calculated as a deviation from 1 amount q_i, each several value being adjusted for weight v_i:

$$q = 1 - \sum_{i=1}^{4} q_i \cdot v_i \qquad (2)$$

Depending on the obtained value q, it is reasonable to determine the discount rate k_c. An example of such a translation is shown in *Table 3*.

Table 3.

Correlation between the careful driving factor and the discount rate

q	<0.7	0.7–0.8	0.8–0.9	0.9–1
k_c	1.0	0.9	0.85	0.8
Discount rate, %	0	10	15	20

After updating, it is suggested that the net premium value determined by formula (1) be calculated as follows:

$$T_n = T_o \cdot k_c + T_p \qquad (3)$$

It is reasonable to consider two examples of insurance rate calculation in the case of careful and aggressive driving as an illustration. The initial data (conditional) is shown in *Table 4*. For convenience, let us take a number of telematic measurements for each parameter equal to 100.

Table 4.

Initial data for net-rate calculation

s.n.	Telematically measured parameter	Careful driving		Aggressive driving	
		Number of measurements 'beyond the red line'	q_i	Number of measurements 'beyond the red line'	q_i
1	Driving speed	10	0,10	60	0,60
2	Acceleration speed	20	0,20	75	0,75
3	Hard braking	5	0,05	55	0,55
4	Lane change frequency	15	0,15	70	0,70

To simplify the calculations, let us take telematically defined parameters equivalent in calculations of the discount value. Then the careful driving factor is calculated by formula (2) as follows:

1) in case of careful driving:

$$q^{crf} = 1 - (\frac{0,10 + 0,20 + 0,05 + 0,15}{4}) = 0,875;$$

2) in case of aggressive driving: .

$$q^{agr} = 1 - (\frac{0,60 + 0,75 + 0,55 + 0,70}{4}) = 0,350.$$

Using rules of translating q to k_c at $T_o = 100$ and $T_p = 10$, we get the following results by formula (3):

1) in case of careful driving: $T_n^{crf} = 100 \cdot 0,85 + 10 = 95$;

2) in case of aggressive driving: $T_n^{agr} = 100 \cdot 1,00 + 10 = 110$.

As a result, in the first case the total discount was 13.6%, and in the second case the discount is not given. Projecting the calculation results to actual CASCO rates, we obtain a substantial economy for the consumer when purchasing the insurance product.

Conclusion

This paper addresses basic aspects of insurance of motor vehicles in Russia, and it provides examples of a modern approach of foreign companies to calculating policy value. Modern foreign companies calculate the insurance rates on the basis of telematic data of the driver defining such individual indicators as acceleration and driving speed, hard braking and lane change frequency. Currently these indicators are not taken into account by domestic companies when calculating the insurance rate. According to interpretation of the data obtained, the majority of drivers consider that individual road skills and carefulness should be important factors in calculating the insurance rate. In addition, on the basis of the poll results, one can determine the required functionality of mobile services, by means of which a driver can individually adjust his driving habits, thereby having an impact on the size of a possible discount when concluding the motor insurance contract.

Global track records in vehicle insurance and passage of the Federal Law No.3 95-FZ of 12.28.2013 "On the National Automated System ERA-GLONASS" predetermined the emergence of several interaction models of parties concerned in the implementation of a new motor insurance technology: the public-private model, the autonomous-private model, the industrial and the outsourcing model.

At present, several domestic insurance companies are proposing that their clients install a telematic sensor in their vehicles and, in the case of careful driving, receive a discount on the insurance product. The discount offered is fixed. This paper describes the method of calculating a discounted insurance rate based on the calculation procedures approved by decree of Russian Insurance Supervision Service No. 02-03-36 of 08.07.1993. ∎

References

1. Towers Watson & Co. (2014) Telematics and usage-based insurance. *Towers Watson & Co.* Available at: http://www.towerswatson.com/en/Services/Services/telematics (accessed 01.03.2015).

2. Morrison R. Usage-based insurance (UBI). Environmental and economic benefits and effects of state laws and intellectual property claims. *True Mileage*. Available at: http://www.lewis.ucla.edu/wp-content/uploads/sites/2/2014/03/140303-UBI-UCLA-Lewis-Center-March-20-2014.pdf (accessed 01.03.2015).

3. Golia N. Usage-based insurance: 5 reasons this is the year. *DriveFactor, Inc.* Available at: http://www.drivefactor.com/usage-based-insurance-5-reasons-this-is-the-year/ (accessed 01.03.2015).

4. Insurance guru (2015) Umnoe KASKO [Smart CASKO]. *Insurance guru* (electronic journal). Available at: http://www.inguru.ru/kalkulyator_kasko/stat_umnoe_strahovanie (accessed 01.03.2015) (in Russian).

5. Harbage R. What makes a UBI score predictive enough? *Towers Watson & Co.* Available at: http://www.towerswatson.com/en/Insights/Newsletters/Americas/americas-insights/2014/what-makes-a-ubi-score-predictive-enough (accessed 15.03.2015).
6. *Ob objazatel'nom strahovanii grazhdanskoj otvetstvennosti vladel'cev transportnyh sredstv* [On compulsory insurance of civil liability of owners of vehicles]. RF Federal Law No. 40-FZ, 25.04.2002 (in Russian).
7. *O gosudarstvennoi avtomatizirovannoi informacionnoi sisteme ERA-GLONASS* [On the state automated information system ERA-GLO-NASS]. RF Federal Law No. 395-FZ, 28.12.2013 (in Russian).
8. *Zakon ob osnashhenii legkovyh avtomobilej sistemoj ERA-GLONASS proshel vtoroe chtenie v Dume* [Act equipping cars system ERA-GLO-NASS passed second reading in the Duma]. Motor Page (electronic journal). Available at: http://www.motorpage.ru/magazine/news/zakon_ob_osnashhenii_legkovih_avtomobilej_sistemoj__jera_glonass__proshel_vtoroe_chtenie_v_dume.html (accessed 01.03.2015) (in Russian).
9. Denisova I.P. (2011) *Strahovanie. Uchebnoe posobie* [Insurance. Textbook]. Moscow: MarT (in Russian).
10. Samarov E.K. (2007) *Strahovaja matematika v primerah i zadachah. Uchebnoe posobie* [Insurance math examples and problems. Textbook]. Moscow: Alfa (in Russian).
11. *Metodiki rascheta tarifnyh stavok po riskovym vidam strahovanija* [Methods of calculation of tariff rates on risky types of insurance]. Approved by Rostekhnadzor, Act No. 02-03-36, 08.07.1993 (in Russian).

Развитие информационно-коммуникационного сектора как фактор развития справедливого автострахования: От замысла к тарифу

В.Н. Щербаков
магистрант кафедры автоматизации обработки информации
Томский государственный университет систем управления и радиоэлектроники
Адрес: 634050, г. Томск, проспект Ленина, д. 40
E-mail: sherbakovasiliy@gmail.com

А.А. Сидоров
кандидат экономических наук, доцент кафедры автоматизации обработки информации
Томский государственный университет систем управления и радиоэлектроники
Адрес: 634050, г. Томск, проспект Ленина, д. 40
E-mail: astroasregion@gmail.com

Аннотация

В работе рассмотрены особенности предоставления услуг автострахования в России, в том числе определены факторы, влияющие на стоимость страховых продуктов. Представлен обзор действующего законодательства в области страхования транспортных средств: согласно принятому закону «О государственной информационной системе «ЭРА-ГЛОНАСС» производители транспортных средств категорий B и C будут обязаны оснащать выпускаемые на территории России автомобили этой системой слежения. Данный факт явился предпосылкой к проведению исследований, ориентированных на изучение возможности адаптации зарубежной системы «умного страхования» на российском рынке. Современные процессы автострахования и технологии определения стоимости страхового полиса, применяемые в таких странах как США, Великобритания и Австралия, основаны на использовании телематических датчиков, определяющих индивидуальные показатели стиля вождения. Такими датчиками могут служить приборы системы «ЭРА-ГЛОНАСС».

Предложены различные модели взаимодействия участников при внедрении новой технологии в процесс предоставления страхового продукта КАСКО в России: государственно-частная модель, автономно-частная модель, отраслевая и аутсорсинговая модели. Определены основные механизмы повышения конкурентоспособности страховой компании при повсеместном внедрении «умного страхования»: уменьшение стоимости тарифа КАСКО за аккуратное и безопасное вождение и расширение функционала ИТ-сервисов для водителей. Проведен опрос и представлена интерпретация результатов анкетирования владельцев транспортных средств, целью которого являлось выявление готовности водителей к переходу на предлагаемую модель автострахования и определение функционала сопровождающих мобильных сервисов.

На основании анализа методики расчета тарифных ставок по рисковым видам страхования, утвержденной распоряжением Росстрахнадзора, предложена модель расчета страхового тарифа с учетом предоставляемой скидки в случае аккуратного вождения.

