REVIEWED PROCEEDINGS

Ninth International Scientific Web-conference of Scientists and PhD. students or candidates

Trends and Innovations in E-business, Education and Security

May 26, 2021

TIEES 2021

Trends and Innovations in E-business, Education and Security

Ninth International Scientific Web-conference of Scientists and PhD. students or candidates

Editors: Schmidt Peter - Rajnai Zoltán - Jurík Pavol

Publisher : The Slovak Society for Economic Informatics

ISSN: **2729-8493**

Date: May 26, 2021

OUR PARTNERS

UNIVERSITY OF ECONOMICS IN BRATISLAVA

FACULTY OF ECONOMIC INFORMATICS

DEPARTMENT OF APPLIED INFORMATICS

ÓBUDA UNIVERSITY BUDAPEST

DONÁT BÁNKI FACULTY OF MECHANICAL AND SAFETY ENGINEERING

THE SLOVAK SOCIETY FOR ECONOMIC INFORMATICS

REMOK FOUNDATION

ŠTEFANA KASSAY FOUNDATION

REVIEWED PROCEEDINGS

Ninth International Scientific Webconference of Scientists and PhD. students or candidates

„Trends and Innovations in E- business, Education and security "

Held under patronage

of dean of Faculty of Economics Informatics

prof. Ing. Ivan Brezina, CSc.

and of

dean of Donát Bánki Faculty of Mechanical and Safety Engineering

prof. Dr. Zoltán Rajnai,

INTERNATIONAL SCIENTIFIC COMMITTEE

REVIEWERS

Erzsébet Ancza
Magdaléna Cárachová
Nguyen Dai phuoc
Nurasil Kerimbaev
Gábor Kiss
Csaba Kollár
Silvia Komara
Tünde Kovács
Jaroslav Kultan
Arnold Őszi
Jan Pittner
Peter Procházka
Zoltán Rajnai
Eva Rakovská
Meruert Serik
Jozef Stašák
Lajos Szabó
Endre Szűcs
Vladimír Zuev

ORGANIZING COMMITTEE

Peter Schmidt
Jaroslav Kultan
Endre Szűcs
Pavol Jurík

EDITORS

Peter Schmidt
Zoltán Rajnai
Pavol Jurík

Table of Contents

DEVOPS DRIVEN COMPUTER SCIENCE EDUCATION

Igor Bandurič[1]

Abstract

DevOps concept is gaining a lot of attention in the business. Commercial subject could bring their product faster to the market and attract more customers. This paper will elaborate two main impact of DevOps in education - how to teach DevOps and how to Educate in DevOps style.

Key words

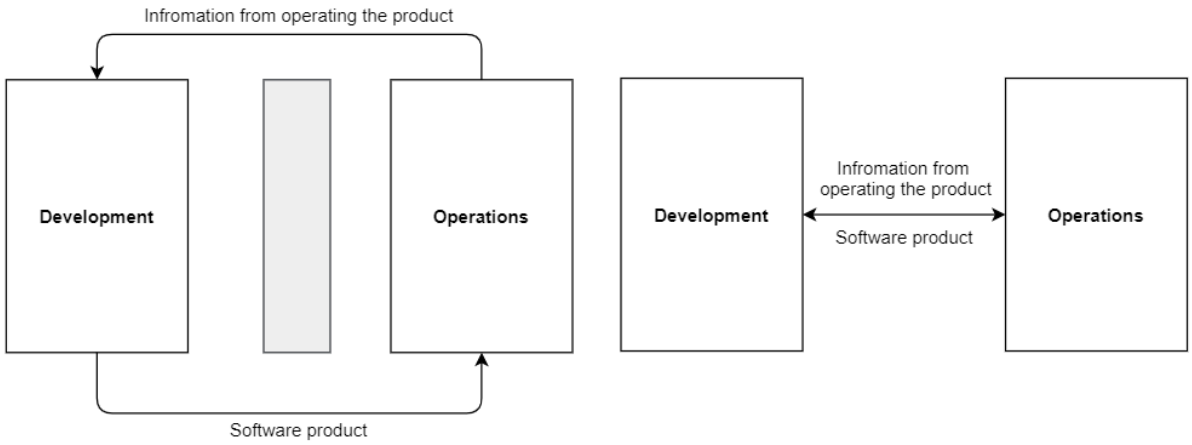
Business, Informatics, DevOps, Education

1. Introduction

DevOps is a term which is used to describe tools and culture which allows company to deliver software system to production fast a reliable. The main idea of DevOps could be simplified to a statement that DevOps is removing fence between development and operation.

Traditional approach to deliver software product is to strictly separate development from operations. This approach makes easier to manage the process however it has also some drawbacks.

Fig. 1 Classical and DevOps style



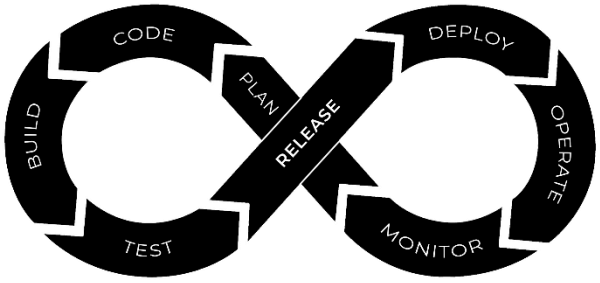
Simplified schema is presented in Fig. 1 where Classical approach has some kind of format, organizational and also technical fence between Development a Operations. This fence makes communication between left side and right side slower and isolated. DevOps style removes this fence and makes this process faster and more complex.

There are several studies which concludes that DevOps style has a competitive advantage for the companies which change their processes from the classical one.

[1] University of Economics, Faculty of Economic Informatics, Department of Applied informatics, Dolnozemska cesta 1, Bratislava, igor.banduric@euba.sk

Most common description of DevOps is visualized by in Fig 2., which denotes standard and usual phases of the process.

Fig. 2 DevOps phases



Description of the phases are:

Plan

The Plan stage covers everything that happens before the developers start writing code. Requirements and feedback are gathered from stakeholders and used to build a product roadmap. The product roadmap can be recorded and tracked.

Code

Code stage is standard coding and software development with the focus on avoid common security flaws and code anti-patterns.

Build

The build phase starts when developer has committed their code to a shared code repository. Commits triggers an automated process which builds the codebase and runs a series of end-to-end, integration and unit tests to identify any regressions. By continuously checking code changes into a shared repository and running builds and tests, developers can minimize integration issues that arise when working on a shared codebase and highlight breaking bugs early in the development lifecycle.

Test

Once a build succeeds, it is automatically deployed to a staging environment for deeper, out-of-band testing. The staging environment may be an existing hosting service, or it could be a new environment provisioned as part of the deployment process.

Once the application is deployed to the test environment, a series of manual and automated tests are performed. Automated tests run security scanning against the application, check for changes to the infrastructure and compliance with hardening best-practices, test the performance of the application or run load testing.

Release

The Release phase is a milestone in a DevOps pipeline — it's the point at a build is ready for deployment into the production environment. By this stage, each code change has passed a

series of manual and automated tests, and the operations team can be confident that breaking issues and regressions are unlikely.

Depending on the DevOps maturity of an organization, they may choose to automatically deploy any build that makes it to this stage of the pipeline. Alternatively, an organization may want to have control over when builds are released to production. They may want to have a regular release schedule or only release new features once a milestone is met. Manual approval process at the release stage which only allows certain people within an organization to authorize a release into production can be added.

Deploy

A build is released into production. There are several tools and processes that can automate the release process to make releases reliable with no outage window.

Operate

The operations team is now making sure that everything is running smoothly. Based on the configuration of the hosting service, the environment automatically scales with load to handle peaks and troughs in the number of active users.

The organization has also built a way for their customers to provide feedback on their service, as well as tooling that helps collect and triage this feedback to help shape the future development of the product.

Monitor

The 'final' phase of the DevOps cycle is to monitor the environment. this builds on the customer feedback provided in the Operate phase by collecting data and providing analytics on customer behavior, performance, errors and more.

All this information is then fed back to the first phase and the development team to close the loop on the process.

Application in Computer science education

As DevOps approach has significant impact on software delivery process in many corporations, we aimed to answer the question how the CS education should respond to those changes in companies.

It is necessary for CS student to understand the implication of DevOps on the process as the process becomes less rigorous and has many forms. There are plethora of DevOps definitions and the best way to describe this approach is by examples and description of typical situation in project and software lifecycle.

2. Methodology

The DevOps method does not have the fixed definition. We investigated current research for the definition which could be used for the purpose of CS education.

Next step was to define how the DevOps could be applied and what are the mappings between educational environment and the practical environment. We have used the analogy method where we looked on education process as standard software delivery process. We replaced software entity with educated student as substitution. So, where DevOps delivers working software which is easily changed the education process delivers student who is easily educated. Applying this method also required to substitute other parts of DevOps cycle by education process.

One of the common definitions of DevOps is “a set of practices intended to reduce the time between committing a change to a system and the change being placed into normal production, while ensuring high quality” (Bass, Weber, & Zhu, 2015)

3. Results

After application of described methodology, we have identified to approach for DevOps in Computer Science. One was expected and we did the investigation to find it on purpose. The second one was not planned and was discovered during the process driven by the methodology. The two axes of the results are:

I. Teaching DevOps
II. Teaching in DevOps

The difference in names we used are small. However, it is a different approach. The first one is a name for using tools supporting DevOps, but the teaching process do not change. The second one (II.) is a term for changing the teaching process without any implication on used tools.

Teaching DevOps

As the DevOps does not have the unified definition the best way to teach DevOps is by example and own experience. For this purpose, the DevOps games are suitable. Vast majority of this games are developed for companies. We have identified a game “The Phoenix Project” (Schilt, 2016) developed by Netherland company which can be used in CS classrooms and labs.

Tab. 1 Game parameters

Game parameter	Value	CS lab Adjusted
Duration of the simulation:	8 hours	5 weeks
Number of participants:	8 - 12	Two groups per lab

Standard duration of the simulation game is 8 hours, one working day. CS labs are usually (in University of Economics) 1,5 hour long. So, one day session could be replaced by five labs. This will cover almost half of the semester and the rest of the semester could be dedicated to other improvements and class labs.

Usual group for commercial game is eight to twelve people which means that common lab should be split to two groups. Which means that there will be to game instances running in parallel. This parallel running should be further investigated as the interference between group could be significant.

As the feedback information from production should be straightforward and automatic the best environment to create student assignment is cloud and its tools which support DevOps approach. We have checked the status for educational tiers in biggest cloud providers. The result can be seen in the

Tab. 2 Cloud for Education program comparison

	Amazon WS	Google CP	MS Azure	Oracle Cloud	IBM Cloud
Program name	AWS Educate	GCP For Education	Azure for Students	Oracle Academy	IBM Academic Initiative
Teacher role	Yes	Yes	No	Yes	No
Credit (€)	100, -	100, -	100, -	300, -	Various
Automatic registration	No	No	Yes	Yes	No

Teacher roles enable teacher to see students project and resources. This allows her to interfere with student work and fix the problem or investigate the solution. If this role is not present student have to intentionally show teacher, the results or to include teacher as a member of the member.

Credit is the amount which as available per registered user/student. Depending on the cloud provider the amount could be joined by group of students to form one project. Google CP even allow to group all the amount to the teacher and teacher can allocate funds to teams also with ability to set the budget limit.

Automatic registration allows to register student just by providing university email address. The opposite is to create all accounts manually or by students themselves.

Teach in DevOps

To describe the teaching process driven by DevOps we have chosen one of the researches which described concrete practices in company do successfully implement DevOps (Forsgren, Humble, & Kim, 2018). This release has measured many others KPI but only those four were correlated to successful companies.

Tab. 3 DevOps KPI (Forsgren, Humble, & Kim, 2018)

DevOps KPI	Description
Lead Time	Time it takes to go from a customer making a request to the request being satisfied.
Deployment Frequency	Frequency as a proxy for batch size. Smaller batches correlate with higher deploy frequency.
Mean Time to Restore (MTTR)	Given software failures are expected, it makes more sense to measure how quickly teams recover from failure.
Change Fail Percentage	A proxy measure for quality throughout the process.

To be able to use this concrete DevOps KPI in education we did mapping to a teaching process and tools.

Tab. 4 Mapping KPI to Education tools

DevOps KPI	Description
Lead Time	• Moodle (e.g., wikis) • LeanPub Books
Deployment Frequency	• Incremental assignments
Mean Time to Restore (MTTR)	• Team workshops • More delivery points for course assignments
Change Fail Percentage	• Moodle (e.g., quizzes) • Kahoot

The mapping is done by common base for activities in companies and in education process.

Lead Time

To deliver educational content to the student faster in high quality a LMS platform is considered as suitable alternative or analogy. Using standard and classical form of educational content like books, textbooks is too long to be suitable to DevOps approach. To deliver just one content platform as a Leanpub are considered. Leanpub allow teacher to publish eBook very fast and in iterative process. This platform also allows to get early feedback on the book by first readers

Deployment frequency

To have feedback about student knowledge very fast it is necessary to assign task or assignment to the students. Like software development in which in each iteration the software is bigger the assignments should also grow and should be iterative. So, one assignment is built on the previous.

Mean time to restore

One way to get fast feedback on what the student has learned is to work with her classmate and get the first feedback from them. In this way the assignment and workshops should be built. For example, we have created assignment where each group was giving feedback to another group – which was much faster than waiting for the lecturer feedback. Lecturer has stepped into only if there were some discrepancies.

Change fail percentage

Fail in study environment is to fail in exams or test. This KPI is the proxy to measure how successful the previous changes were. To measure this indicator, we must do exams and test continually. On possibility is to use before mentioned LMS the second option is to use platform for rapid quizzes creation and running. One can run the quiz before each lecture or lab. This type of platforms allows very attractive forms of execution.

Target KPI

For the proper analogy to (Forsgren, Humble, & Kim, 2018) the absolute values for the KPI should be established and researched. In the following table the absolute values for software development process are stated.