Ключевые слова: автострахование, умное вождение, телематический датчик, мобильный сервис, страховой тариф, методика расчета.

Литература

1. Telematics and usage-based insurance // Towers Watson & Co. [Электронный ресурс]: http://www.towerswatson.com/en/Services/Services/telematics (дата обращения 01.03.2015).
2. Morrison R. Usage-based insurance (UBI). Environmental and economic benefits and effects of state laws and intellectual property claims // True Mileage. [Электронный ресурс]: http://www.lewis.ucla.edu/wp-content/uploads/sites/2/2014/03/140303-UBI-UCLA-Lewis-Center-March-20-2014.pdf (дата обращения 01.03.2015).
3. Golia N. Usage-based insurance: 5 reasons this is the year // DriveFactor, Inc. [Электронный ресурс]: http://www.drivefactor.com/usage-based-insurance-5-reasons-this-is-the-year/ (дата обращения 01.03.2015).
4. Умное КАСКО // Страховой Гуру. [Электронный ресурс]: http://www.inguru.ru/kalkulyator_kasko/stat_umnoe_strahovanie (дата обращения 01.03.2015).
5. Harbage R. What makes a UBI score predictive enough? // Towers Watson & Co. [Электронный ресурс]: http://www.towerswatson.com/en/Insights/Newsletters/Americas/ americas-insights/2014/what-makes-a-ubi-score-predictive-enough (дата обращения 15.03.2015).
6. Об обязательном страховании гражданской ответственности владельцев транспортных средств: Федеральный закон РФ № 40-ФЗ от 25.04.2002.
7. О государственной автоматизированной информационной системе «ЭРА-ГЛОНАСС»: Федеральный закон РФ № 395-ФЗ от 28.12.2013.
8. Закон об оснащении легковых автомобилей системой «ЭРА-ГЛОНАСС» прошел второе чтение в Думе // Motor Page. [Электронный ресурс]: http://www.motorpage.ru/magazine/news/zakon_ob_osnashhenii_legkovih_avtomobilej_sistemoj__jera_glonass__proshel_vtoroe_chtenie_v_dume.html (дата обращения 15.03.2015).
9. Денисова И.П. Страхование. Учебное пособие. М.: МарТ, 2011. 240 с.
10. Самаров Е.К. Страховая математика в примерах и задачах. Учебное пособие. М.: Альфа, 2007. 95 с.
11. Методики расчета тарифных ставок по рисковым видам страхования: Утв. Распоряжением Росстрахнадзора от 08.07.1993 № 02-03-36.

Synergetics and development processes in socio-economic systems: Search for effective modeling constructs

Natalya N. Lychkina

Associate Professor, Department of Information Systems and Technologies in Logistics
National Research University Higher School of Economics
Address: 20, Myasnitskaya Street, Moscow, 101000, Russian Federation
E-mail: nlychkina@hse.ru

Abstract

This paper explores the genesis of socio-economic processes of strategic development of social transformations in a transition economy. The author studies the system patterns and structural and dynamic aspects of developing socio-economic systems, defining the need for appropriate modeling methods. Restrictive features of traditional mathematical modeling are analyzed.

Synergetics is applied as a theoretical and methodological fundamental of research and modeling of developing socio-economic systems. Modeling social behavior, individual choice, various forms of social interaction and self-organization processes are recognized to be significant aspects. Searching for methods of model description is based on carrying out interdisciplinary research in the field of economic and social sciences, using modern paradigms (system dynamics, agent-based computer simulation) and simulation technologies and analysis of their possibilities in the study of the dynamic aspects of the development processes in socio-economic systems. A general concept of simulation of developing socio-economic systems based on the principles of stratification is proposed: the micro-level reproduces individual decisions of social and economic agents, as well as the collective organizational forms; and the macro-level describes processes of evolution. The interpretation of the interactions between socio-economic configurations is based on analysis of causal dependencies and dynamic manifestations of the interpenetration of phenomena occurring in different strata of the social system. It offers methods of combining composite system-dynamic and agent-based models, allowing us to investigate the dynamics of socio-economic processes by a cyclical interaction of processes of individual and group behavior of economic and social agents at the micro level with the basic processes of socio-economic system at the macro level.

The basic directions of development of technologies of simulation in procedures and strategic decision support systems are defined as follows: a computer dynamic scenario analysis based on a generalized simulation model of the control object; a model of balance of interests in the procedures of coordination of scenarios and interests of the participants of the process of social design; a stratified description of the model set with the use of ontologies; methods of parameterization of models of socio-economic systems and the specifications of agents.

Introduction

Formation of the foundations of strategic planning in the Russian Federation in practice sets the task of creating and implementing long-term development strategy at all levels of public administration: federal, regional, sectoral and corporative. The multiplicity of possible scenarios of such development requires solution of the tasks of analysis of such scenarios and analysis of potential trends of socio-economic system development, taking into consideration numerous factors that influence these processes, including the internal potential and the processes of self-organization in socio-economic systems (SES), the synthesis of effective development scenarios, analysis of attainability of the goals set and evaluation of the risks of their unattainability.

What are the specific features of the current situation? Most economists define Russia's economy as an economy undergoing transformations in the process of shifting to market relations: a country with a market being formed and a society moving from a command economy to a market-oriented economy. This results in the determination of development of liberal democratic forms in socio-political life of the society and, first of all, establishment of social institutions, an active role of personality and citizens' freedom of will. Change of the economic model and the social structure connected with it defines the current process as a transitional process connected with reformation of the main spheres of economy and society, the step-by-step introduction of new institutional forms and mechanisms of economic, financial and other kinds, which in general corresponds to transitional processes in economic and social life of society.

We cannot solve this task without studying the genesis of socio-economic phenomena and development processes in SES which are characteristic for transition economies, and the ways of modeling them. In a transitional economy, SES shifts from its previous condition to a qualitatively new one. Potential trends of such development depend not only on the measures taken, administrative resolutions and the type of structural reforms but also on a large number of internal and external factors, stochastic, chaotic and turbulent occurrences in external environment, on the capability of the SES itself to be self-organized and self-reproduced.

1. Economic paradigms shift — computer simulation of paradigms shift

The modeling methods that were functioning efficiently not so long ago — in the socialist economy planning, under static conditions — are not suitable for studying transition processes in the economy under conditions of social transformations that are typical for the present socio-economic situation.

Why are system analysis and economic cybernetics and their traditional apparatus of economic and mathematical modeling not enough for the modeling of socio-economic processes in case of a transition economy and social transformations?

Recent publications stating that the processes which can be seen in society and in socio-economic phenomena are somewhat similar to the processes being researched by such a branch of systemological sciences as synergetics [1] resemble philosophical discussions about the possibility of implementing theoretical synergetics postulates in the field of socio-economic research into current events in Russia. Are there any new practical approaches, constructive model designs that are useful for studying socio-economic processes in the new economic situation? Yes, there are. However, due to the inertness of academic schools researchers still try to approach developing systems with the old mathematical apparatus. Rebuilding the system of society results in the need for reconstruction of the models that are used to describe it. The goal of this research is to demonstrate model designs and approaches based on using modern paradigms and technological solutions in the field of simulation modeling of socio-economic processes and social forecasting that allow us to study complex dynamic occurrences in the development of socio-economic systems which are characteristic for the real economic situation in Russia.

Traditional econometric modeling methods do not allow us to describe the processes of development and self-organization in SES. Hypotheses stating that the tendencies we have observed in the past, on retrospective data, should be continued in the future are not valid under the conditions of transition processes. Statistical data on socio-economic processes, even in case of representative samples, trend models, cannot be the basis for describing a unique, sole and specific trend, as proven by the experience of the societal system in its development process. Statistics and the experience of other countries do not help draw the details of possible SES trends when they are applied to particular socio-economic conditions and historical periods. The state of the system in the future (according to J.Forrester) depends not only on its current state but also on the entire past history, i.e. on the way that has led the system to such a state. However, an econometric model helps to reveal and explain the relationship between particular factors, explain certain stable phenomena — but does not form the whole picture.

Attempts to reveal and give a timeframe for cycles in the SES life cycle do not explain the mechanisms of changes and structural reformation taking place in complex systems. They can teach us to "follow the tides" but not to manage them. Management should be performed at bifurcation points – but modern management science does not provide methodological recommendations on this issue. Changing the political and socio-economic mechanisms of development predetermines the establishment of new principles and conditions for organizing SES development management.

Actuarial evaluations are perplexed by the stability of financial systems of the state but problems are located in other units of the socio-economic system.

Economists create models of economic systems in which they use Western instruments with the hypotheses of the "perfect market", the so-called models of general economic equilibrium where the equilibrium in markets is reached instantly, not taking time into consideration. Their macroeconomic reference points do not include such important aspects of societal systems as social behavior, population, its motivation, individual choice and activity; these reference points are shown aggregated, as rational economic agents (e.g., households that maximize the utility function). The rational choice theory has been criticized for a long time in economic and management research in respect to organizational systems. The issues of bounded rationality of the decision-maker are thoroughly studied in H.Simon's works [2].