Tab. 5 Absolute values for KPI in sofware delivery (State of the Devops 2018, 2018)

Aspect of Software Delivery Performance	Elite	High	Medium	Low
Deployment Frequency	Multiple Deploys Per Day	Between once per hour and once per day	Between once per week and once per month	Between once per week and once per month
Lead time for changes	Less than one hour	Between one day and one week	Between one week and once month	Between one month and 6 months
Mean time to restore service	Less than one hour	Less than one day	Less than one day	Between one week and one month
Change failure rate	0-15%	0-15%	0-15%	46-60%

Fore each KPI the similar absolute values should be established. This research and calculation we plan to do in the following research.

4. Conclusion

The next steps in our research should be validation on both results presented in section 3. The game should be played, and evaluation form should be provided to the students. Also, the understanding of the DevOps among students should be compared with a control group.
KPI which are valid for companies should be validated in university CS labs. We do expect this to be a longer process. And as we mentioned before we should address the absolute target value for the KPI in educational process.

5. Resources

Bass, L., Weber, I., & Zhu, L. (2015). *DevOps: A Software Architect's Perspective* (1st. ed. ed.). Sydney: Addison-Wesley Professional.

Forsgren, N., Humble, J., & Kim, G. (2018). *Accelerate: The Science of Lean Software and DevOps: Building and Scaling High Performing Technology Organizations* (1st edition ed.). Portland: IT Revolution Press.

Schilt, J. (2016, February 25). The Phoenix project - DEVOPS SIMULATION. GamingWorks. https://www.gamingworks.nl/the-phoenix-project-devops-simulation/.

THE NEED AND IMPORTANCE OF PROJECT MANAGEMENT EDUCATION AT UNIVERSITIES

Veronika Horniakova[1]

Abstract

Project management is a complex activity that combines several areas of technical and soft skills. Education in the field of project management has an important place for the needs of graduates. From the surveys we conducted last year and are discussed in this article, it is possible to see what the market demands and how cooperation with universities should look like.

Key words

Project management, Education, Certifications

1. Introduction

Education in the field of project management is an important part of the preparation of graduates at the Faculty of Business Informatics of the University of Economics in Bratislava. Taking into account the current situation in the labour market, it is necessary to reflect the needs of employers. Almost every employee encounter project management.

Project management is a very dynamically developing area and it is also related to the 2 biggest problems of universities in connection with education in this field:

- development of new study programs and
- updating of existing ones to meet market requirements.

According to the PMI organization, one of the latest trends in project education is the use of so-called a hybrid model that connects academia with practice. The commercial area is increasingly aware of the value of quality project management to meet its goals and the need to invest in training and certification of future project managers. (Pappas, 2005)

As a problem of education in the field of project management can be also a fact that the process of creating a quality study program takes at least 1-2 years, students can study it for another 2-3 years at the earliest and therefore the study program is outdated when the first students after 4-5 years graduate. The situation calls for a more agile approach. Also, many faculties primarily focus only on the core subjects of their study program, but the principles of project management should be at least marginally approximated, as most students will face project tasks in their careers. (Obradovic, 2020)

Education of PM at Slovak universities

The structure of education systems varies across countries around the world. In the past there was a need to create an international framework for collecting, reporting and statistically

[1] University of Economics in Bratislava, Faculty of Business Informatics, Department of Applied Informatics, Dolnozemska cesta 1, 852 35 Bratislava, veronika.horniakova@euba.sk (1st author)

comparing data on curricula and outcome qualifications with a similar level of educational content. ISCED is an international standard classification for the organization of educational programs and related qualifications according to individual levels and disciplines. The aim of the framework created in this way is to set the criteria for the distribution of educational programs in the world so that they are mutually comparable. (OECD-EUROSTAT-UNESCO, 2015)

Universities in Slovakia are characterized by a high degree of autonomy and therefore direct management tools cannot be applied to their management from the point of view of the state. The state may use, in particular, legislative means, economic instruments, control competencies, while the operation and competencies of the state administration in higher education are defined by Act no. 131/2002 Coll (National Council of the SR, 2002).

The national classification of education in Slovakia is based on the System of Current Study and Teaching Fields and their specializations, which are stated in decree issued by the Ministry of Education and their connection to the ISCED classification. Study programs at universities compile individual universities separately, but each study program must belong to the field of study, or in the case of an interdisciplinary study program must be a combination of study fields from the System of Study Fields of the Slovak Republic for universities.

Based on the examination of the content of individual fields of study, 38 fields of study, which means 79%, provide space for education in the field of project management. However, there is no consensus in Slovakia that discusses how effective teaching should take place. It largely depends on the teacher's approach.

Education of PM at universities abroad

Germany

According to a professor at the University of Würzburg, current leader of a special interest group for project management education at universities in Germany, from 2008 to 2019, the share of project management jobs across all sectors increased from 29% to 41%. (Wehnes, 2020) Therefore, according to him, it is necessary to address this area, especially in the form of linking practical experience and new theories and methodologies developed at universities themselves, which ensures an effective mix of experience and information between academia and practice itself.

In general, project management is taught at German universities depending on the scope in three ways: (Wehnes, 2020)

- Separate subject - lectures, exercises and projects dedicated to project management;
- Group of related subjects - follow-up subjects with lectures, exercises and projects dedicated to project management;
- Comprehensive master's degree program - the possibility of obtaining a master's degree in Master of Project Management.

One of the priorities is to include project management at least briefly in all areas and orientations of the study programs. Current trends include more and more study programs taught in English and, finally, students' cooperation on projects with external partners in the field of project management. In general, agile or hybrid models are now preferred for practical

projects over the methods of the traditional waterfall model of project management. This approach brings results and is also appreciated by students for their feedback.

Russia

There are 741 accredited universities in Russia, with 496 public universities and 245 private universities. SOVNET, a non-profit project management organization has been involved in project management training since 1990. Due to the rapid progress of the project management area, in 2018 this organization established the Council for Professional Development in the field of project management for higher education and research organizations. (Tovb, 2020) Its main tasks include:

- Conduct surveys to define and refine the specifics of education in the field of project management and develop a universal model of the education system;
- Provide tools, techniques and methodologies of education in the field;
- Organize and participate in webinars, discussions, competitions, conferences, forums and other events related to the field.

The result of these activities is the establishment of cooperation with 700 representatives from more than 40 universities for 2 years, with whom the SOVNET organization regularly communicates. In cooperation with the University of Tyumen, the government and other partners, the West Siberian Technological Innovation Center was established within the national program, at which the School for Leaders of Scientific and Technological Projects was created. A project for the training of project managers for the development of project skills in today's social and technological conditions was successfully implemented at the said school. The school program consists of the following modules:

- Strategic project management,
- Expert evaluation and analysis of projects,
- Organization of work with project partners,
- Management of project stakeholders,
- Specifics of project management,
- Economic view of projects,
- Data protection and intellectual property.

The result is 45 successful graduates 2 years after the introduction. (Tovb, 2020)

China

According to a professor at Shandong University, a member of the Management Education Committee of the Ministry of Education Dr. Ding Ronggui (2020) was defined by the local Ministry of Education, the so-called a map of education in the field of project management, which consists of 5 interconnected areas for the education of a new generation of project managers.

The development of project management and the importance of training in this area is also supported by the fact that between 2000 and 2019, the number of PMP certification holders in China increased from 316 to 340,000 holders, representing more than a third of all PMP certified people worldwide. Including the holders of certification from IPMA, the number would increase to 400,000 certified persons. (Ronggui, 2020)

Role of the Project manager

The project manager, as a critical person for project management and the person who manages them all project processes and other members of the project team, responsible primarily for

- overall coordination of the project and successful completion of the project;
- resource management (work efficiency and meeting deadlines, meeting consumption of materials, equipment and use of information technology in accordance with approved sources);
- risk management, control, and minimization;
- optimization of problem solving and conflict resolution;
- communication with project team members and communication with the external environment (managing authorities, company management, suppliers, etc.).

According to Kerzner, finding a candidate with the right qualifications is not an easy task, because the selection of a project manager is more about personal assumptions than descriptions of the work (Kerzner, 2017).

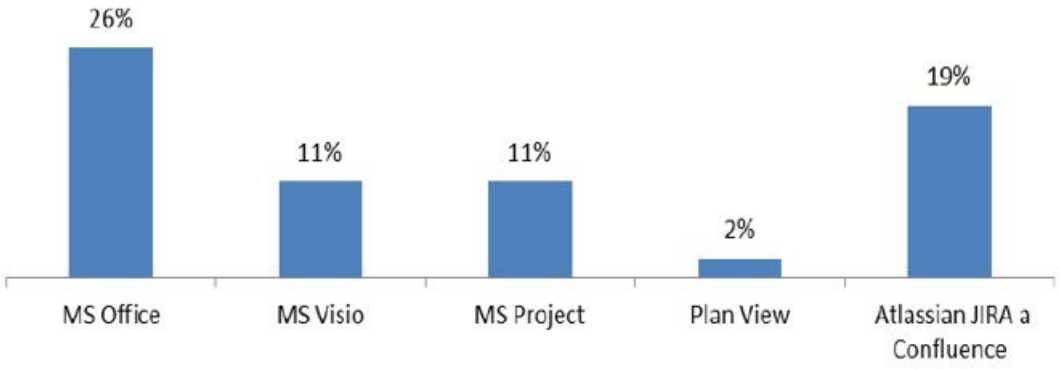
Kerzner further refers to Russell (Russel, 1976), who defines the important personal characteristics of the project manager, including:

- Flexibility and adaptability;
- Initiative and leadership;
- Ambition, activity, force;
- Good communication skills;
- Wide range of personal interests;
- Spontaneity, imagination, enthusiasm;
- Well organized and disciplined;
- More general than a specialist;
- Able and willing to devote most of the time to planning and control;
- Can identify problems;
- Willing to make decisions;
- Can maintain a good balance in the use of time (time management).

In addition to the proper qualification of the person, the project manager is also required to master technical skills, especially in the form of working with tools supporting project management. In a survey we conducted in 2020 (Turansky, 2021) and was focused on market analysis in the field of project manager's job offers, the job offer published on the Profesia portal expressed the need for practice and the ability to effectively use some of the products to manage project portfolios. The contracting authorities expressed such a request in more than 30% of job offers. Of the classic project management and project tracking tools, the preference was for MS Project or Plan View products. Project managers positions on IT projects with a

focus on the development of software solutions, especially in international companies, a requirement to practice using products from Atlassian, Trello or Confluence. The Figure 1 shows preferences of project management tools in job offer analysis.

Fig. 1: Expected knowledge of tools for project managers



Source: TURANSKÝ, P., 2021. Competences and Profile of a Project Manager in an ICT Project Environment

Survey about Project management in Slovakia

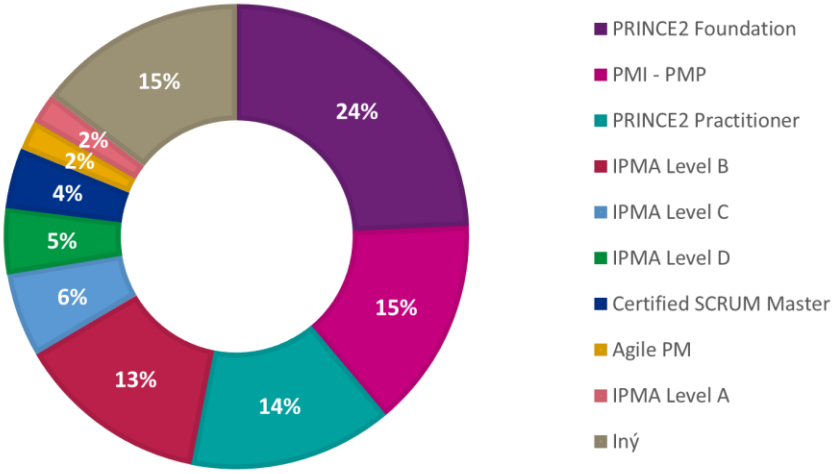
The field of project management is developing very dynamically worldwide. Therefore, in 2020, a survey on project management in Slovakia was conducted in cooperation with IPMA Slovakia, PMI Slovakia, Next level Consulting and KPMG Slovakia (2020). The survey was attended by 237 respondents from the project community across various sectors - information and communication, financial and insurance services, education, professional, scientific, and technical activities, state administration, public administration and self-government, etc. Note: The author of this article was a member of the solvers team.

The results of the survey provided several interesting findings. One of them is that 85% of respondents recommend certification in project management, with the following certifications being most preferred in Slovakia:

- PRINCE2 Foundation - basic certification based on the methodology of PRojects IN Controlled Environments from the British organization Axelos. Verifies knowledge of project management and PRINCE2 terminology, with the aim of verifying assumptions in practice for the position of project manager or another member of the project team; (Machal et al.,2015)
- Project Management Professional ("PMP") - A Guide to the Project Management Body of Knowledge ("PMBOK") certification from the US Project Management Institute ("PMI"). It has the largest representation of certificate holders among all certifications in the field of project management and confirms the ability to manage projects and lead project teams; (PMI, 2021)
- PRINCE2 Practicioner - higher certification based on the PRINCE2 methodology based mainly on case studies, which verifies the ability to effectively use the PRINCE2 methodology in practice; (Machal et al.,2015)
- IPMA Four Level Certification in levels A, B, C and D - certifications according to the IPMA Competence Baseline from the Swiss organization International Project Management Association. (Dolezal et al., 2018)

Of course, there are many more options for obtaining certificates, whether from the above-mentioned or other companies engaged in project management. Representation of holders of individual certificates among certified respondents of the survey shows Figure 2.

Fig. 2: Most preffered certifications in Slovakia



Source: IPMA Slovakia – KPMG – Next Level Consultng – PMI Slovakia. 2020. Survey about project management in Slovakia

Another interesting result is that up to 95% of certified respondents consider a certificate in the given area to be beneficial. Certification was most beneficial for the respondents in terms of their professional development and also in terms of knowledge of project management methodology and procedures, while respondents confirmed that 88% of them continue to provide at least occasional further training in project management and also train employees in the field. in its organization, whether in the form of external training, internal training, self-education, e-learning, conferences, etc.