The social sciences try to understand the way people act, the way they interact. Traditional quantitative tools of sociologic research reveal superficial occurrences in society but they cannot help reveal system patterns in the process of societal development. Social behavior emerging at the micro level can lead to global changes in the societal system. In order to study it, we need extensive interdisciplinary research and communications in a broad specter of the humanities on society, economy and psychology.

2. Genesis of the processes of strategic development of SES.
Synergetics of socio-economic systems and dynamic aspects of modeling the developing SES

Strategic management is the systemic development of a complex system with time [3] which requires analysis of structural changes and dynamic aspects of its development. Strategic management stipulates the transition of the system from its current state to a desirable goal state in a rather long period of time. The term 'development' here means the movement of the system in a phase space,

the change of the system states stipulated by external and internal causes, in the process of which new structures and systems can originate. To determine the character of this movement, the path of the system's advancement with time, which depends on the current state and the potential of a system, on external impact and the character of managerial decisions – this is the main task of analysis and dynamic modeling. The dynamic model of a system is nothing other than a time-ordered sequence of its states, the last of which is equivalent to the goal of the system [4]. It is important to remember that in case of a developing system the dynamic model should describe the structural changes, and also the mechanisms of origination of new forms and structures that emerge in the process of its development.

3. Structural and dynamic complexity of SES, specific aspects of developing SES

We shall consider some aspects of researching and modeling socio-economic systems and processes that are stipulated by structural and dynamic complexity of this particular class of simulation objects and shall determine specific aspects of developing SES:

✦ structural complexity. Non-uniform structure, structural changes;

✦ numerous cause and effect relations, direct and inverse relations within the system structure. As a rule, cause and effect are separated in time and space;

✦ a large number of influencing factors (multiple-factor character). Non-linear character of occurrences.

✦ dynamic complexity (delays, different process speeds, oscillating processes) complicates the interpretation of system behavior (interaction);

✦ management via inverse relations, adaptability;

✦ dependency on history: the state of a system in the future depends not only on its previous state but on its entire past history as well [5];

✦ developing SES. Analysis and choice of a development path. The rules of decision making (development scenarios) can also change with time; the points of application of managerial efforts can change with time;

✦ counter-intuitivity of complex SES [5, 6], limited opportunities for expert evaluation;

✦ conflicts between long-term and short-term decisions;

✦ behavioral aspects connected with an entity of particular elements of a complex system;

✦ uncertainty, including uncertainty in determination of development goals. Random and chaotic factors, stochastic process. Uncertainty of the environmental influ-

ence on the system. Chaos. Uncertainty of the development of the system with time;

✦ evolution and transitional processes. Instability, phase transitions in the system;

✦ self-organization, dynamic behavior can emerge spontaneously, depending on the internal structure and impact from the external environment. Generation of the internal organization of SES through self-organization (and self-reproduction).

Understanding all the processes of economic development and societal development is more complex than equilibrium [9, 10]. SES can be both in equilibrium state and demonstrate specific dynamic behavior in the development process. If the change of the external environment exceeds a particular limit, the system can become unstable and start moving to another phase or form of internal relations. Stability is replaced by instability and phase transition. There are several phases in the process of SES development. Under the influence of exogenous factors of stochastic and turbulent character with unidentified dynamics the system loses its stable state, its structure is being changed. Under the influence of impulses produced by the transforming structure, self-organization processes are started at the micro level; these processes can lead to origination of new forms and organizations and, thereby, the development processes — and the transition to a new state - in such systems are initiated. Its further stability or instability depends on a large number of factors. The most important case is the phase transition, i.e. bifurcation of the most probable ways and trends of the SES sectors being examined. Transition through bifurcation points can be accompanied by negative trends, degradation, system destruction, protracted stay in chaotic state. The further way of development of a complex SES depends on correct management. It is known that management should be performed at bifurcation points — but modern management science does not provide methodological recommendations on this kind of management.

What if we test the managerial decisions we make with the help of computer models? However, the established practice of making managerial decisions is to make decisions "here and now", without being responsible for the long-term consequences of such decisions. Classification of modeling approaches on the basis of the object and subject principle (some models study the object of management, analyze its functioning process; other models (which study operations) create, synthesize optimal management and managerial decisions) only complicate the procedures of making long-term forecasts. To look into the distant future, to form a strategic vision — such a task can be solved better with the help of qualitative judgments and philosophic generalizations than with the help of scrupulous quantitative analysis. Uncertainty in forming the goals of strategic development, in making forecasts in a complex environment impedes social planning and development of scenarios of the future, even in case of massive expertise and foresight technologies. Of course, we can increase the amount of experts' responsibility but their capabilities are limited. It is necessary to help experts, to provide them with computer tools, some kind of "computer foreteller" that could help them evaluate and compare development alternatives, evaluate long-term consequences of the decisions being made. The modern theoretical and methodological basis of system modeling and high-tech solutions of simulation modeling can solve this task which is overwhelming for the expert community, and can also help create consolidated (for different expert groups, ensuring the conciliation of the interests in the triad: civil society, state, business) scenarios of movement to the desired future. The main thing is to include the expert and analytical community, the persons responsible for developing managerial decisions and all the interested parties (civil society, business) in the interactive process of social modeling of the future.

4. Let's touch upon the issue of categories and essence of simulation modeling. It's like nothing on earth…

From the standpoint of history, simulation modeling (SM) has had a relatively short development cycle in comparison with other types of modeling, starting from the 1960s [12]. Its establishment brings modeling theory to a qualitatively new level. As a result of the development of information technologies and computational methods (simulation as a type of modeling never existed without computers) new high technologies appeared in this sphere in the 1990s [13]. However, attempts to relate simulation to the modeling types which existed before its origination led to misunderstanding of the simulation method and restrained its application for solving practical problems.

Scientific and academic books abound in a large number of interpretations and incorrect definitions of simulation modeling (from a type of analog simulation to Monte Carlo method extensions). In order to understand why simulation modeling, in the author's opinion, can be used for describing processes in a developing SES, it is necessary to determine the categories and comprehend the character of simulation modeling [14].

We shall determine the following categories:

✦ methodological foundation and theoretical and

methodological fundamentals of simulation modeling;

✦ simulation modeling method;

✦ character and essence of simulation model;

✦ simulation modeling technology;

✦ simulation modeling paradigms (concepts);

✦ simulation modeling tools;

✦ computational methods used in simulation modeling.

The theoretical and methodological fundamentals of simulation modeling have been formed in the process of establishing a broad specter of system methods and system approaches, general systems theory, cybernetics and synergetics, adaptive regulation and automatic control, information theory and, finally, modeling theory - the system modeling technology [15], according to the definition given by the Russian classical scientific school. The methodological foundation of simulation modeling is applied system analysis which facilitates practical application on the basis of the system analysis methodology for the purpose of solving various complex problems in different fields. The central procedure of system analysis is the construction of a generalized model which reflects all the substantial factors and interconnections of a real system. Attempts to formulate a general theory of simulation modeling [16] have always led to particular generalizations that had some intersections with all the branches of systemology and a search for a universal methodology for describing processes in complex systems. In practice it has always been a way of solving particular applied problems. Military people were the first ones who used simulation in order to reduce the costs of expensive experiments. They were followed by engineers who needed reasoning for project decisions in complex technical systems; now it's the turn of economists, sociologists and managers — because in these fields experiments with real-world systems are wholly inadmissible.

The establishment of synergetics and its variations studying objects of different character [1, 10], in comparison with the complex of systemological sciences, its predecessors, has updated dynamic systems analysis, focusing it on research of specific structural and dynamic changes in complex systems, studying the transition processes from chaos to order and back, including the processes of self-organization and self- disorganization in open, non-linear environments of different character. Though cybernetics and automatic control theory, as comprehensive management sciences, included negative inverse relations, they generally studied the issues of ensuring system stability. Moreover, synergetics is a theory of non-stationary, developing systems for which the influence of fluctuation becomes a reason for substantial changes in the system behavior. The synergetic and information approach can be regarded as a further development of the system approach [11] but it provides new opportunities for studying processes and events in society not only in stationary state (homeostasis) but also allows us to analyze the processes of development and disharmony in complex systems.

The simulation modeling method is an experimental method of research into a complex system on the basis of its counterpart — a computer model directly replicating structural and dynamic properties of the object being modeled, which combines the specific features of the experimental approach and the specific conditions of using computer technologies. This definition emphasizes two important characteristics of simulation modeling (and it is also the source of non-scientific and incorrectly translated name of this method in Russian):

◇ marked isomorphism of structural and dynamic characteristics of the object and its computer counterpart;

◇ experimental character of simulation (in order to obtain data about the object being modeled, we need to carry out a goal-directed computing experiment on a simulation model; its content and method are stipulated by research tasks).

"Direct replication of structural and dynamic properties of the object" in this definition means that at the stage of creation of a simulation model system analysts perform structural analysis and dynamic description of basic interaction processes, and it also means that the simulation model created always visualizes, allows us to observe (in the form of computer-generated imagery, flow charts, diagrams etc.) the structure and dynamics (state, behavior, development path) of the object being modeled. This is a substantial advantage of simulation modeling in comparison with other models and methods.

Simulation models belong to the class of dynamic and stochastic models, but determination of this classification characteristic does not allow us to understand their specificity, in comparison with other types of mathematical modeling. Simulation models — logical and mathematical ones — form a subclass of algorithm models. Construction of a simulation model implies combined use of mathematical, heuristic, expert methods and other methods of analysis without substantial deformation of structures, elements, processes, relations of order and connections of modeling objects [16].

Inherently a simulation model is always a dynamic model and it reflects the transition of a system from its particular state (state variables is the most important class of variables in simulation models) to another

state under the impact of a multitude of influencing factors; simulation modeling is a dynamic representation of changes of the system states with the lapse of time, a kind of a "dynamic portrait" of the system being modeled. We can state that this is a specific feature of simulation modeling, characteristic only for this class of algorithm models, which allows us to describe the processes of functioning and development in complex systems. The interaction of elements in simulation modeling of complex systems is defined with the help of special operational rules, agreements in the field of a large number of abstract elements forming one or another class of systems; numerous parallel processes are synchronized with the help of special mechanisms of model time advancement which are tied to a unified simulated time scale.

The simulation model is always a stochastic model and it allows us to research complex systems under the influence of numerous stochastic factors, random events, risks and other occurrences determined with the help of stochastic variables. Due to this characteristic and the aforementioned properties, running a simulation model is always a kind of random implementation of the modeling processes. In order to obtain statistically significant results, we have to repeat simulation model running many times. This feature often leads to misinterpretations and drawing analogies to the Monte Carlo method. But this illusion vanishes immediately when it comes to the structural and dynamic complexity of the system being modeled (a large number of heterogeneous elements and processes running with different speeds in different units of a complex system).

It is known that simulation modeling is used in case there is a lack of theoretical knowledge or formal mathematical models, or if classic issue definition in the research field leads to computational problems [12]. Simulation modeling allows us to surmount the limitations of axiomatics of any mathematical method [16].

Simulation modeling technology as a method of creating computer models has some specific features connected with the importance of the stage of problem structuring and conceptualization on the basis of the system reasoning of the experts on this issue (the simulation model is created on the basis of the structured verbal (conceptual) description of the system being modeled). The process of determining a generalized structure of the SES being modeled is an informal procedure performed by system analysts and experts specializing in this issue, in the environment of expert revisions and extensive interdisciplinary communications. It allows us to describe the process of developing system-dynamic models of complex systems as a method of structuring

the experts' knowledge of this issue and refer the models of this class to the sphere of cognitive analytics and knowledge management technologies which are connected with some areas of artificial intelligence.

Simulation modeling paradigms are the different ways of conceptualizing the object being modeled on the basis of the methods of determination and implementation of basic abstract elements and model variables and the rules of their interaction, and also the ways of model time advancement. The most popular paradigms that are already included in the modern instruments of simulation modeling: the process approach, system dynamics, agent-based modeling and simulation. Let us take a brief look at the content of these paradigms because it is important for the further reasoning.

Process-oriented simulation (or discrete event simulation DES) has become a tool for engineering and projecting not only technical systems but also business processes, logistics, transport systems and many other types of systems. This approach was initially offered by J.Gordon as a way of movement of dynamic systems in networks, queuing systems. It is characterized by a high level of discreteness, extraction of features of all the entities, description of a set of random events implemented in such systems. Initially it was oriented to the description of parallel and interacting processes; today it is a universal method of process description which allows us to study, measure time and cost of these processes, reveal bottlenecks, synchronize and optimize processes in complex systems.

System dynamics (SD) is a method of describing nonlinear dynamic systems with inverse relations which is based on flow stratification or representation of the system being modeled as an aggregate of interacting flows of different character. This method was offered by the American scientist J.Forrester [5—7]. There is no discreteness here for representation of particular processes — the system is modeled at a high level of aggregation. According to the fundamental concepts of system dynamics, the key element in describing the system dynamics is its structure represented as interacting flows, and also the interaction of inverse relation contours in its structure. Emphasis is made on the managerial aspect and on understanding the system behavior but not on obtaining precise quantitative estimates. The language of system flow charts and the specialized graphic technique for structuring the dynamic objects being modeled make this approach a very expressive and efficient tool for system analysis procedures — decomposition and subsequent composition (synthesis) of a complex dynamic system on the basis of cause-effect analysis and the principles of inverse information relations.

Agent-based modeling and simulation (ABMS) [18] emerged not so long ago as a kind of a specific branch in a broad specter of sciences dedicated to artificial intelligence and computer technologies. Presently it draws its conceptual view and existence philosophy in social sciences, behavioral economics, management, cognitive psychology etc. Agent-based models represent the real world as separately specified active elements which are called agents; these elements interact between themselves and with their environment. The activity of the agents is expressed in their capability to form individual behavior. The behavior of a complex system emerges as the result of interaction of agents in which they perform their behavior; this allows us to observe and study the patterns and features characteristic for the system in general.

There are also other paradigms of simulation modeling; the process of searching for a universal methodology of structuring dynamic systems is not finished yet. Presently the simulation modeling paradigms are the different views of the description processes and events occurring in complex systems. In practice, structuring of the object being modeled is performed in the context of the paradigm being applied.

Modern simulation systems implementing the modeling algorithm that reproduces the functioning of a complex system in which a particular concept of structuring is fixed are high-tech tools of system modeling. The merits of computer sciences and the development of information technologies and their intensive use in business applications and socio-economic research have made the simulation modeling tools available not only for experienced IT specialists but also for managers, administrators, data domain specialists and decision-makers, thanks to the user-friendly visual interfaces, tools for forming models with the help of chart representation and other representation types. Simulation modeling tools provided additional opportunities in the area of carrying out experimental research, visualization and interpretation of the results of simulation modeling etc. [13, 14]. Simulation modeling technologies and tools continue developing due to reference models and library solutions for different application fields. Promising lines of development of simulation modeling technologies and tools are connected with the search for universal schemes and concepts of structuring, stratification support, scenario approach development. At present, particular trends of these high-tech solutions are aimed at transferring simulation modeling to those users who are meant to create and use them – decision-makers and experts on specific issues!

In spite of the aforementioned user-friendly character of modern simulation systems, according to their computational character they are extremely complex analytical tools which allow us to perform quantitative analysis of SES functioning (in contrast to structure-functional modeling and cognitive analysis) with the help of various computational methods (time advancement mechanisms and scheduling for dealing with temporal collisions, numerical methods, random action generators and mechanisms of statistical information gathering and processing by objects and time marks of the system being modeled, and many more), and a number of analytical applications (statistical procedures, optimization methods etc.) that support the goal-directed computing experiment on a simulation model.

The stages of source data gathering and processing, the procedures of dynamic models parameterization, specification of processes and agents require analysts to be competent in different areas of applied mathematics. Simulation model creation is always a well-organized project, and the process of its development requires collaboration of experts on specific issues, system analysts, IT specialists and programmers, mathematicians and project managers. Procedures of organizational interaction between project participants are established. Modern consulting has a lack of experience in managing such projects [19]; this is one of the influential factors holding back the implementation of SM-based solutions in the practice of public and corporative administration, social projecting of the future, in the infrastructure of information, expert and analytical activities in the sphere of strategic planning and scenario analysis, foresight technologies support.

5. General concept of simulation of developing socio-economic systems

Modeling of a complex SES has a problem of stratification of SES structural layers and interpretation of interaction between the layers. Different layers of a complex system are characterized by different degrees of organization and the character of dynamic processes in different strata of such a system. We can conventionally determine particular strata in SES description and examine the cyclical transitions between socio-economic configurations:

✦ the micro level – the main focus of research is the individual solutions of economic and social agents;

✦ the meso level – the collective organizational forms (and social groups in the society system);

✦ the macro level – the processes of SES evolution and development.

Internal dynamics and the processes occurring at the micro level and meso level of SES have a significant impact on the behavior of the entire system and determine the path (trajectory) of further development of the system. At the macro level new system characteristics of society emerge. On the contrary, the processes at the macro level form the environment for the lives of many individuals at the micro level, where they implement their decisions depending on the current socio-economic situation. Approaches to SES stratification based on structural approaches need to be supplemented with interpretations of interactions between downward and upward strata of the socio-economic system that describe the cause-effect relations and dynamic manifestations of the interpenetration of phenomena occurring in different strata of the socio-economic system. Consistency in the consideration of the societal system and socio-economic system is enhanced by the cyclical character of descending and ascending interaction between the main strata of the simulated system, highlighting aspects of such interaction in systems of various types. It turned out that a verbal description of such interaction is not enough to reveal the dynamic aspects of manifestation of such interaction between different strata of the socio-economic system.