6. Conclusion

The final output of the survey (DRAHOŠOVÁ et al., 2021) was the identification of areas in which respondents see room for improvement in the level of project management in Slovakia and would be able to help the organizations that participated in the creation of the survey. These areas are:

- Introduction and support of project management education at secondary schools and universities and greater connection with the academic community.
- Expanding the possibility of project management training outside the IT sector and project-oriented companies.
- Raising awareness of the importance of experienced project management professionals with an emphasis on a clear definition of this position and related competencies and responsibilities.
- Creating a platform, resp. community space for project managers and graduates of project training to share practical experience and support the connection of theory with practice in the field.

We can see, the labor market prefers cooperation in the education of new project managers. As in other countries, in Slovakia we see room for improvement in the teaching of the subject of Project Management and the development of cooperation with employers.

7. Resources

DOLEŽAL a kol. *Projektový management.* 1. vyd. Praha: Grada publishing, 2018. 424 s. ISBN 978-80-247-5620-2.

DRAHOŠOVÁ, S., PODOLSKÝ, D., SOCHNA, P., SLOCÍK , L., PEVNÁ, L., SCHREIBEROVÁ, R., PETRUFOVÁ, L., HORNIAKOVÁ , V., & HANUŠ, J. (2020). Prieskum o projektovom riadení na Slovensku. Bratislava; International Project Management Association Slovakia.

KERZNER, H. (2017). *Project Management: A Systems Approach to Planning, Scheduling, and Controlling.* 12. edition, John Wiley & Sons, Inc, 2017. 814 p. ISBN 978-1-119-16537-8.

MÁCHAL, Pavel. – KOPEČKOVÁ, Martina. – PRESOVÁ, Radmila. (2015). *Světové standardy projektového řízení pro malé a střední firmy*. 1. vyd. Praha : Grada publishing, 144 p. ISBN 978-80-247-5321-8.

OBRADOVIĆ, V. (2020). *Striving for project excellence at universities: Cases and Concepts.* In: The state of project management in universities: Online Conference - IPMA SmarterU SIG, https://youtu.be/dKtcho8T-qk.

OECD – EUROSTAT – UNESCO Institute for Statistics. (2015). *ISCED 2011 Operational Manual: Guidelines for Classifying National Education Programmes and Related Qualifications.* 1. edition. Paris: OECD Publishing, 2015. 14 p. ISBN 978-92-9189-174-0.

PAPPAS, L. (2005). *The state of project management training.* https://www.pmi.org/learning/library/state-project-management-training-3210.

PMI. (n.d.). *PMI® certifications*. PMI. Retrieved April 28, 2021, from https://www.pmi.org/certifications/types

RUSSEL, D.A. (1976). Managing High-Technology Programs and Project. New York: John Wiley & Sons, 1976, 396 p. ISBN 978-0-471-26557-3.

RONGGUI, D. (2020). *The PM Education in China.* In: The state of project management in universities: Online Conference - IPMA SmarterU SIG. https://youtu.be/dKtcho8T-qk.

TOVB, A. (2020). *Experience of the national project management association SOVNET in the development of project management at universities.* In: The state of project management in universities: Online Conference - IPMA SmarterU SIG. https://youtu.be/dKtcho8T-qk.

TURANSKÝ, P. (2021). *Competences and Profile of a Project Manager in an ICT Project Environment.* University of Economics in Bratislava. Bratislava: FHI EU. 51p.

WEHNES, H. (2020). *PM at German Universities*. In: The state of project management in universities: Online Conference - IPMA SmarterU SIG. https://youtu.be/dKtcho8T-qk.

National Council of the SR. (2002). Slovakia > Slovakia - Law No. 131 on Higher Education and on Changes and Supplements to Some Laws. http://www.ilo.org/dyn/natlex/natlex4.detail?p_lang=en&p_isn=83784.

PARALLEL COMPUTING IN PYTHON

Tomáš Jalč[1]

Abstract:

By default, programming languages are sequential, which means they run commands and command blocks one after the other (line by line). Many problems are so large or complex that it is difficult or impossible to solve them with a sequential program, especially due to the limited memory space of the computer. For example modeling (galaxy formation, planetary motion, traffic, weather), big data, data mining, artificial intelligence, ... Thanks to a parallel approach, we can divide a complex problem into several small problems (subtasks) and distribute these problems to available computing resources. With a parallel approach, we can make better use of available hardware. Modern computers, including laptops, are parallel with multiple processors / cores in their architectures. Parallel software is specially designed for parallel hardware with multiple cores or processors. In most cases, serial programs running on modern computers only "waste" potential computing resources. Sometimes we just need to increase the performance of the algorithm (application responsiveness) for complex operations by creating a parallel algorithm. Python is one of the few programming languages that allows us to create parallel algorithms, to split a large and complex problem into subtasks, distributed to threads or processes, with better use of hardware. In this work I introduce how can be used and what limitations can be met using Python programming language for parallel computing.

Key words:

Python, parallel, threads, processes

1. Introduction

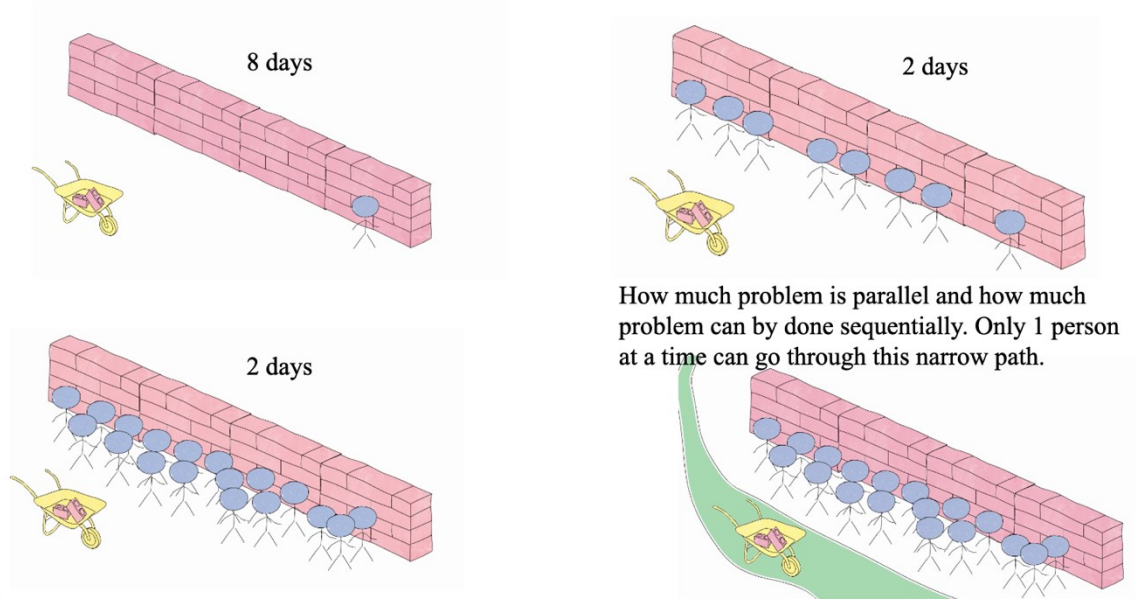
Parallel computing is a science of using multiple processing units to solve problem faster. Sometimes it is also used to improve program responsiveness. It applies in real life as well.

Let's make simple example. We want to build a wall around our house. We go to store and buy various building materials and tools for project. The contractor, we just hired, will build the wall around our house. In this project, it will took 8 days to complete entire project. Instead of hiring 1 person, we hired 2 people to do exact job. Instead of completing job in 8 days, 2 people would complete the same job in 4 days. Let's repeat this exercise and instead of hiring 2 people, we hired 4 people. We would find out that our project is completed in 3 days. Next time we hire 8 people, 8 people would complete the same job in 2 days and in the next iteration instead of 8 people we would hire 16. However, the amount of time to complete the project was still 2 days and we did not speed up any further. As you can imagine, there is a limit to speed up. We can get particular problem (just hiring more and more people) the wall will not result faster completion date. The speed up is limited some factors:
Laying of the bricks itself (we can not lay one brick before the one underneath has been finished), and

And how much problem is parallel and how much problem can be done sequentially. We can have wheelbarrow having to pass through a narrow passage, making it the part that is sequential. Only 1 person at a time can go through this narrow path using the wheelbarrow to bring the building materials. See the figure 1.

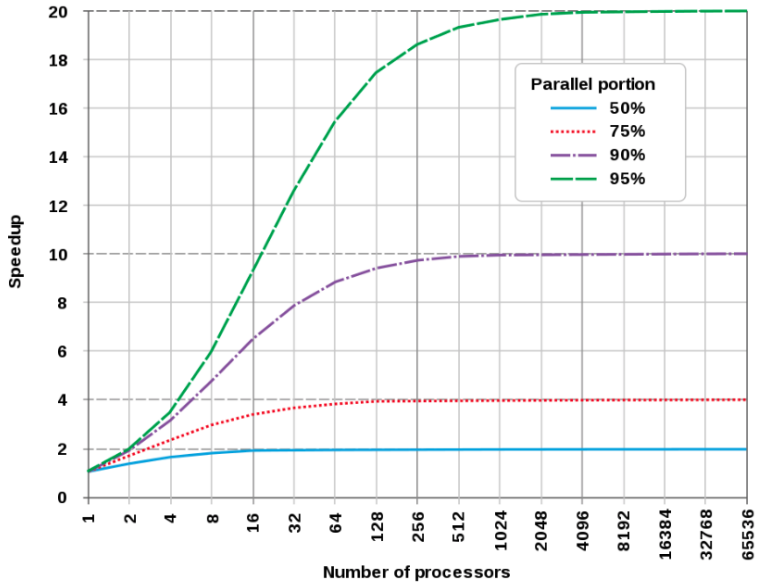
[1] University of Economics in Bratislava, Faculty of Economic Informatics, tomas.jalc@euba.sk

Fig. 1. How much problem is parallel and how much is sequential. Source: (CUTAJAR, 2021)



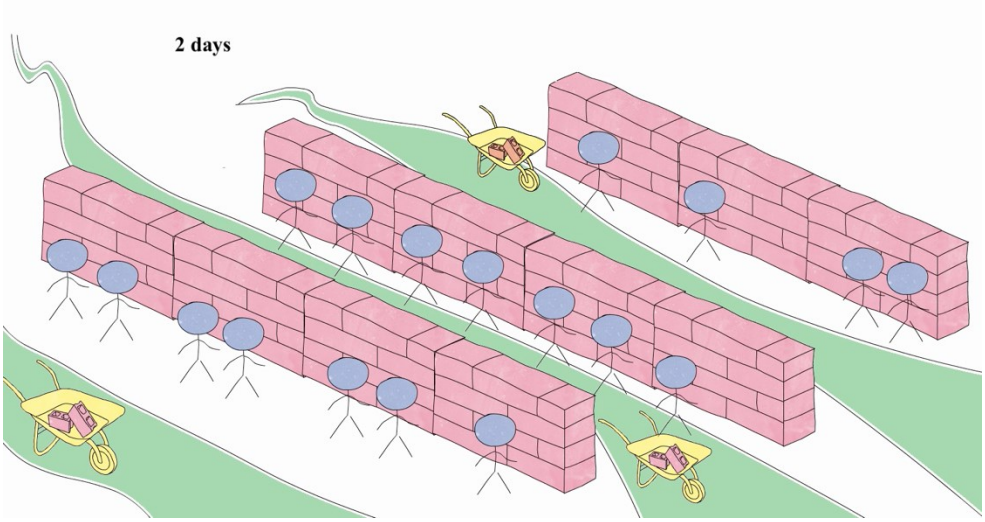
The ratio of how much of our problem is parallel and how much is sequential will dictate the limits our speed up. If the parallel parts of our program is large, if most of our program is done in a parallel fashion, we will get the most speed up. We will hit the speed up limit much later. If, on the other hand, we have sequential part of our algorithm which is quite big, we will hit the speed up limit much sooner. We call it Amdahl's law to predict the theoretical speedup when using multiple processors, see the figure 2. It tells us how scalable our problem will be with regards to speed up. We can see on the green plot if we increase the number of processors to 16, we get a tenfold speed up. The speed up of the program is much worse on the blue plot. We can see that with 16 processors, we only get twice the speed up. This is when we spend 50 % of our program executing in the parallel portion. Another 50 % is sequential part.

Fig. 2, Amdahl's law. Source: (Amdahl's law, 2021)



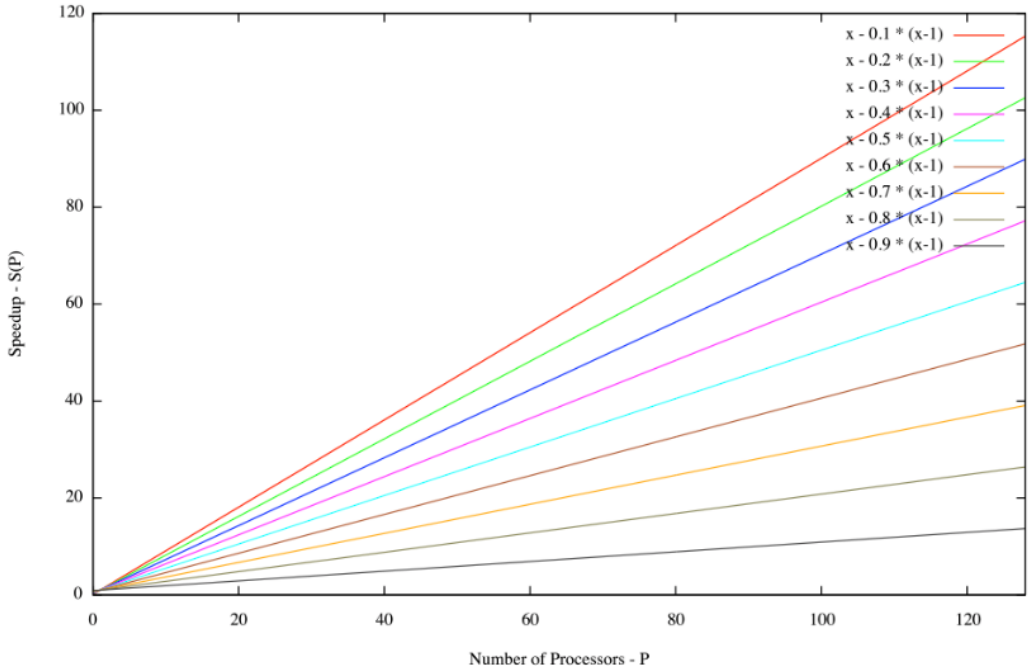
As soon as we realize that we are not going to go any faster, we can decide to increase the problem size. And instead of building one wall, we build 3 walls, for example, in the same amount of time. 2 days that will require 1 wall to be completed. We have increased the problem size and also resources, because now we have more than one passageway to build of these walls. And at this point, we have put those extra people that we hired to use. See the figure 3.