Let us consider the general approach to the creation of simulation models which describe such phenomena in socio-economic systems. The simulated model of a society system should connect the micro level at which individuals make decisions and take actions, and the macro level which describes the state, the basic structure and the development of such a system. All the model variables are constantly changing during the long-term period under the impact of external and internal factors, in the environment of transforming system structures and socio-economic system characteristics.

In the process of creation of a generalized simulation model of a socio-economic object, system dynamics models and methods are used. At the macro level model designs are produced by means of the aggregated system-dynamic models describing the main elements and processes of development of a societal system: population, economy, production and social infrastructure, environment and other factors of social life. In terms of non-linear dynamics, formation of this model stratum is an attempt to present the fundamental equation of a system, with the main phase variables describing evolution, dynamics of socio-economic system in aggregated form. However, Markov hypothesis does not allow us to set up formalisms and mechanisms of development in such systems, due to the complex cause-effect relations,

the aim to describe the entire retrospective of such development and its influence on current and future states of a developing SES.

The methodology of system dynamics allows us to simulate dynamic processes at a high level of aggregation; it is based on the idea of functioning of a dynamic system as an aggregate of flows (of cash, people etc.). There are two sections in the general structure chart of system dynamics models: flow chart and information chart. The latter joins together numerous managerial decisions and scenarios tested in a simulated model. System-dynamic simulations of SES are the flow type models: resources (natural, human, financial etc.) are being depleted, resources are being replenished, and they can be described as a network of heterogeneous flows. The state of SES is described by specific variables − "levels" (the number of people of different categories, production facilities, financial resources etc.), and it defines the accumulations forming within a system. External influences and managerial decisions determine the rate (of reproduction, resource consuming, dynamics) of the system being modeled. On the basis of processing of the experts' knowledge all the factors functioning in the particular system and cause-effect relations between them are determined. System-dynamic model is aimed at conceptualization and structuring of problems, understanding of the occurring processes, revealing of the sense, - system dynamics methods are close to the reasoning of strategic managers. The societal systems modeling process is performed in the environment of broad interdisciplinary communications, participation of specialists in different fields in the expert revision procedures. The methods of ontological modeling and cognitive analytics are efficient tools of such work which help reveal implicit knowledge in the expertise procedures.

System dynamics offers a paradigm, a methodology and a technological approach which differ from the traditional ones. It allows us to analyze complex dynamics, non-stationary socio-economic systems being in a process of transition, a structural change, with uncertain and dynamic external environment. Generalized description of a SES presented at the macro level should be based on the terms of correspondence of its internal structure, functional organization (through the relation of its constituent elements, key units of development) and its development dynamics. The SES structure, being changed and transformed under the impact of internal impulses, influences the dynamics (behavior) of SES and its development processes. The causes of destabilization and changeover to other dynamic modes are located in interacting dynamic processes occurring in different SES

units and at different rates. It is the possible paths and the route of long-term evolution of a SES that are the main subjects of analysis in system-dynamic simulated models. The major task of administration of a developing system is to manage the process of development through controlling time, timely revealing moments and twists in development, branching of dynamic processes and moving through bifurcation points and "points of no return", performing analysis of trajectories of the SES sectors being studied, identifying the points of application of managerial efforts. Model designs of system dynamics are based on reliable econometric estimates which are used for bringing the system flow charts of the model to well-defined quantitative proportions at the stage of its parameterization. Parameterization of dynamic models of SES is performed on the basis of mathematical models according to the results of monitoring socio-economic processes, with the use of the intellectual data analysis method, the results of examination and the results of sociologic studies.

Time is a special variable in the scenario task of managing a dynamic system as well. Development of such scenarios, taking into consideration their dimension (numerous control parameters), distribution in time (character and time of application of managerial impacts which perform the changeover of system from its current state to the desired goal state in a rather long period of time), should be based on the information about the possible paths of SES development and analysis of the impact of managerial decisions. Methods and technologies of generation of possible SES management scenarios require the active participation of experts and decision-makers in procedures of development of such scenarios according to the results of possible behavior trajectories analysis, which are performed with the use of the simulation model of the object of management. Strategic management analytics is set up in the form of procedures of dynamic computer-based scenario analysis using a generalized simulation model of the object of management. The scenario approach allows experts to form possible development scenarios or SES movement paths based on the information about the SES state and structure as well as action programs (plans), and to analyze them with the help of a simulation model.

At the level of micro processes description, aggregated system-dynamic models of SES are supplemented with agent-based models and also with models describing the interaction of a multitude of social groups.

The human factor in both its individual and collective manifestations is essential in the study of socio-economic processes. The active elements of the economic and

social system are people, individuals. They are a rather complex system as well. An individual can have certain rationality if he is acting in a market, but he never possesses complete information. He competes or collaborates, changes his behavior under the influence of changes in his institutional environment and the strategies being implemented by other participants. He accumulates experience and knowledge, enriches his mental models on the basis of which he makes decisions. The operations of agents in the market are predetermined by personal interest and individual behavior, institutional influence, relations which are formed between the agent and the institutional environment. There are a number of sciences — first of all, organizational behavior, cognitive psychology, — proving that individuals are irrational, that they have mentality, emotionality; their actions are guided by their own rules, by implementation of their personal mental models. An important stimulus for the development of this branch was the replacement of the economic paradigm about the rational behavior of economic agents and the limitations of the mathematical apparatus supporting it by the concepts of bounded rationality of decision-makers, the development of the ideas of learning organization, the search for methods of describing intellectual economic and social agents' learning on the basis of experience.

An individual in a new societal system (civil society, liberal democratic forms) is not only an economic agent maximizing the utility function but a person who has freedom of choice and implements ways of individual behavior. His activity vector to a certain extent is characterized by "passionarity" which varies in different social groups. The composition of social groups is changeable. It is formed under the influence of general socio-economic trends taking place in the society. These mental models are formed not in the market but in the process of social agents' interaction with society, with other individuals, and they include not only the rational choice mechanisms but also the values formed in the process of social reproduction of value. Different individuals have a multitude of potential behavior types. Social sciences study the phenomena of origination of social and human capital which are formed in the process of social interaction and social reproduction. Therefore, at the micro level of SES we take into consideration the decisions and actions of individuals, motivation and behavior standards which are the characteristics of certain social groups.

The behavior of such a person, representative of the society, individual choice and his communications in society and economic life can be described using multi-agent simulation. Algorithmic constructions of such models can

reproduce the individual behavior of such active agents at the micro level of the societal system. An agent can be an intellectual one (or not too intellectual), learning from his experience. Characteristics of an agent change with time. He alters his decisions under the influence of changes in his environment. He interacts, exchanges information with other participants of a socio-economic system. Groups and structural formations are set up. Changes in the organization of the socio-economic group itself arise. They influence the socio-economic environment in which the agent lives and makes his choice.

It is at the micro level where the processes of self-organization and self-reproduction are started. These processes determine stability and other dynamic manifestations in particular elements of a socio-economic system. It is possible to define a meso level at which we describe how people behave and interact; and where social groups forming as a result of such interaction are determined. Social behavior emerging at the micro level can lead to global changes in the societal system. The properties of a complex economic system at the macro level are formed as a result of agents' interaction at the micro level where their behavior is carried out. This allows us to observe and study patterns, properties and dynamics characteristic for the system in general.

The agent-based model allows us to investigate the individual behavior of different groups of agents, the specificity of their adaptation to the changing environment, and the way the processes of self-organization influence the evolution and development of the socio-economic system as a whole. The computational power of modern computers and the scaling technique allow us to produce a system with practically any level of complexity from a large number of interacting agents. The results of numerous sociological studies, theoretical knowledge in the social and economic sciences allow us to carry out clusterization, the specification of agents of the computer model. The key tasks of formation of this model level are connected with the procedures of identifying agents, their environment; the determination of the rules of their behavior (specification) and corresponding representation of the agents' interaction. The processes formed at the micro level allow us to determine classes and characteristics of agents, the rules of their decision-making, the character of interaction and information exchange between the system agents and with the external environment, etc.