Fig. 3, increased problem size (build 3 walls in 2 days). Source: (CUTAJAR, 2021)



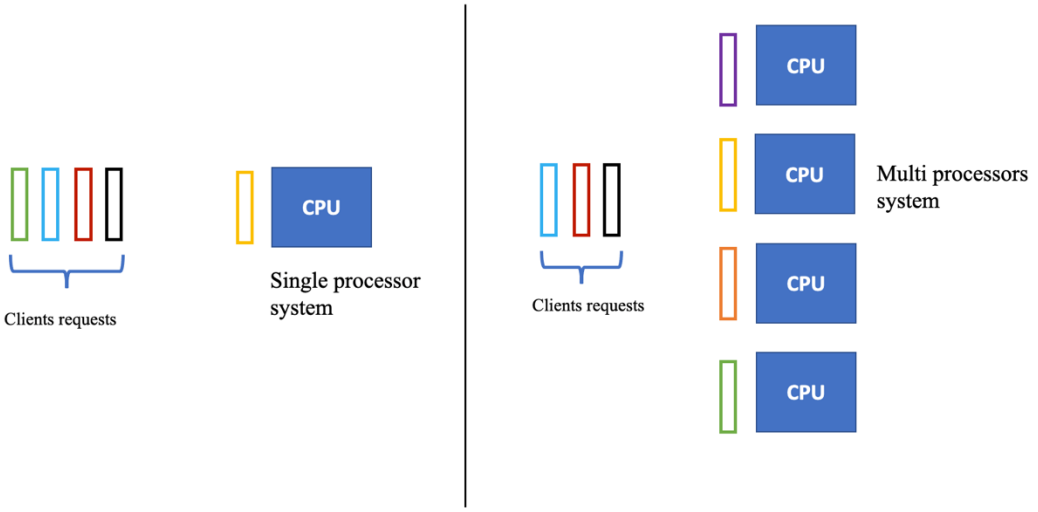
This is the basis of Gustafson's law and it states that if we take into a count that we also increase, the scalability is not limited at all. This also applies to computer programming, for example, we are developing computer game and we notice that the current video card we have, which have multiple processors, that the frame rate can only be increased up to certain point, let's say 60 frames per second. We need only half of processor for to get the same frame rate. The another half of processors we can use on something else (to increase resolution). Basically we increase the problem size to make use of all the resources that we have. If we take into a count that we can actually increase the problem size as we scale the number of processors, the speed up will be theoretically linear. The amount of speed up that we get will still be determined by the ratio of how much of our program is parallel and how much is sequential. At the figure 4 we have multiple plots, each line shows the theoretical speed up of different ratios between parallel and sequential, assuming that we can always increase the problem size. All lines are linear and they increase proportionately with the number of processors that we have. It means that as long as we keep on increasing the problem size, as long as we keep on finding jobs for those extra processors to do, we can be assured that we will still get more work done and will improve the speed up of our programs.

Fig. 4, Gustafson's law. Source: (Gustafson's law,2021)



It is important to realize that parallel computing and concurrent programming is not just about speeding things up. Sometimes it is useful to improve responsiveness of our services. For example some service which handles clients requests. If we have just 1 process that handles clients requests, each one of our client waits a long time until they got served, especially if there is a lot of load on the system. If we have multiple processors, we will get that person served faster, see the figure 5.

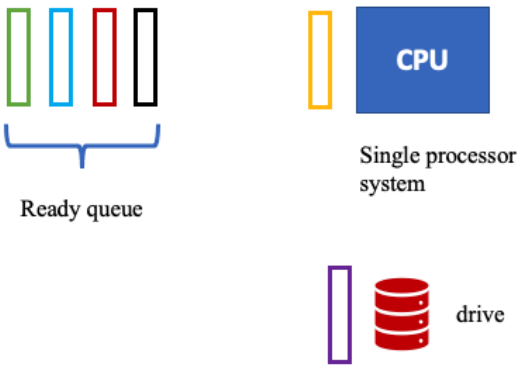
Fig. 5, single processor vs. multi processors system. Source:author.



For single processor system, processor is constantly switching which processor's task is currently handling, that gives impression that we are running on a multiple processor system. That multiple processes are all running concurrently. This constant switching of tasks is called context switching and it is handled by the operating system (scheduler) whether we are running on a multi processors system or even a single one. It is done to improve responsiveness of each program or task that we have running on that system. Context switching does not just happen on a timer, it can also happen if the process that is currently executing requests to do any

input/output (reading file on a disk or downloading something from the network). When this happens, process is removed from the execution, put into a pool of processes waiting for a response from input/output, and another process is taken from the queue and puts executing. Where all processes are waiting to be executed, it is called ready queue. When input/output operation is completed, the process goes back to ready queue waiting for its turn to be executed as well, see the figure 6.

Fig. 6, context switching with I/O operation. Source:author



2. Motivation of parallel computing

There are some facts why parallel computing to be used:

Increase algorithm performance (improve application responsiveness) for complex operations by creating a parallel algorithm.

Many problems are so large or complex that it is impossible to solve them with a sequential program, especially due to the limited memory space. We can create an application with time-consuming operations that can run parallel fashion without freezing the application. Example: copying a large file in the operating system, web search engines / databases processing millions of transactions every second.

Make better use of hardware: Modern computers, including laptops, are parallel with multiple processors / cores in their architecture. Parallel software is specially designed for parallel hardware with multiple cores, threads. In most cases, serial programs running on modern computers "waste" potential computing capacity.

3. Disadvantage of parallel computing

Point out negative facts of parallel computing:

Communication between threads: the cost of communication between threads across cores is too high. Parallel programs are sometimes slower than sequential programs. For example for small databases (we have to put together the results from different threads, from different cores into one). For databases with tens of millions of records, this is different.

Load balance: Tasks must be distributed evenly between processors. Let's make the calculation of the factorial, numbers 1 - 1000. We can not divide the interval into subsets 1 - 500 and then 501 - 1000, run these subsets on two processors, because that is not correct. The first processor terminates much earlier than the second,

Parallel algorithms are very, very difficult to implement (easy to make errors, difficult to detect errors).

4. Threads and processes

Processes and threads are independent sequences of execution.

5. Processes

A process is an instance of a running program. Processes are more heavyweight than threads. If we create process, the entire memory is copied from the current process. We make exact copy of the process. The current process continues its own independent execution. Also creating new process, it takes more time, because we copy everything. If we run software or a web browser, they are all independent processes. Operating system assigns different registers, stack memory and heap memory (RAM) for each process individually. Each process runs on completely separate processor with their own memory.

6. Threads

Thread is a lightweight process, no entire memory space is copied, data in memory, less time required, memory space saved. It is a unit performing a certain task within a given process (a process can have several threads). Creating new thread, only new execution context is created, but threads share same memory space as the original one. Threads are created a lot of faster than processes and a lot of memory space is saved. Threads share the same heap memory and each thread has its own stack memory.

Thread:

- Same heap memory
- Lightweight
- No isolation
- GIL

Process:

- Separate memory
- Resource heavy
- Isolation
- No GIL

See the figure 7 creating 2 threads with the same operation they call. Due to time-slicing[2] and GIL (explained in section 2.4), algorithm executes concurrently. If we want to create processes, then we have to import module multiprocessing and from that create object Process.

Fig. 7, creating two threads and performing the same function

```
import threading as thr

def count_operation():
    for i in range(100):
        print(thr.current_thread().getName() + " " + str(i))

t1 = thr.Thread(target=count_operation, name="Thread 1")
t2 = thr.Thread(target=count_operation, name="Thread 2")

t1.start()
t2.start()
```

[2] Thread or process that is executed in 1 core CPU or in 1 core for little period of time. The CPU starts thread 1 for a little period of time, then thread 2, then thread 1 again, then thread 2, etc.

Every time we run Python application, Python virtual machine creates 2 threads (Cutajar, 2021):

- Main thread;
- Daemon thread.

Main thread initializes our application, it is also the last thread at the end of the program because it performs various termination operations. Daemon thread is intended as auxiliary thread (for example Garbage collection process or I/O operations). We can create daemon thread implementing parameter daemon=True in our Thread object. Besides them, there are worker threads, which are created from main thread. Each thread created from the. main thread is worker thread, if parameter daemon is false. Worker threads are used for other operations, for example CPU intensive operations.

If we create processes in Python, it is recommended do not exceed number of available cores or CPUs we have available (1 process is executing in 1 CPU or in 1 core).

Python processes are started with the start() operation, which has 3 modes (Cutajar, 2021):
Spawn: available for Windows, Unix, Mac OS

Fork: available for Unix platforms only

Forkserver: available for Unix platforms only[3].

The difference between Spawn and fork

Fork copies all memory (consumes more memory), starting the process faster than spawn mode. Default mode for Unix platforms.

Spawn does not copy all memory - it does not copy key file names and others that are not needed in new processes, which is also less optimization - saving little memory. It takes longer to run the process in spawn mode. Spawn is the default mode for Windows, Mac OS.

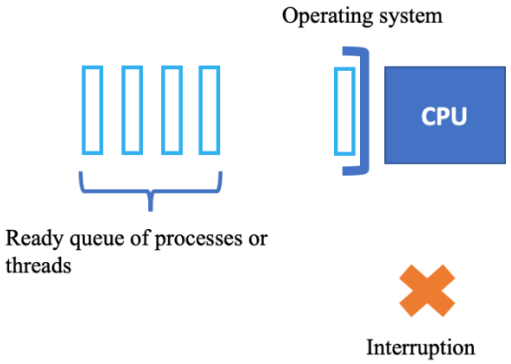
Forkserver - it is a hybrid between Fork and Spawn. It tries to save memory and at the same time be fast.

8. Context switch overhead

How to determine which process or thread to be executed. Operating system (scheduler) is responsible for scheduling the processes and threads. We have 1 CPU (central processing unit) and we have threads or processes waiting in ready queue. Scheduler will pick process or thread from the queue for CPU to be executed. If some process or thread is interrupted (for example waiting on input/output), this process needs to be swapped out and for another process replace it. Operating system runs scheduler that swaps process out of the execution a then will pick another process and replace it. The time that operating system spends on doing this, on choosing another process and swapping old one and replacing it with a new one, is called the context switch overhead, see the figure 8. Context switch is usually a very fast operation, however it is still wasted time. It is still the time that we have not spent executing our programs. It is not ideal if we are spending a lot of time doing this context switching. This is usually not a problem, if we have small number processes or threads. The proportion of time we are spending on this context switch overhead is small relative to the time we are spending in executing our code in the processed or threads. However, if we have tens or thousands of threads or processes, we are spending a lot more time doing this context switching between different processes or threads. If this happens, we say that system is trashing its spending a lot more time doing context switch overhead instead of executing our code.

[3] We can set process mode: multiprocessing.set_start_method("spawn")

Fig. 8, context switch overhead for 1 CPU. Source:author

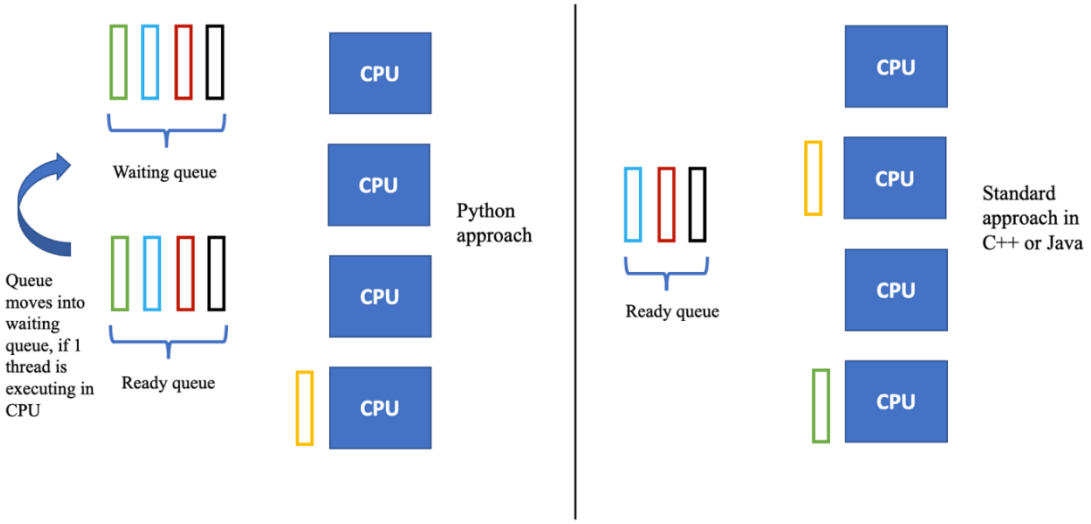


The limit how many processes or threads we can have before it becomes unacceptable for context switching depends on our system. The more processors and cores resources our system has, the more is able to support many processes and threads. If we run code that is processor's intensive and have a lot of processes and threads, it is more likely that our system will need more resources to handle the load.

9. Global interpreter lock

Global interpreter lock (GIL) is kind of limitation built in the Python interpreter. GIL is a mutex (or a lock) that allows only one thread to hold control of a Python interpreter in a given period of time. This means that only 1 thread can be executed in a certain period of time. GIL prevents multithreaded applications from taking full advantage of multiprocessor systems. There is no real parallel programming in Python (program execution). See the figure 9, if 1 thread is executing in CPU, other threads in ready queue are moved to waiting queue. If some interruptions happens (for example input/output), thread is removed from the execution and put back into the ready queue with the remaining threads (on the end of the queue). When this happens, all other threads that were on the waiting queue for this particular program, will get moved into ready queue as well. Operating system chooses another one of these threads and moves it to a processor to be executed. And it repeats. The Python interpreter locks the remaining threads for that program, moving them to the waiting queue.

Fig. 9, parallel computing using threads in Python and in other programming languages. Source:author



Python is there since the days when operating systems did not have concept of threads. Python was designed to be easy to use in order to make development faster. Over time, more and more developers began using it and it became popular in many communities. GIL allows us to increase the performance of single-threaded programs because the GIL is a single lock on the interpreter itself which adds a rule that execution of any Python bytecode requires acquiring the interpreter lock. This prevents deadlocks (there is always only one lock that is distributed by different Python objects) (Ajitsaria, 2018). Single threaded programs are faster than multithreaded in Python because the effectiveness of GIL makes a multithreaded application single threaded at some point.

GIL has significant impact on CPU bound operations (operations pushing CPU to its limit – mathematical computations, gaming / graphics, de/encryption, video/audio editing, machine learning, etc.). See the figure 10. For I/O bound operations (operations that spend time waiting for Input/Output which can come from user, file, database, webserver, etc.), GIL does not have much impact on the performance (Ajitsaria, 2018).

GIL was not removed from the Python version 3 because (Ajitsaria, 2018):

speed down existing single threaded programs,

programs developed in Python v3 would be slower than programs developed in Python v2,

it is not easy to remove GIL.

If we want to run a Python program on multiprocessor systems in parallel fashion, instead of parallel threads, we have to use parallel processes. Each Python process gets its own Python interpreter and memory space. Python has a multiprocessing module.

Or use alternative Python interpreters: Python has multiple interpreter implementations. CPython (standard), Jython, IronPython, Pypy and Cython, written in C, Java, C# and Python, are the most popular ones. Jython and IronPython have no GIL implemented and can fully exploit multiprocessor systems. PyPy currently has a GIL like CPython. In Cython the GIL exists, but can be released temporarily using a "with" statement. If our program, with its libraries, is available for one of the other implementations, then we can try them out as well.
Or we can easily create multiple threads and let them run in parallel fashion on separate processors using Java or C++ programming languages. They do not have this limitation of global interpreter lock.

See at the figure 10, CPU bound operation is faster in parallel fashion using processes instead of threads. On 2 cores CPU using processes operation finished in 6 secs. While using threads, operation is slower, it finished in 10.5 secs.

Fig. 10, CPU bound operation using threads and processes

```python
from threading import Thread
import time

def do_operation():
    print("Starting operation")
    i = 0
    for val in range(100000000):
        i += 1
    print(f"Finished operation {i}")

t1 = Thread(target=do_operation, args=())
t2 = Thread(target=do_operation, args=())
start = time.time()
t1.start()
t2.start()
t1.join()
t2.join()
end = time.time()
print(f"Time taken {end - start}")
```

```python
from multiprocessing import Process
import time

def do_operation():
    print("Starting operation")
    i = 0
    for val in range(100000000):
        i += 1
    print(f"Finished operation {i}")

if __name__ == "__main__":
    p1 = Process(target=do_operation, args=())
    p2 = Process(target=do_operation, args=())
    start = time.time()
    p1.start()
    p2.start()
    p1.join()
    p2.join()
    end = time.time()
    print(f"Time taken {end - start}")
```

"If __name__" condition for processes is required because we have to check if we are in main module (python processes run in separate Python interpreters).

10. Memory management

Let's imagine we do building a wall without any communication. For example, if we are using the same building material such as wheelbarrow or shovel, we need to talk together in order to know who is using what. And also at the end of our task, we need to communicate with the rest of the team to tell them that we have finished our job, we can merge the different pieces together. The same applies where we are doing parallel computing, our different processes or threads they need to communicate together to effectively solve the problem.
The mechanism by which different threads or processes talk together is called inter process communication or inter thread communication. All the different manners of communicating between threads or processes fall under 2 categories (Cutajar, 2021):

- Message passing (f.i. when we talk to each other),
- Shared memory (f.i. if we exchange information on the whiteboards without talking to each other)

11. Memory sharing between threads

Memory sharing is that one thread would write something into a memory location and the other thread would read it. We can share our information with multiple threads **more effectively**. Which type of communication we choose, it depends on the problem we are solving.

Communicating using memory sharing is easier when we are using threads because threads are sharing the same memory space. If 1 thread writes something in its memory, another thread executing in the same program, it will be able to access to the same memory space.

Without using thread synchronization, different threads step over each other. Why we need synchronization: If there are more threads which share the same variable in the same time, one threads overwrites variable while another thread reads value of variable. The part of our program where multiple threads or processes are using the same resources is called race condition (which are very hard to debug, because we can have running program without any issues and then one fine day it crashes because of the race condition). We have to use synchronization, to synchronize access to common resources. One of the tool is Lock or Mutex lock. The lock guards a particular chunk of code, only 1 thread at a time can access it.

Look at the figure 11. The operation increment() increments global variable x (common resource) to 1 million. When program finished, variable x would be equal 2 millions. Using Lock in or function increment(), we secure only 1 thread can access to variable x (common resource) at a time. Without using Lock, the result of our program can vary, but it will not be 2 millions, because more threads can access to variable x at a time. Function lock.acquire() acquires the lock (only 1 thread can hold the lock at a time and common resource is locked for 1 thread), function lock.release() releases the lock.

Fig. 11, synchronization between threads

```
from threading import Thread, Lock

x = 0
lock = Lock()

def increment():
    global x
    lock.acquire()
    x += 1
    lock.release()

def operation():
    for i in range(1000000):
        increment()

t1 = Thread(target=operation, name="Thread #1")
t2 = Thread(target=operation, name="Thread #2")
t1.start()
t2.start()
t1.join()
t2.join()

print(f"Operation finished, {x}")
```

12. Memory sharing between processes

Memory sharing between threads is easy because they are occupying the same memory space. With processes this is a little bit more complicated to achieve because they are running on completely separate processors and they have their own memory, they do not share memory (it is very hard for them to communicate). Typically, the communication between 2 separate processes is achieved by message passing. Other way that processes can communicate to each other is using file on our HDD and we can have 1 process to write to this file while the other one is reading it. Using this filesystem for communication works but it is slow.

Do it faster we can allocate a chunk of memory where both processes have access to this memory and use it as a communication medium between processes inside this chunk of memory. We need to make sure that the access to this memory is synchronized (only 1 process can update the value at a time). This shared memory concept creates shared memory space and also way to synchronize access to it.

Using multiprocessing.Array (Multiprocessing, 3.9.5), we can create a shared memory space in Python that can be accessed by processes (it is a memory outside the process's own memory). Using shared memory space with a lock (synchronization - only 1 process at a time) is less efficient than using thread synchronization (slower process switching). We can only store a simple one-dimensional array or variable in shared memory space for processes. See the figure 12, we want to create application where 1 process creates one-dimensional array with values "-1" and the other one will increment these values by +1 (0, 1, 2, …)|. Using multiprocessing.Array we can do that because process will share the same chunk of memory. Without it, we can not because the second process (child process) will only create copied memory from the parent process with the same data and the result will be always -1. For multiprocessing.Array the first argument is data type (i = integer), the second is data, the third is lock (default is false). We use lock=true only if one process at a time would use shared resource.

Fig. 12, shared memory space between 2 processes

```
import multiprocessing
from multiprocessing.context import Process
import time

def print_array_data(array):
    while True:
        print(*array, sep = ", ")
        time.sleep(1)

if __name__ == "__main__":
    # arr = [-1] * 10
    arr = multiprocessing.Array("i", [-1] * 10, lock=False)

    p = Process(target=print_array_data, args=([arr]))
    p.start()

    for j in range(10):
        time.sleep(2)
        for i in range(10):
            arr[i] = j
```

13. Message passing with Pipes and Queues

In this section we describe threads or processes communication using message passing with Pipes or Queues. Message passing is just another way to threads or processes can communicate with each other like natural way for humans to communicate with each other. If we send the message to our friend using phone or e-mail, there is this kind of social contract where we are waiting for a message, we receive it, we understand it and we reply back and the other person does the same from his end. This is kind of applying that pattern to our threads and processes.

If we have 2 or more threads/processes that want communicate with each other, they usually do it via medium. The thread or process that sends the message is typically called the producer. At the other end of the thread or process that receives the message is called a consumer. There 2 tools that implement this channel between thread or processes, Queues and Pipes.

14. Queue

The queue (as a channel between the producer and the consumer) has front and rear. On the producer side there is a put operation that accepts the message we want to send to the consumer. The consumer reads the message. If the consumer does not read the messages that it is receiving, the messages are piled up on the queue every time the producer sends another message. That message is just hold before the previous one. At a certain point, the capacity of the queue is reached. We can specify capacity of the queue (f.i. 100 messages) or we can decide not to specify any capacity, in which case the queue will continue to fill up while we still have memory (Multiprocessing, 3.9.5). If we have specified a fixed capacity, when we put another massage on a full queue, the producer thread or process will be blocked until there is some space on that queue. On the other end, the consumer to receive a message, we call the get operation, which reads 1 message from the queue. If capacity is specified, we will not run out of memory. If the queue is completely empty, the consumer thread or process will be blocked until there is a message that is placed on the queue. We can have multiple producers sending messages and even multiple consumers consuming from the same queue. Each producer will

just handle 1 consumer and this is the same: each consumer will read 1 message and take it out of the queue.

15. Pipe

Pipes are a little bit more heavy-weighted than they are typically used to communicate between processes. Pipes are similar to queues in a way, that there is open channel between 2 processes. Pipe has 2 ends. The difference with a queue is that pipe has not front or rear, there are just 2 opposite ends. The process or thread that wants to send the message calls the operation send and the messages are sent to whoever is consuming from that pipe. Unlike a queue, we can not specify a capacity for a pipe (Multiprocessing, 3.9.5). It will get filled whenever we run out of resources in our system and on the other end, the consuming process call receive operation to consume 1 message from the pipe. If there are no messages, the receiver will block until there is a message. A pipe can be used as what is know as half duplex, where we have the producer processed that is producing messages and the consumer that is just receiving it, or we can have full duplex as well (the messages flow in both directions). In full duplex mode, there is no point calling our processes as a producer and consumer. The other big difference from a queue is that there can only by 1 process at one end and another process or thread on the other end. Pipes have faster performance than Queues.

Queue:
Rear and front
Half duplex mode (flow in 1 direction)
Max capacity
Multiple processes or threads in both sides
Slower

Pipe:
No rear and front
Full duplex mode (flow in both directions)
No capacity
1 process or thread at one end
Faster than Queue

See at the figure 13 Queue and Pipe examples between 2 processes. Pipe is not available for threads.

Fig. 13, communication between processes using Queue and Pipe

```python
from multiprocessing import Process, Queue
import time

def consumer(val):
    while (True):
        txt = val.get()
        print(f"Received message: {txt}")
        time.sleep(1)


def producer(val):
    while (True):
        val.put("Hello, my friend...")
        print("Message sent")


if __name__ == "__main__":
    q = Queue(maxsize=15)
    p1 = Process(target=consumer, args=(q,))
    p2 = Process(target=producer, args=(q,))
    p1.start()
    p2.start()
    p1.join()
    p2.join()
```

```python
from multiprocessing import Process, Pipe
import time

def ping(pipe_conn):
    while True:
        pipe_conn.send(["Ping and time: ", time.time()])
        ping = pipe_conn.recv()
        print(ping)
        time.sleep(1)

def pong(pipe_conn):
    while True:
        pong = pipe_conn.recv()
        print(pong)
        pipe_conn.send(["Pong and time: ", time.time()])
        time.sleep(1)

if __name__ == "__main__":
    pipe_end_a, pipe_end_b = Pipe()
    p1 = Process(target=ping, args=([pipe_end_a,]))
    p2 = Process(target=pong, args=([pipe_end_b,]))
    p1.start()
    p2.start()
```

Queue example: Process p1 puts message to Queue (queue capacity is 15 messages) and another Process p2 receives message. If Queue limit is achieved, not massage can be put on Queue.

Pipe example: Process p1 sends Ping message to Process p2 and Process p2 receives it and replies back Pong message to Process p1.

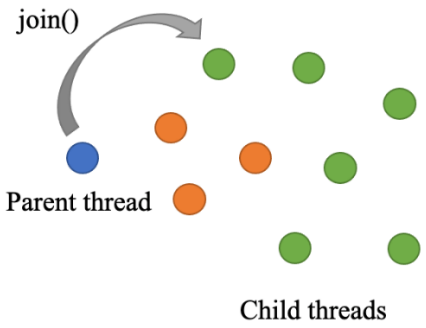
16. Waiting for completed tasks

In this section we describe simple tools (join and barriers) that allows us to synchronize different threads or processes together, in other words tools allowing us to wait for completed tasks.

17. Join operation

Join operation is that 1 thread or process can call on another thread or process to wait until thread or process finishes from its work. See at the figure 14 join operation.

Fig. 14, join operation. Source:author



This join operation blocks parent thread (it goes to sleep) until the child thread completes its work. When child thread completes its work and then it terminates, the join will release and the parent thread will wake up or unblock and continue with its execution (Threading, 3.9.5).

We can adapt the join operation to wait for multiple threads or processes completing particular task. We can do this using loop. For example if we want to render some picture, we divide picture into multiple sections and we assign a different thread to each section to render it. We start with parent thread that creates a number of these child threads and then this parent thread needs to wait for this whole group to finish their rendering. One way to wait for all the threads to finish is to use the join. The parent thread will simply have for loop that iterates over every single thread that it has created. Once particular thread is also finished, it will also terminate and the join will return. And it continues this fashion. The parent thread will call a join on another thread and wait for it to finish. If other threads have terminated without our calling join on them and we move on to them and we call join, the join will return immediately, so we move on to another thread and we will have to wait there for nothing. If all our child threads finished, the parent thread can display the picture.

See at the figure 15 very simple example with join operation, where parent thread (main thread) is waiting or blocked while child thread is running. We can simulate some kind of work calling the sleep operation. When child thread finished (after 5 secs.), parent thread is unblocked and finishes its operation. If join operation has parameter set in secs, f.i. 2 secs., the parent thread will not wait longer than 2 seconds and after that, it will continue to finish its work.