Therefore, in the process of creation of an agent-based model we specify the individual logics of behavior of the process participants, and the tendencies, patterns and new properties of the entire system emerge as integral characteristics of behavior of a set of the agents that form the sys-
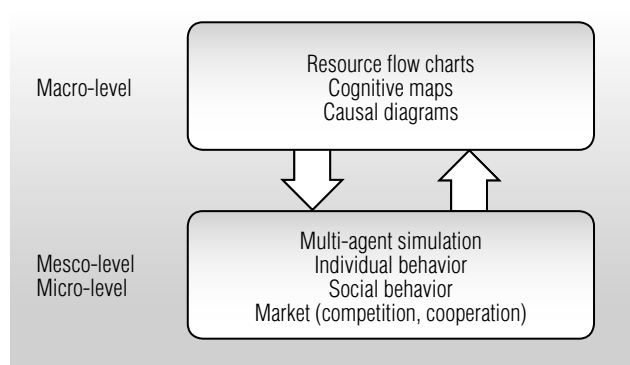

Fig. 1. The analysis of the dynamics of socio-economic processes through the cyclical interconnection between micro level and macro level in SES. Modern model design

tem, which can be revealed both at the meso level (group or social behavior, the formation of new structures) or directly at the macro level, in form of aggregated characteristics of SES. The emerging behavior and mutations in the process of functioning of an agent-based model transfer the signals to the macro level. The processes occurring at the macro level of a SES form the environment for existence of a multitude of such agents. Control actions at the macro level should be oriented to the manifestations in socio-economic processes that emerge at the micro level.

Thus, effective model designs of developing SES are created upon the principles of composite combination of system-dynamic simulation models and agent-based simulation models (*Figure 1*). Composite dynamic models of SES function on the basis of a united model and information frame which allows organizing the processes of information exchange and the mechanisms of interaction between the macro level and the micro level of the SES being modeled. SES at the macro level is an external environment in which social and economic agents carry out their individual behavior, and which to a large extent predetermines the rules of decision-making, experience and knowledge of the agents. In turn, the emerging social behavior starts the processes of self-organization, development or stagnation that define functioning and control of the SES in whole. Such an approach to creation of multi-model complexes on the basis of composite system-dynamic and agent-based simulation models allows us to study the dynamics and development of socio-economic processes through the cyclical interconnection between micro level and macro level in a particular SES.

Model construction by the socio-economic systems are considered and applied by the author of this article in the construction of dynamic models in the social sphere (health, education, housing, pension system), a regional system [20-26], organizational system, supply chain [27, 28]).

The idea of creating composite and hierarchical models is not a new one, but in practice it leads to the problems of coordination of heterogeneous models and computational problems. They make it difficult to implement such models in informational and analytical practice. But it is no problem for the simulation modeling technology. The capacities of hierarchical modeling, support of stratification and evolution of the models are easily implemented with the help of solutions based on object-oriented programming and engineering, because the simulation model is a kind of a dynamic information system. Multiple sub-models of a generalized simulation model of an object function on the basis of a single research and information frame. They are united to form a single information and analytical infrastructure which makes it possible to encapsulate it into the structure of information and analytical center of any purpose.

6. The infrastructure of information and analytical activity in strategic management

The current level of development of information technologies allows us to create the infrastructure of cognitive centers and systems of support of strategic decision-making, where a simulation model is integrated with visual and mathematic models, ontology, monitoring systems, display panels, network expertise and other infrastructural components of the decision-making procedures (*Figure 2*).

An SES simulation model is a system-forming link in the procedure of strategic decision-making in information and analytical centers (IAC), along with the monitoring system, data analysis, scenario generation methods, technologies of conducting scenario research and analyzing its results [20, 23, 29]. Strategic management analytics on the basis of IAC is set up in the form of procedures and landscape for conducting dynamic computer scenario analysis on the basis of a generalized computer model of the control object, supplemented with the methods of generation of possible SES development scenarios, expert analysis of the consequences of scenario implementation, methods and models of coordination of interests of the participants in the social planning process.

Conclusion

Created simulation models of SES are based on reliable econometric estimates in identification of socio-economic processes and specification of social and economic agents. Analytical monitoring and situation analysis form an informational basis for describing the current state of a system-dynamic model, parameterization and specification of its elements (processes and agents). The procedures of expert revisions and expert cognitive analysis are used for stratification, ontology engineering of the socio-economic systems being modeled, formation of possible development scenarios tested in simulation models, and creation of the "balancing of interests" models.
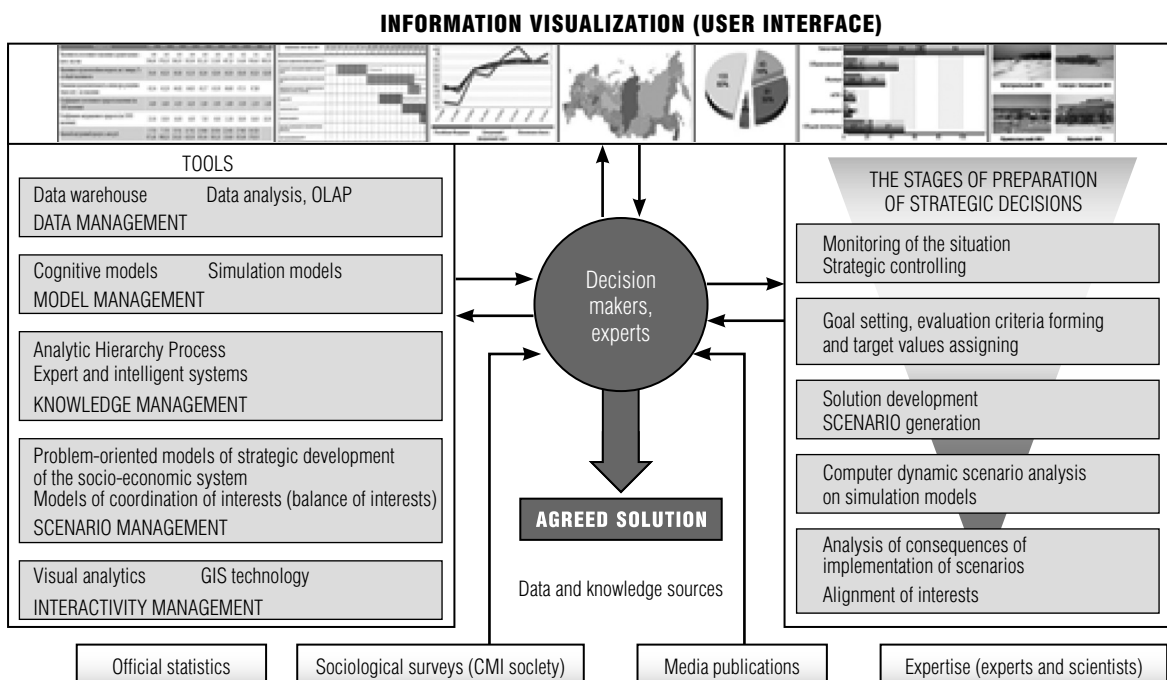

Fig. 2. The infrastructure of information and analytical activities in strategic decision support systems

The scenario approach allows experts to form possible scenarios or trajectories of SES movement on the basis of the information about SES state and structure, and action programs (plans). They can analyze them with the help of a simulation model. An important aspect in forming strategic decisions is coordination and balancing of interests of all participants of this process: state, business, people. The discussions going on in the experts and analysts community, which is trying to offer a large number of scenarios of such development, create favorable conditions for forming a consolidated development scenario or a long-term "balancing of interests of all participants" on the basis of a simulation model.

Improvement of the technologies of system modeling and scenario planning on the basis of information and analytical centers in the framework of the tasks of strategic planning in public and corporative management requires improvement of the methods of conceptualization of systems being modeled and stratification of modeling complexes [30, 31] on the basis of ontologies; enhancement of the convergent component of the process of making agreed decisions on the basis of procedures of organization and expert revisions; creation of "balancing of interests" models with the use of network expertise technologies and expert assessment and visual modeling. ■

References

1. Haken H. (1980) *Sinergetika* [Synergetics]. Moscow: Mir (in Russian).
2. Simon H.A. (1995) *Teoriya prinyatiya resheniy v ekonomicheskoy teorii o nauke i povedenii: Teoriya firmy* [Theory of decision making in economics and behavioral science: Theory of the firm]. Saint Petersburg: School of Economics (in Russian).
3. Kleiner G.B. (2008) *Strategiya predpriyatiya* [The strategy of the company]. Moscow: Delo (in Russian).
4. Drogobytsky I.N. (2013) *Sistemnyy analiz v ekonomike* [System analysis in economics]. Moscow: UNITY-DANA (in Russian).
5. Forrester J.W. (1969) *Urban dynamics*. Waltham, MA: Pegasus Communications.
6. Forrester J.W. (1971) *World dynamics*. Waltham, MA: Pegasus Communications.
7. Forrester J.W. (1961) *Industrial dynamics*. Waltham, MA: Pegasus Communications.
8. Meadows D.H., Meadows D.L., Randers J., Behrens III W.W. (1972) *The limits to growth*. New York: Universe Books.
9. Weidlich W. (2005) *Sotsiodinamika. Sistemnyi podkhod k matematicheskomu modelirovaniyu v sotsial'nykh naukakh* [Sociodynamics. A systematic approach to mathematical modeling in the social sciences]. Moscow: Editorial URSS (in Russian).
10. Zhang V.B. (1999) *Sinergeticheskaya ekonomika. Vremya i peremeny v nelineynoy ekonomicheskoy teorii* [Synergetic economy. Time and changes in the nonlinear economic theory]. Moscow: Mir (in Russian).
11. Prangishvili I.V. (2000) *Sistemnyy podkhod i obshchesistemnye zakonomernosti.* [System approach and system-wide patterns]. Moscow: SINTEG (in Russian).
12. Shannon R. (1978) *Imitatsionnoe modelirovanie system: Iskusstvo i nauka* [Simulation modeling of systems: Art and science]. Moscow: Mir (in Russian).
13. Lychkina N.N. (2001) Sovremennye tendentsii v imitatsionnom modelirovanii [Current trends in simulation]. *University Bulletin*, no. 1, pp. 135–141.
14. Lychkina N.N. (2011) *Imitatsionnoe modelirovanie ekonomicheskikh protsessov* [Simulation modeling of economic processes]. Moscow: INFRA-M (in Russian).
15. Avramchuk E.F., Vavilov A.A., Emelyanov S.V., et al. (1988) *Tekhnologiya sistemnogo modelirovaniya* [Technology of system modeling]. Moscow: Mechanical Engineering; Berlin: Technician (in Russian).
16. Kobelev N.B., Polovnikov V.A., Devyatkov V.V. (2013) *Imitatsionnoe modelirovanie* [Simulation modeling]. Moscow: INFRA-M (in Russian).
17. Pavlovsky Yu.N. (2000) *Imitatsionnye modeli i sistemy* [Simulation models and systems]. Moscow: Fazis, CCAS (in Russian).
18. Makarov V.L., Bakhtizin A.R. (2013) *Sotsial'noe modelirovanie − novyi komp'yuternyy proryv (agent-orientirovannye modeli)* [Social simulation − a new computer breakthrough (agent-based models)]. Moscow: Economics (in Russian).
19. Lychkina N.N. (2009) Retrospektiva i perspektiva sistemnoi dinamiki. Analiz dinamiki razvitiya [Retrospective and perspective of system dynamics. Analysis of the dynamics of development]. *Business Informatics*, no. 3 (9), pp. 55–67 (in Russian).
20. Lychkina N.N. (2004) Komp'yuternoe modelirovanie sotsial'no-ekonomicheskogo razvitiya regionov v sistemakh podderzhki prinyatiya resheniy [Computer modeling of socio-economic development of regions in decision support systems]. Proceedings of the *III International Conference "System Identification and Control Problems" (SICPRO'04). 28–30 January 2004, Moscow.* Moscow: IPU RAS, pp. 1377–1402 (in Russian).
21. Lychkina N.N. (2012) *Simulation modeling of socio-economic systems (system-dynamic models of the city and the agricultural area)*. Germany: LAP LAMBERT Academic Publishing (in Russian).
22. Lychkina N.N., Morozova Y.A. (2013) Dinamicheskoe modelirovanie protsessov razvitiya pensionnoy sistemy [Dynamic modeling of the pension system]. *Applied Informatics*, no. 3 (45), pp. 99–100 (in Russian).
23. Lychkina N.N., Shults D.N. (2009) Simulation modeling of regions' social and economic development in decision support systems. Proceedings of the *27th International Conference of the System Dynamics Society, 26–30 July 2009, Albuquerque, New Mexico, USA.* Available at: http://www.systemdynamics.org/conferences/2009/proceed/papers/P1068.pdf (accessed 01 February 2016).
24. Lychkina N.N., Andrianov D.L., Morozova Y.A. (2011) Social sphere modeling based on system dynamics methods. Proceedings of the *29th International Conference of the System Dynamics Society, 24–28 July 2011, Washington, DC.* Available at: http://www.systemdynamics.org/conferences/2011/proceed/papers/P1211.pdf (accessed 01 February 2016).
25. Lychkina N.N., Morozova Y.A. (2014) Dynamic simulation of Pension system development processes. Proceedings of the *32nd International Conference of the System Dynamics Society. 20–24 July 2014, Delft, Netherlands.* Available at: http://www.systemdynamics.org/conferences/2014/proceed/papers/P1180.pdf (accessed 01 February 2016).
26. Lychkina N.N., Morozova Y.A. (2015) Agent based modeling of pension system development processes. Proceedings of *SAI Intelligent Systems Conference 2015 (IntelliSys 2015), 10–11 November 2015, London, UK,* pp. 857–862.

27. Lychkina N.N. (2007) Imitatsionnye modeli v protsedurakh i sistemakh podderzhki prinyatiya strategicheskikh resheniy na predpriyatiyah [Simulation models of processes and strategic decision support systems in the enterprise]. *Business Informatics*, no. 1, pp. 29–35 (in Russian).

28. Lychkina N.N. (2013) Innovatsionnye paradigmy imitatsionnogo modelirovaniya i ih primenenie v upravlencheskom konsaltinge, logistike i strategicheskom menedzhmente [Innovative paradigm of simulation and their application in management consulting, logistics and strategic management]. *Logistics and Supply Chain Management*, no. 5 (58), pp. 28–41 (in Russian).

29. Gorbunov A.R., Lychkina N.N. (2007) Problemy, aktual'nye zadachi i prioritety v sozdanii sistem podderzhki prinyatiya resheniy i primenenii imitatsionnogo modelirovaniya v sfere upravleniya i biznesa [Issues related with objectives and priorities of decision support systems and the application of simulation in the field of management and business]. Proceedings of the *Third All-Russian Scientific-practical Conference on Simulation and its Application in Science and Industry "Simulation. Theory and Practice" (IMMOD 2007), 17–19 October 2007, Saint Petersburg, Russia*, vol. 1, pp. 27–36 (in Russian).

30. Lychkina N.N., Morozova Y.A. (2012) Stratifikatsiya kak osnova inzhenerii tekhnologiy komp'yuternoy podderzhki prinyatiya gosudarstvennykh resheniy v pensionnoy sfere [Stratification as a basis of engineering of technology for computer support of decision-making in the pension field]. *Business Informatics*, no. 2 (20), pp. 20–28 (in Russian).

31. Idiatullin A.R., Lychkina N.N. (2011) Instrumental'naya realizatsiya arkhitekturnykh modeley predpriyatiya na osnove ontologiy [Instrumental implementation of enterprise architecture models based on ontologies]. *Business Informatics*, no. 5 (15), pp. 31–42 (in Russian).

Синергетика и процессы развития в социально-экономических системах: Поиск эффективных модельных конструкций

Н.Н. Лычкина
кандидат экономических наук, доцент кафедры информационных систем и технологий в логистике
Национальный исследовательский университет «Высшая школа экономики»
Адрес: 101000, г. Москва, ул. Мясницкая, д. 20
E-mail: nlychkina@hse.ru

Аннотация

В работе исследуется генезис социально-экономических процессов стратегического развития, характерных для общественных трансформаций в условиях переходной экономики. Исследуются системные закономерности, структурные и динамические аспекты развивающихся социально-экономических систем, определяющие потребность в адекватных способах модельного описания, приводится анализ ограничительных возможностей традиционного аппарата математического моделирования.

В качестве теоретико-методологического базиса исследования и моделирования развивающихся социально-экономических систем применяется синергетика. Существенным аспектом признается отображение в моделях социального поведения и индивидуального выбора различных форм социального взаимодействия и процессов самоорганизации. Поиск способов модельного описания основывается на проведении междисциплинарного исследования в области экономических и социальных наук, применении современных парадигм (системная динамика, агентное компьютерное моделирование) и технологических решений имитационного моделирования, а также анализе их возможностей в исследовании динамических аспектов процессов развития в социально-экономических системах. Предложена общая концепция моделирования развивающихся социально-экономических систем на основе принципов стратификации: микро-уровень воспроизводит индивидуальные решения социальных и экономических агентов, а также коллективные организационные формы, макро-уровень — процессы эволюции. Интерпретация взаимодействий между социо-экономическими конфигурациями осуществляется на основе анализа причинно-следственных зависимостей и динамических проявлений взаимопроникновения явлений, происходящих в различных стратах общественной системы. Предложены методы композитного сочетания системно-динамических и агентных имитационных моделей, позволяющие исследовать динамику социально-экономических процессов посредством циклической взаимосвязи процессов индивидуального и группового поведения экономических и социальных агентов на микроуровне с базовыми процессами развития и эволюции социально-экономической системы на макроуровне.

Обозначены основные направления совершенствования технологии имитационного моделирования в процедурах и системах поддержки принятия стратегических решений: динамический компьютерный сценарный анализ на основе обобщенной имитационной модели объекта управления, модели баланса

интересов в процедурах согласования сценариев и интересов участников процесса социального проектирования, стратифицированное описание модельного комплекса с применением онтологий, методы параметризации динамических моделей социально-экономических систем и спецификации агентов.

Ключевые слова: стратегическое развитие, социально-экономические системы, синергетика, имитационное моделирование, системная динамика, агентное моделирование, стратификация.