Fig. 15, join operation for threads and processes

```python
from threading import Thread
import time

def child():
    print("Child thread is working...")
    time.sleep(5)
    print("Child thread finished.")

def parent():
    t = Thread(target=child, args=([]))
    t.start()
    print("Parent thread is waiting...")
    t.join()
    print("Parent thread is unblocked.")

parent()
```

```python
from multiprocessing import Process
import time

def child():
    print("Child process is working...")
    time.sleep(5)
    print("Child process finished.")

if __name__ == "__main__":
    p = Process(target=child, args=([]))
    p.start()
    print("Parent process is waiting...")
    p.join()
    print("Parent process is unblocked.")

parent()
```

18. Barriers

Think about the barriers as being starting line before the race where every thread comes to that starting line and then once they are all in position, release them together.

If we initialize one of these barriers, we specify how many threads or processes we want to support. In this example there are 4 (threads) they are executing their code in their own pace. One of them will get to the code where it hits one of these barriers. In the code thread 2 will hit the line, which says barrier dot wait (barrier.wait()). When this happens, the thread is blocked, it is taken out of the execution. It waits for all the other threads to arrive at that same line of code (barrier.wait()). And the internal count of the barrier goes down by 1. Another thread will arrive to that barrier and it is also blocked when it call barrier.wait(). The internal count of the barrier goes down by 1. This example is the same for thread 3 and 4. When the barrier count is 0, the barrier releases all of these threads together and the count is reset back to 4. Now these threads are free to continue with their execution.

Let's have another example. We have 2 threads and a barrier of size 2. Thread 1 starts running, until it hits the barrier and it calls operation barrier.wait(). Because, it is the only thread that called this barrier, it will be blocked for some time until another thread calls this operation. Thread 2 spends longer time running until it calls itself this wait function on the barrier. And now because the barrier has 2 threads that have caused this wait operation, they are both released and they continue running. See at the figure 16, 2 threads (thread #1 and thread #2) performing some task (represented by sleep operation) and waiting on the barrier. Once the barrier count is 0, barrier will immediately release 2 threads.

Fig. 16, one barrier waiting for 2 threads and 3 processes

```python
from threading import Thread, Barrier
import time

barrier = Barrier(2)

def wait_on_barrier(name, time_to_sleep):
    for i in range(2):
        print(f"{name} running")
        time.sleep(time_to_sleep)
        print(f"{name} hit barrier and is waiting on barrier")
        barrier.wait()
    print(f"{name} finished")

t1 = Thread(target=wait_on_barrier, args=("Thread #1", 4))
t2 = Thread(target=wait_on_barrier, args=("Thread #2", 12))
t1.start()
t2.start()
```

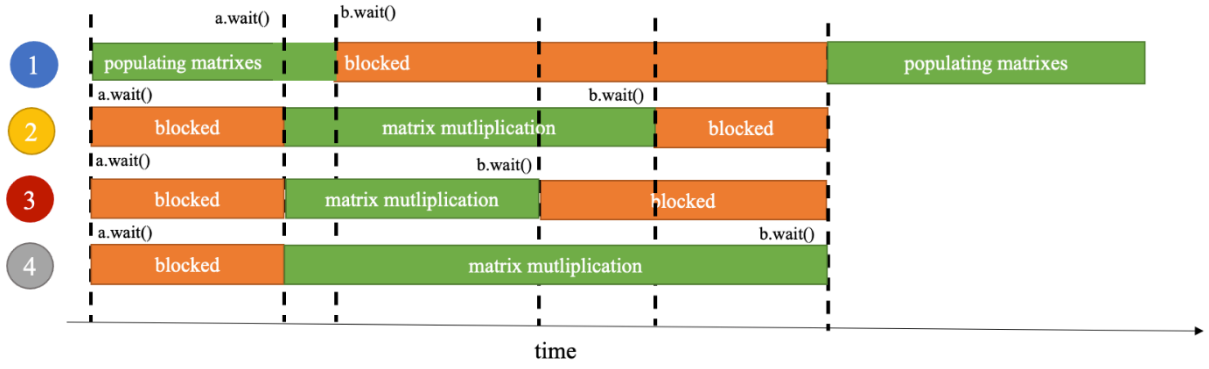
```python
from multiprocessing import Process, Barrier
import time

def wait_on_barrier(barrier, name, time_to_sleep):

    print(f"{name} running")
    time.sleep(time_to_sleep)
    print(f"{name} hit barrier and is waiting on barrier")
    barrier.wait()
    print(f"{name} is finished")

if __name__=="__main__":
    barrier = Barrier(3)
    p1 = Process(target=wait_on_barrier, args=(barrier, "Process #1", 4))
    p2 = Process(target=wait_on_barrier, args=(barrier, "Process #2", 12))
    p1.start()
    p2.start()
    barrier.wait()
```

Let's have more complicated example – matrix multiplication which is a very CPU intensive operation and it is a problem that has a very poor runtime complexity, so we can distribute the processing over multiple processors. This is cubic runtime complexity (as our input increases, the time taken to compute the matrix multiplication increases cubically). We have 3 matrixes (A, B, C), each matrix size of 3 x 3. It takes 1 second for multiplication to be done, if we have 6 x 6 matrix, it takes 8 seconds for multiplication to be done, if we have 12 x 12 matrix, it takes 64 seconds for multiplication to be done. The idea is to assign a separate thread for each row of our output.

If we have 3 x 3 matrixes (A, B, C) and 3 threads, each thread will be responsible to solve a particular row of our output (matrix C). We use memory sharing, each one of these threads can read input values, but also update the same structure as output. We also have main thread that will generate the input matrixes. In order to synchronize between main thread and all the other threads, we will use 2 sets of barriers (A, B). Each set of this barrier will be initialized with a count of 4 (4 threads). Initially, when we start the child threads, these 3 threads, we will get them to wait on the barrier A (all of them will wait for new work to be available). Then main thread will populate input matrixes. Once it has finished, it will also go to wait on the barrier A itself. Then barrier will became unblocked (it has 4 threads and the count of barrier is 0), all the threads will be unblocked. At this point, the main thread will go back to wait on the second barrier B. The barrier will wait for the work to be completed and 3 child threads will

start working each row separately. And once they are done, they will go wait on this second barrier B, the second barrier B will unblock. Once this unblocks, this process will repeat again. See the whole process at the figure 17.

Fig. 17, matrix multiplication using 4 threads and 2 barriers (A, B). Source:author



19. Effective threads and processes handling

In this section we describe pools and semaphores as tools used for effective threads and processes handling in Python.

20. Pools

It is not that convenient to create 100 threads or 100 processes. We have to create one thread, then the second, then the third, fourth, ..., 100 thread. Then we have to run these threads, join them, which is not always effective. Moreover, it is quite expensive to create and destroy threads and it is even more expensive to create and destroy process. Thanks to pools, we can easily and conveniently create, run as many threads or processes as we want (Holczer, 2020). Pools do not destroy threads or processes that have finished the operation, but will reuse them in other operations.

Let's have an example. If we have 5 threads and 5 tasks, using pools they will execute at the same time. We use submit function which executes given operation. There is a shutdown function which is similar to join function from previous section. The shutdown function makes sure that executor will not execute more operations. If function parameter wait=True, then executer will wait until all threads finished tasks (default value is true).

At the figure 18 we can see the pool. The pool size is 5 that means we create 5 threads. We call do_operation six times. What happens. If the pool size is equal to 1, the operations start in 1 thread sequentially. If the pool size is equal to 5, the operations start in 5 threads together with time-slicing. If the number of operations (6) is greater than the pool size (5), the executor will wait until any thread finished operation and assigns another operation to it. Operation shutdown waits until all operations (threads) finished. The right example shows the same program using context manager (no need to call shutdown operation, it is included in context manager).

Fig. 18, using pool for threads

```python
from concurrent.futures import ThreadPoolExecutor
import time

def do_operation(num1, num2):
    time.sleep(2)
    result = num1 * num2
    print(f"Operation finished: {result}")

executor = ThreadPoolExecutor(max_workers=5)

executor.submit(do_operation, 5, 6)
executor.submit(do_operation, 5, 8)
executor.submit(do_operation, 10, 2)
executor.submit(do_operation, 15, 4)
executor.submit(do_operation, 20, 3)
executor.submit(do_operation, 25, 8)

executor.shutdown()
```

```python
from concurrent.futures import ThreadPoolExecutor
import time

def do_operation(num1, num2):
    time.sleep(2)
    result = num1 * num2
    print(f"Operation finished: {result}")

with ThreadPoolExecutor(5) as executor:
    executor.submit(do_operation, 5, 6)
    executor.submit(do_operation, 5, 8)
    executor.submit(do_operation, 10, 2)
    executor.submit(do_operation, 15, 4)
    executor.submit(do_operation, 20, 3)
    executor.submit(do_operation, 25, 8)
```

If we want to use processes, we have to import module ProcessPoolExecutor instead of ThreadPoolExecutor. At the same time processes have to be created in main module (if __name__ == "__main__").

21. Semaphores

Semaphores are simple variables (or abstract data types) that are used for controlling access to common resources, for example if we have a common resource and more threads trying to work with that resource. Semaphore is a record of how many units of a particular resource are available, they do not store information about which of the resource is available. We have to wait until a unit of the resource becomes available again.
There are (Holczer, 2020):

- Counting semaphores: allows an arbitrary resource count
- Binary semaphores: semaphores that are limited to the values 0 and 1

Suitable for solving the problem of the producer and the consumer. Standard problem when programming concurrent processes. The producer produces certain data and stores it in a stack of limited size. The consumer is retrieving data from the stack. The producer cannot add more data if the stack is full, and the consumer cannot take the data if there is no data in the stack.

If we have a complex problem that is very expensive for available resources, for example computer vision (training database has over 1 million images). At first, it is very difficult to process this data, the second it is difficult to process machine learning on 1 device or on a few devices. The HPC (High Performance Computing) cluster can be solution. The HPC cluster has limited capacity, let's say no more than 5 users (1 user = 1 thread) can access the compute operation at a time. How can we solve the problem ? Using counting semaphores.

See this example at the figure 19. We simulate this scenario in compute operation. We initialize the semaphore size on 5. We create „infinite“ number of threads (1 thread = 1 user). Only 5 threads (users) at a time can access to common resource. If semaphore counter is less than the semaphore size, thread can access to common resource and internal counter goes up

+1 (calling operation acquire). If thread finished, semaphore counter goes down -1 (calling operation release) and another thread can access to common resource if semaphore counter is less than the semaphore capacity.

Fig. 19, using semaphore by Threads and Processes for accessing to common resource. Maximum 5 units can access it

```python
from threading import Thread, Semaphore
import random
import time

semaphore = Semaphore(5)
operation_counter = 0

def compute():
    global operation_counter
    semaphore.acquire()
    print(f"Executing operation in cluster...")
    operation_counter += 1
    time.sleep(random.randint(3, 8))
    print("Operation finished")
    semaphore.release()
    operation_counter -= 1

while True:
    time.sleep(0.1)
    t = Thread(target=compute)
    t.start()
```

```python
from multiprocessing import Process, Semaphore
import random
import time

def compute(semaphore):
    semaphore.acquire()
    print(f"Executing operation in cluster...")
    time.sleep(random.randint(3, 8))
    print("Operation finished")
    semaphore.release()

if __name__=="__main__":
    semaphore = Semaphore(5)
    while True:
        time.sleep(0.1)
        t = Process(target=compute, args=(semaphore,))
        t.start()
```

22. Parallel application example in Python

In this section we can see application in parallel fashion in Python. We want to create a parallel application that will open the URLs one by one using Processes and Pipes and will find the occurrence of the words: video, covid, sport, auto. List of word occurrences will print on the console. Process 1 (the main process) will open the URL and send the content of the page to Process 2. Process 2 will extract the searched words in the received data. Extracted data (found words) will send to Process 3. Process 3 will create a dictionary of searched words, then will increment value by +1 for each found word. Then Process 3 will send the output to Process 1, which will print it to the console. It is I/O bound operation (read data from URLs) and CPU bound operation as well (search and count words). See the example at the figure 20.

Fig. 20, parallel application using processes and pipes

```python
import certifi
import ssl
import json
import time
import re
from urllib.request import Request, urlopen
from multiprocessing import Process, Pipe

urls = ['https://www.aktuality.sk','https://dennikn.sk','https://plus7dni.pluska.sk','https://www.pravda.sk',
'https://www.sme.sk','https://sport.aktuality.sk','https://zive.aktuality.sk','https://www.nextech.sk',
'https://autobild.pluska.sk','https://www.stop.sk']

R_WORDS = "video|covid|sport|auto"
words = ['video', 'covid', 'sport', 'auto']

def parse_input(text_conn, data_conn):
    text = text_conn.recv()
    while text is not None:
        lines = text.lower()
        lines = lines.splitlines()
        s_data = []

        for line in lines:
            line = line.split()
            for word in line:
                if re.search(R_WORDS, word):
                    s_data.append(word)
        data_conn.send(s_data)
        text = text_conn.recv()

    data_conn.send(None)
```

```python
def count_words(data_conn, data_dist_conn):
    count = {}
    for word in words:
        count[word] = 0
    s_data = data_conn.recv()
    while s_data is not None:
        for value in s_data:
            if re.search(R_WORDS, value):
                index_w = re.search(R_WORDS, value).group()
                count[index_w] += 1
        s_data = data_conn.recv()

    data_dist_conn.send(count)

if __name__ == "__main__":
    text_conn_a, text_conn_b = Pipe()
    data_conn_a, data_conn_b = Pipe()
    data_dist_conn_a, data_dist_conn_b = Pipe()

    Process(target=parse_input, args=(text_conn_b, data_conn_a)).start()
    Process(target=count_words, args=(data_conn_b, data_dist_conn_a)).start()

    start = time.time()
    for url in urls:
        response = Request(url, headers={'User-Agent': 'XYZ/3.0'})
        text = str(urlopen(response, context=ssl.create_default_context(cafile=certifi.where())).read())
        text_conn_a.send(text)
    text_conn_a.send(None)
    data_dist = data_dist_conn_b.recv()

    end = time.time()
    print(json.dumps(data_dist, indent=4))
    print(f"Time taken {end - start}")
```

23. Conclusion

Our research shows that using Python programming language in parallel fashion is useful and makes a sense to think about it for our projects. Although, Python standard interpreter has implemented GIL, which limits our multithreaded program, we can get around GIL by implementing different Python interpreter (Jython, IronPython) or using processes instead of threads (each process has separate interpreter). GIL does have significant impact on the speed up CPU bound operations, but for I/O operations it does not. Python has many tools/options how we can do our program in parallel fashion from using threads, processes, join, to implementing pools, semaphores, etc. There are other options related with parallelism in Python, which were not described in this work, such as condition variables or using GPUs instead of CPUs. The choice is always up to us, but Python is a tool, which can make our program faster, more responsive and it makes better hardware utilization.