Литература

1. Хакен Г. Синергетика / Пер. с англ. М.: Мир, 1980. 406 с.
2. Саймон Г.А. Теория принятия решений в экономической теории о науке и поведении: Теория фирмы / Пер. с англ., под ред. В.М.Гальперина. СПб: Экономическая школа, 1995.
3. Клейнер Г.Б. Стратегия предприятия. М.: Дело, 2008. 568 с.
4. Дрогобыцкий И.Н. Системный анализ в экономике. М.: ЮНИТИ-ДАНА, 2013. 423 с.
5. Forrester J.W. Urban dynamics. Waltham, MA: Pegasus Communications, 1969. 285 p.
6. Forrester J.W. World dynamics. Waltham, MA: Pegasus Communications. 1971. 144 p.
7. Forrester J.W. Industrial dynamics. Waltham, MA: Pegasus Communications, 1961. 464 p.
8. Meadows D.H., Meadows D.L., Randers J., William W. Behrens III W.W. The limits to growth. New York: Universe Books, 1972. 203 p.
9. Вайдлих В. Социодинамика. Системный подход к математическому моделированию в социальных науках / Пер. с англ., под ред. Ю.С.Попкова, А.Е.Семечкина. М.: Едиториал УРСС, 2005. 480 с.
10. Занг В.Б. Синергетическая экономика. Время и перемены в нелинейной экономической теории / Пер. с англ., под ред. В.В.Лебедева и В.Н.Разжевайкина. М.: Мир, 1999. 325 с.
11. Прангишвили И.В. Системный подход и общесистемные закономерности. М.: СИНТЕГ, 2000. 528 с.
12. Шеннон Р. Имитационное моделирование систем: Искусство и наука. М.: Мир. 1978. 418 с.
13. Лычкина Н.Н. Современные тенденции в имитационном моделировании // Вестник университета. 2001 №1. С.135—141 .
14. Лычкина Н.Н. Имитационное моделирование экономических процессов. М.: ИНФРА-М, 2011. 254 с.
15. Технология системного моделирования / Е.Ф.Аврамчук и [др.], под общ. ред. С.В.Емельянова. М.: Машиностроение; Берлин: Техник, 1988. 520 с.
16. Кобелев Н.Б., Половников В.А., Девятков В.В. Имитационное моделирование: Учеб. пособие. М.: КУРС: ИНФРА-М, 2013. 368 с.
17. Павловский Ю.Н. Имитационные модели и системы. М.: ФАЗИС: ВЦ РАН, 2000. 134 с.
18. Макаров В.Л., Бахтизин А.Р. Социальное моделирование — новый компьютерный прорыв (агент-ориентированные модели). М.: Экономика, 2013. 225 с.
19. Лычкина Н.Н. Ретроспектива и перспектива системной динамики. Анализ динамики развития // Бизнес-информатика. 2009. № 3 (9). С. 55—67.
20. Лычкина Н.Н. Компьютерное моделирование социально-экономического развития регионов в системах поддержки принятия решений // Труды III Международной конференции «Идентификация систем и задачи управления» (SICPRO'04). 28—30 января 2004 г., Москва. М.: Институт проблем управления им. В.А.Трапезникова РАН, 2004. С. 1377—1402.
21. Лычкина Н.Н. Имитационное моделирование социально-экономических систем (Системно-динамические модели города и сельскохозяйственного региона). Germany: LAP LAMBERT Academic Publishing, 2012. 181 с.
22. Лычкина Н.Н., Морозова Ю.А. Динамическое моделирование процессов развития пенсионной системы // Прикладная информатика. 2013. № 3 (45). С. 99—110.
23. Lychkina N.N., Shults D.N. Simulation modeling of regions' social and economic development in decision support systems // Proceedings of the 27th International Conference of the System Dynamics Society, 26—30 July 2009, Albuquerque, New Mexico, USA. [Электронный ресурс]: http://www.systemdynamics.org/conferences/2009/proceed/papers/P1068.pdf (дата обращения 01.02.2016).
24. Lychkina N.N. Andrianov D.L., Morozova Y.A. Social sphere modeling based on system dynamics methods // Proceedings of the 29th International Conference of the System Dynamics Society, 24—28 July 2011, Washington, DC. [Электронный ресурс]: http://www.systemdynamics.org/conferences/2011/proceed/papers/P1211.pdf (дата обращения 01.02.2016).
25. Lychkina N., Morozova Y. Dynamic simulation of pension system development processes// Proceedings of the 32nd International Conference of the System Dynamics Society, 20—24 July 2014, Delft, Netherlands. [Электронный ресурс]: http://www.systemdynamics.org/conferences/2014/proceed/papers/P1180.pdf (дата обращения 01.02.2016).
26. Lychkina N.N., Morozova Y.A. Agent based modeling of pension system development processes // Proceedings of SAI Intelligent Systems Conference 2015 (IntelliSys 2015), 10—11 November 2015, London, UK. P. 857—862.
27. Лычкина Н.Н. Имитационные модели в процедурах и системах поддержки принятия стратегических решений на предприятиях // Бизнес-информатика. 2007. № 1. С. 29—35.
28. Лычкина Н.Н. Инновационные парадигмы имитационного моделирования и их применение в управленческом консалтинге, логистике и стратегическом менеджменте// Логистика и управление цепями поставок. 2013. №5 (58). С. 28—41.
29. Горбунов А.Р., Лычкина Н.Н. Проблемы, актуальные задачи и приоритеты в создании систем поддержки принятия решений и применении имитационного моделирования в сфере управления и бизнеса // Труды Всероссийской научно-практической конференции по имитационному моделированию и его применению в науке и промышленности «Имитационное моделирование. Теория и практика» (ИММОД—2007), 17—19 октября 2007, Санкт-Петербург. Т.1: Пленарные доклады. СПб: 2007. С. 27—36.
30. Лычкина Н.Н. Морозова Ю.А. Стратификация как основа инженерии технологий компьютерной поддержки принятия государственных решений в пенсионной сфере // Бизнес-информатика. 2012. № 2 (20). С. 20—28.
31. Идиатуллин А.Р., Лычкина Н.Н. Инструментальная реализация архитектурных моделей предприятия на основе онтологий // Бизнес-информатика. 2011. №5 (15). С. 31—42.

AUTHORS GUIDELINES

Articles should be topical and original, should outline tasks (issues), describe key results of the author's research and appropriate conclusions.

Manuscripts are submitted via e-mail: bijournal@hse.ru.

MANUSCRIPT REQUIREMENTS

TEXT FILES should be submitted in electronic form, as a MS Word document (version 2003 or higher).

LENGTH. Articles should be between 20 and 25 thousand characters (incl. spaces).

FONT, SPACING, MARGINS. The text should be in Times New Roman 12 pt, 1.5 spaced, fit to the width, margins: left − 25 mm, all other − 15 mm.

TITLE of the article should be submitted in native language and English.

AUTHORS' DETAILS are presented in native language and English. The details include:

- ✦ Full name of each author
- ✦ Position, rank, academic degree of each author
- ✦ Affiliation of each author, at the time the research was completed
- ✦ Full postal address of each affiliation (incl. postcode / ZIP)
- ✦ E-mail address of each author.

ABSTRACT are presented in native language and English.

- ✦ The abstract should be between 200 and 300 words.
- ✦ The abstract should be informative (no general words), original, relevant (reflects your paper's key content and research findings); structured (follows the logics of results' presentation in the paper)
- ✦ The recommended structure: purpose (mandatory), design / methodology / approach (mandatory), findings (mandatory), research limitations / implications (if applicable), practical implications (if applicable), originality / value (mandatory).
- ✦ It is appropriate to describe the research methods/methodology if they are original or of interest for this particular research. For papers concerned with experimental work the data sources and data procession technique should be described.

- ✦ The results should be described as precisely and informatively as possible. Include your key theoretical and experimental results, factual information, revealed interconnections and patterns. Give special priority in the abstract to new results and long-term impact data, important discoveries and verified findings that contradict previous theories as well as data that you think have practical value.
- ✦ Conclusions may be associated with recommendations, estimates, suggestions, hypotheses described in the paper.
- ✦ Information contained in the title should not be duplicated in the abstract. Authors should try to avoid unnecessary introductory phrases (e.g. «the author of the paper considers...»).
- ✦ Authors should use the language typical of research and technical documents to compile your abstract and avoid complex grammatical constructions.
- ✦ The text of the abstract should include key words of the paper.

KEYWORDS are presented in native language and English. The number of key words / words combinations are from 6 to 10 (separated by semicolons).

FORMULAE should be prepared using Math Type or MS Equation tool.

FIG.S should be of high quality, black and white, legible and numbered consecutively with Arabic numerals. All Fig.s (charts, diagrams, etc.) should be submitted in electronic form (photo images − in TIF, PSD or JPEG formats, minimum resolution 300 dpi). Appropriate references in the text are required.

REFERENCES should be presented in Harvard style and carefully checked for completeness, accuracy and consistency.

The publication is free of charge.