24. Resources

Ajitsaria, A (2018). What Is the Python Global Interpreter Lock (GIL)? Real Python. https://realpython.com/python-gil/

Amdahl's law. (2021). Microsoft Academic.

Amdahl's law. Microsoft academic. (n.d.). Retrieved December 29, 2021, from https://academic.microsoft.com/topic/114595137/publication/search?q=Amdahl%27s+law&qe=And%28Composite%28F.FId%253D114595137%29%252CTy%253D%270%27%29&f=&orderBy=0

Andrews G. R. (2000). Foundations of Multithreaded, Parallel, and Distributed Programming. Addison Wesley

Ben-Ari, M. (2006). Principles of Concurrent and Distributed Programming (2nd edition). Addison Wesley

Cutajar, J. (2021, March). *Learn parallel computing in Python*. Udemy. https://www.udemy.com/course/parallel-computing-in-python/.

Gustafson's law. (2021). Microsoft Academic.

Gustafson's law. Microsoft academic. (n.d.). Retrieved April 29, 2021, from https://academic.microsoft.com/topic/203799913/publication/search?q=Gustafson%27s+law&qe=And%28Composite%28F.FId%253D203799913%29%252CTy%253D%270%27%29&f=&orderBy=0

Hanák, J. (2013). Moderné paralelné programovanie (2nd edition). Vydavateľstvo EKONÓM

The PSF. (n.d.). *Multiprocessing - process-based parallelism*. multiprocessing - Process-based parallelism - Python 3.9.7 documentation. https://docs.python.org/3/library/multiprocessing.html.

The PSF. (n.d.). *Threading - thread-based parallelism*. threading - Thread-based parallelism - Python 3.9.7 documentation. https://docs.python.org/3/library/threading.html.

FEATURES OF THE USE OF GEOINFORMATION SYSTEMS IN HIGHER EDUCATION

Nursaule Karelkhan[1], Aknur Kadirbek[2]

Abstract
The use of geoinformation systems as teaching tools in the field of education is closely linked to interdisciplinary and Natural Sciences. Currently, the use of geoinformation systems in the educational process has great potential for higher education institutions. The article discusses the features of using software products for teaching geoinformation systems in higher education institutions.

Key words
geoinformatics, higher education, GIS, geoinformation systems

1. Introduction

In higher education several waves of e-learning development and deployment enthusiasm scudded in the last few years. These "waves" were caused mainly by the fact that there are not precisely defined frames of e-learning itself, because there are many definitions of e-learning and they capture the essence of the issue more or less. The following definition is used very often: "E-learning is an educational process, using information and communication technologies to create courses, to distribute a study content, to enable communication between students and teachers and the study management. ". When we think about this definition we can easily see that it takes us 20 years back in time. It is based on the classical model of teaching and the only real difference is that there are modernized channels of communication through which students can transmit information. The second definition may sound a little sci-fi, but we think that it better captures the essence of today's understanding of e-learning. "Man and machine are a symbiotic system, which is able to effectively learn and solve certain problems. The connection of man and machine is so powerful that man becomes dependent on technology and this interdependence will be even stronger in the coming years. " (Schmidt & Jurík, 2016). Different trends and tendencies in this field such as mobile-learning, responsive design, gamification, creation of personalized interactive courses and video-based learning are described by Jurík (Jurík, 2018).

The state program for the development of education in the Republic of Kazakhstan for 2020-2025 provides for equal access to the best educational resources and technologies for all participants in the educational process and the integration of education, science, industry, introduction of intellectual property and technology products in education (1. State program for the development of education and science in the Republic of Kazakhstan for 2020-2025: - Nur-Sultan, 2019).

Modern technologies are making major changes in the traditional work on the compilation of cartographic materials, their analysis and use. Geoinformation system is a system that allows you to provide all services in orientation in space and time with an information base at a qualitatively new level and solve a number of problems in the field of economic, technical, and natural sciences.

[1] L. N. Gumilyov Eurasian National University, Faculty of Computer Science, Nur-Sultan, Kazakhstan, knursaule@mail.ru

[2] L. N. Gumilyov Eurasian National University, Faculty of Computer Science, Nur-Sultan, Kazakhstan, kadirbek-aknur@mail.ru

The purpose of the article is to identify the features of using software products for teaching geoinformation systems. Currently, developed countries consider the development of GIS to be a matter of state significance.

One of the primary goals of the State GIS Program (State of Hawaii) is to improve overall efficiency and effectiveness in government decision-making. In support of this goal, participating State agencies are developing, maintaining and sharing their respective databases and applications. The centralized database enables agencies to share information while reducing the development of redundant databases, helps standardize the information being analyzed by decision makers and serves as a means for collecting and distributing the best available databases (Hawaii Statewide Gis Program, 2021).

A Geographical Information System (GIS) is a computer system for processing, storing, checking, integrating, manipulating, analysing and displaying data related to positions on the surface of the earth. It is then presented cartographically, graphically or as a report (Jebara, 2021).

Before geographic information systems (GIS) were introduced to the world, and to the world of academic geography, place and the uniqueness of place was considered all-important. In the 1950s and early 1960s the uniqueness of place was challenged by the rise of quantitative and theoretical approaches to geography. In the 1960s the rise of geographic information systems gradually led to an ascendancy of a spatial and scientific approach to geography. By the late 1980s, GIS were providing jobs for students who were educated and trained to use and develop these new software tools. However, throughout the decades since GIS were developed there were those in the discipline who celebrated the uniqueness of place and the value of qualitative geography and a plethora of nonquantitative and aspatial approaches to geography. This entry will show how, during the last two decades, GIS have attempted to accommodate and assimilate these approaches and how those working in GIS have come to respect both the uniqueness and similarities of places (Nigel, 2020)

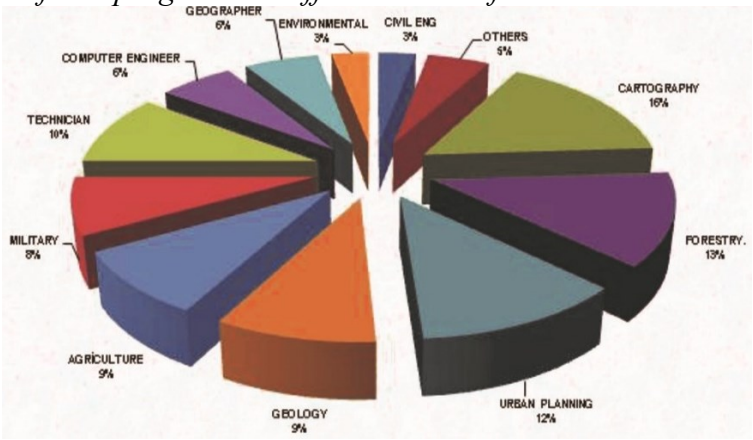
GIS has revolutionized the methods and dimensions of spatial analysis resulting in a dramatic change in the direction of Geography and become a major component in other disciplines with a spatial component (i.e., biology, political science, urban planning, geology, sociology etc.). It has also led to significant reforms in geographic education (Birkin, Clarke G, Clarke M.&Wilson, 1996).

The use of GIS in geographic education first began at the university level. The number of the programs, which were offering GIS lessons at the US and Canadian universities in the beginning of 1980s were approximately 10. By the end of the 1990s, this number exceeded 2000 (Zhou, Smith&Spinelli, 1999).

Currently, GIS has become an integral part of geography in many universities, colleges, and schools in the United States, Canada, and some European and Asian countries, some of which offer special degrees in GIS or geographic informatics at the undergraduate, graduate, and doctoral levels.

GIS is being used as a search tool in the level of universities, in more than 100 different academic disciplines. However, when we look at the GIS education, it can be observed that some disciplines come into prominence. Nowadays, in undergraduate and postgraduate level, the GIS education is given in the departments like geography, geodesy, photogrammetry, and also in ecological sciences, natural resources, forestry, civil engineering, landscape architecture, ecology, urban design and planning. In relation to the developments in computer technology, GIS has improved and the usage area has increased. The increase of software number, decrease in expenditures and usage flexibility in GIS applications caused to solve the problems of design and planning (Mustafa&Korucu, 2012) (Figure 1).

Fig.1 - Application of GIS programs in different areas of education



(Mustafa & Korucu, 2012)

Especially, in recent years, aiming to generalize the use of GIS in educational activities, many people, organizations, institutions and companies have done various studies and the result of these studies have started to be seen. For example, the contributions of GIS in the educational studies which were done at undergraduate level have realized and GIS started to be used in the secondary education institutions of the USA and Europe. This previously took place in yearly scheme of education in countries like U.S.A., Canada, Britain, Sweden, Denmark, Germany, Finland, and the Netherlands. Afterwards, together with geography courses, it was started to be used in courses like; science, chemistry, biology, mathematics, environmental sciences, and social sciences (Yiğit, Ataol & Dinç, 2011).

Let's focus on freely distributed GIS, which are increasingly used around the world, and this is due not only to their cost and open source, but also to the fact that their functionality has recently increased dramatically and some of them, for example QGIS, can compete in a number of aspects even with such a market leader as ESRI ArcGIS. A fairly quick change of versions also attracts the user, who can connect their own libraries or a set of calculation modules for specific tasks, forming a personal information and computing environment for processing spatial data on the basis of such systems. An important role in the development of open software was played by the Open Source Initiative consortium, which determined the compliance of the software license with open software standards (Shekar & Hiong, 2008).

Currently, in Kazakhstan, some universities in the field of education use geoinformation systems in the field of natural sciences, in the field of technical sciences, in the field of economic sciences.

From all the variety of pedagogical applications of the latest innovative technologies, the use of geographical information systems (GIS) should be highlighted in connection with their relevance in relation to geographical education and increasing popularity in the practice of domestic and foreign geographical educational process (Kreider, 2004).

Despite many years of experience in using a variety of GIS for training purposes, their potential capabilities are not fully used and not in full force. Geographical information systems (GIS) for educational purposes are usually intended for use in the educational process, in the preparation, retraining and advanced training of personnel in the field of education, in order to develop the student's personality, and to intensify the learning process (Petrova & Novenko, 2003).

The use of GIS allows you to simplify many functions, in particular, in education, development, information, propaganda, visual and figurative, and also instills skills when working with GIS. Currently, the effectiveness of the use of GIS technologies in the field of

education is becoming increasingly important and relevant, as it contributes to the disclosure of personal qualities of each student.

GIS is very effective to use for quality analysis in the education system in all regions and cities. Such an analysis should be carried out by public authorities on a systematic basis in order to ensure effective and affordable education for the local population, along with monitoring the quality of education. And each educational institution can use it to find new markets for itself and make a decision about opening new professions.

The use of GIS in the field of education, in a word, makes it possible to think, participate in research and solve problems, as well as to display spatial patterns, connections and relationships.

The most popular GIS in education is represented by various software products: ArcGIS, MapInfo, Panorama, Neva, Erdas Imagine, Geo Media, etc., But the most common is Arc GIS, which consists of separate modules responsible for separate operations and functions on spatial and thematic data: Arc Catalog, Arc Globe, Arc Reader, Arc Editor, Arc Map, Arc Tools, Arc Info, combined in a single environment Arc GIS, which we offer for students to study at the Speciality of Natural Sciences.

In addition, an analysis of the best world and Republic of Kazakhstan higher education institutions by QS ranking using GIS technologies was conducted and an indicator of the programs they use was released (Tab 1. and Fig 2.).

Tab. 1- Programs for teaching GIS in Higher education institutions

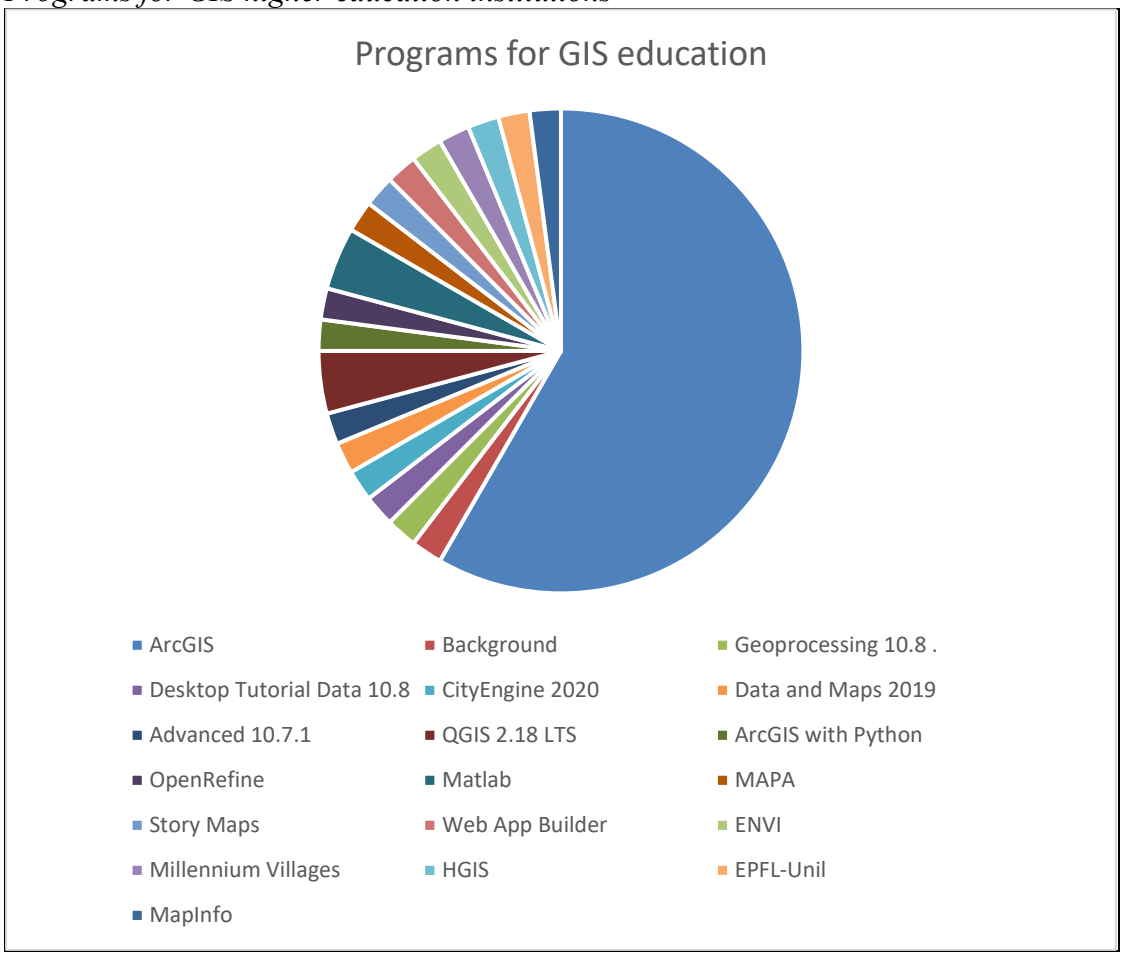
№	Higher education institutions	Programs for teaching GIS
1	Massachusetts Institute of Technology	- ArcGIS; - Background Geoprocessing 10.8; - Desktop Tutorial Data 10.8; - CityEngine 2020; - Data and Maps 2019.
2	Stanford University	- ArcGIS; - Advanced 10.7.1; - QGIS 2.18 LTS; - OpenRefine; - Matlab.
3	Harvard University	- MAPA; - ArcGIS.
4	University of Oxford	- ArcGIS Collector; - ArcMap; - ArcGIS; - Story Maps; - Web App Builder.
5	Imperial College London	- ArcGIS; - ENVI.
6	Columbia University	- ArcGIS; - QGIS; - Millennium Villages.
7	EPFL	- HGIS; - EPFL-Unil.
8	The University of Edinburgh	- ArcGIS.
9	The University of Hong Kong	- ArcGIS.

10	University of Toronto	- ArcGIS.
11	Peking University	- ArcGIS Online.
12	The University of Tokyo	- ArcGIS.
13	The University of Manchester	- ArcGIS; - SimaPro; MATLAB; - R and IES Virtual Environment.
14	The Australian National University	- ArcGIS Pro.
15	Seoul National University	- ArcGIS Desktop.
16	Technical University of Munich	- ArcGIS Desktop ArcGIS Pro; - ArcGIS environment.
17	Al-Farabi Kazakh National University	- ArcGIS.
18	L.N. Gumilyov Eurasian National University	- ArcGIS; - MapInfo .

We found out that GIS universities use different programs in education. In particular, it is established that the ArcGIS software packages are used very often (Figure 2.).

L.N. Gumilyov Eurasian National University uses the ArcGIS program in the educational process.

Fig. 2- Programs for GIS higher education institutions



ArcGIS is a geographic information system (GIS) for working with maps and geographic information maintained by the Environmental Systems Research Institute (Esri). It is used for creating and using maps, compiling geographic data, analyzing mapped information, sharing and discovering geographic information, using maps and geographic information in a range of

applications, and managing geographic information in a database. The system provides an infrastructure for making maps and geographic information available throughout an organization, across a community, and openly on the Web. ArcGIS consists of the following Windows desktop software:

ArcReader, which allows one to view and query maps created with the other ArcGIS products;

ArcGIS Desktop (often referred to as "ArcMap" to distinguish it from ArcGIS Pro), made up of four fundamental applications:

ArcMap, for viewing and editing spatial data in two dimensions and creating two-dimensional maps;

ArcScene, for viewing and editing three-dimensional spatial data in a local projected view;

ArcGlobe, for displaying large, global 3D datasets;

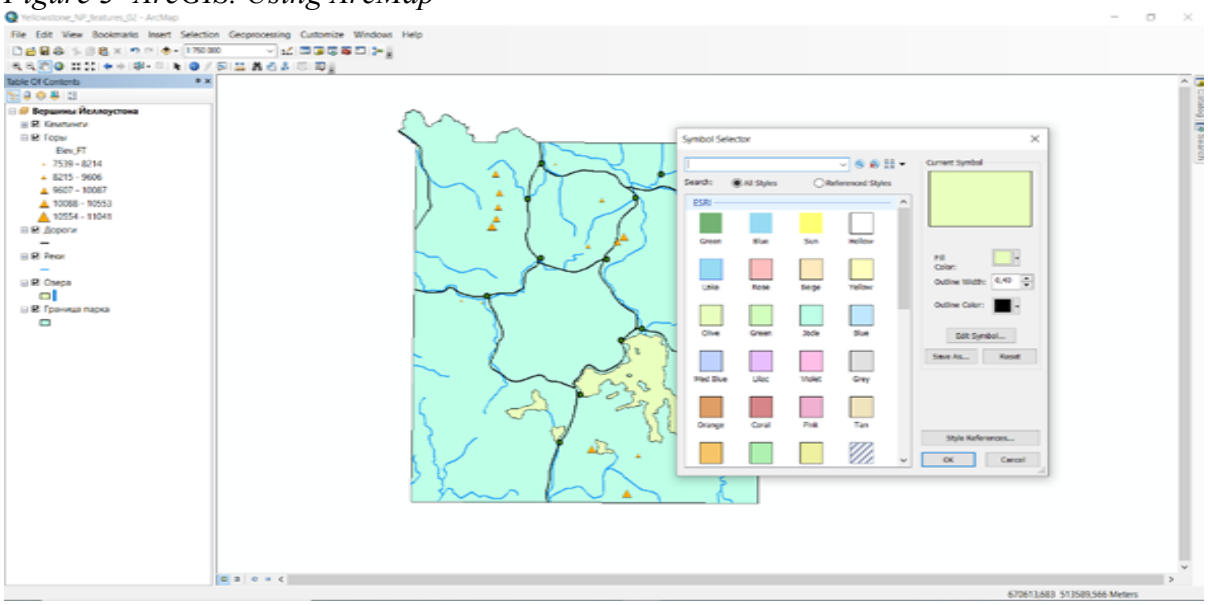
ArcCatalog, for GIS data management and manipulation tasks (Petrova&Novenko, 2003).

Currently, the L.N. Gumilyov Eurasian National University uses the software product ArcGIS 9.1 with additional modules Spatial Analyst, 3D Analyst, Geostatistical Analyst.

The greatest interest in studying GIS is shown by students of the Faculty of Natural Sciences and the Faculty of Architecture and Construction of L.N. Gumilyov Eurasian National University, and especially those who study in the specialties "geography", "ecology" and" geodesy and cartography". The attention of future land managers to this technology is dictated by the introduction of the "Automated Information System of the State Land Cadastre" (AIS GZK) in the Republic and its expected development to the "land information systems" (ZIS) of Kazakhstan. Among future ecologists, this interest is also related to the fact that at the beginning of the century, the Ministry of Environmental Protection and Natural Resources of the country signed a decree on the mandatory conduct of all environmental works related to spatial objects only on the basis of GIS.

Today, on the basis of the faculties of L.N. Gumilyov Eurasian National University, through the efforts of departments, classes of "Geoinformation systems" and "AIS GZK"are organized. Disciplines "Geoinformation systems", "GIS and digital cartography "and" AIS GZK " began to teach on the basis of software products of the ArcGIS family. We are currently considering using ArcGis to study the subject "Digital technologies by industry" (Figure 3).

Figure 3- ArcGIS. Using ArcMap



As a result of the survey, it was found that students have a lack of IT competence in studying this program. Therefore, when studying the discipline "Information Technologies by industry", students must first learn the basics of GIS.

2. Conclusion

In the Republic of Kazakhstan, the number of educational programs in universities that can provide qualified GIS education is calculated in units. The name beginning with "Theoretical and practical fundamentals of GIS application in the training of digital technology in the field of natural Sciences", has great importance here: education in GIS has become a prerogative of the geological and geographical faculties. It is clear that these faculties are not taught by IT specialists. This approach to learning deprives GIS technologies of their main strength and value - the ability to comprehensively process information from different fields of science. These latter tasks require specialists to have a good knowledge of mathematics (probability theory, mathematical statistics, operations research, modeling), programming skills to create custom applications, and database design skills, which usually have a very complex structure for geocoordinated data. For example, the ArcGis program uses the Python programming language. For students of the Faculty of Natural Sciences, these tasks that go beyond their profession and interests are simply impossible and also it is known that the training of GIS requires the competence of information technology specialists. In the educational programs of the Faculty of Natural Sciences, we must first form the basis for the application of GIS.

Therefore, in teaching the discipline "Information Technologies by industry", it is especially important for students to first learn the basics of GIS.

The study of the basics of the theory and practice of geoinformation systems, depending on the specialization, is included in the number of general education disciplines "Information Technologies by industry".

3. Resources

Birkin, M., Clarke, G., Clarke, M., & Wilson, A.(1996) *Intelligent GIS: Location decisions and strategic planning* (1st ed.). Cambridge: Wiley. 292.

K. Ben Jebara *(2007)* The role of geographic information system (gis) in the control and prevention of animal diseases. *Conf. OIE 2007, 175-183.*

Jurík, P. (2018). Ongoing trends in e-learning. Elektronnaja Kazaň – 2018 International scientific-practical conference, 11–18.

Hawaii statewide GIS Program. (n.d.). Retrieved December 28, 2020, from https://planning.hawaii.gov/gis

Legal information system of Regulatory Legal Actsof the Republic of Kazakhstan, (2019), *Retrieved December 28, 2020, from http://adilet.zan.kz/kaz/docs/P1900000988*

Korucu, M. G. (2012). GIS and types of Gis education programs. Procedia - Social and Behavioral Sciences, 46, 209-215. doi:10.1016/j.sbspro.2012.05.095

Waters, N. (2018). Gis : History. International Encyclopedia of Geography, 1-13. doi:10.1002/9781118786352.wbieg0841.pub2

Petrova & Novenko (2003) Complex application of interactive and audiovisual teaching aids in geography lessons in basic school. *Informatics and Education, vol.1*, 10-13.

Shekhar, S., Xiong, H., & Zhou, X. (2017). Encyclopedia of GIS. United States: Springer.

Schmidt. P – Jurík, P. (2016). Use of e-Learning in Higher Education. Elektronnaja Kazaň - 2016. International scientific-practical conference. 5 – 12.

Yigit, I, Ataol, M, Dinc, A, & Dinc, A. (2013). Gis education in geography departments and the need for gis. *Marmara Journal of Geography, 0 (24), 312-331 . Retrieved from https://dergipark.org.tr/en/pub/marucog/issue/471/3842*

Zhou, Y., Smith, B.W. & Spinelli, G. (1999) Impacts of increased student career orientation on American college geography programmes. *Journal of Geography in Higher Education*, 23(2), 157-165.

SEARCHING FOR CONFIGURATION AND PRESET DATA OF MOTION STATEMENTS OF A MAIN CONTROL PROGRAM OF THE KUKA INDUSTRIAL WELDING ROBOT IN A ROBOT CONFIGURATION FILES

Igor Košťál[1]

Abstract
During testing of a given robot program the operator needs to modify some motion statements or to add new statements into the robot program. In this case, he must modify or set the motion data set parameters (velocity, acceleration, etc.) of an existing or a new motion statement and the parameters of the tool and base (workpiece) coordinate system, the interpolation mode etc. of these statements. The values of the majority of the motion statement motion data set parameters are the percentage of the preset configuration data values. If the robot operator does not recognize these preset configuration data, then he must search for them in source files of subprograms and in configuration files of a robot main control program to know on what values he sets particular parameters of this motion data set. We created a .NET search application that for each type of motion statements of the robot main control program searches for configuration data in the configuration files that contain the default values of the motion data set parameters of these statements. This application also searches for preset structure variables defining the default motion data sets and sets that define the tool and base (workpiece) coordinate system, the interpolation mode etc. of these statements. In this way, our .NET search application can be a very effective software support for the operator during testing of a given robot program.

Key words
industrial robot, motion statements, motion data set, configuration data, testing robot

1. Introduction

Testing of the precision of the TCP (Tool Center Point) (KUKA, 2005) (KUKA, 2010) of a robot tool motion through particular points of the TCP programmed path is also a part of the development of a robotic workstation consisting of several robots and manipulators. The robot operator who performs these tests repeatedly is often not the author of a robot control program. During this testing the operator needs to modify some motion statements or to add new statements into the robot program. In this case, as we mentioned it above, he must modify or set the motion data set parameters of an existing or a new motion statement and the parameters of the tool and base (workpiece) coordinate system, the interpolation mode etc. of these statements. The values of the majority of the motion data set parameters are the percentage of the preset configuration data values. If the robot operator does not recognize these preset configuration data, then he must search for them in source files of subprograms and in configuration files of a robot main control program to know on what values he sets particular parameters of this motion data set. The structure of robot subprograms may be complicated and the operator may not know it in detail if he is not the author of the robot program. Therefore,

[1] University of Economics in Bratislava, Faculty of Economic Informatics, Department of Applied Informatics, Dolnozemská cesta 1, 852 35 Bratislava, Slovakia, e-mail: igor.kostal@euba.sk

finding this structure and also searching for configuration data in it and in large configuration files is a very time consuming work.

As we mentioned above, we created a .NET search application that for each type of motion statements of the robot main control program searches for configuration data in the configuration files that contain the default values of the motion data set parameters of these statements. This application also searches for preset structure variables defining the default motion data sets and sets that define the tool and base (workpiece) coordinate system, the interpolation mode etc. of these statements. The paper deals with our .NET search application outputs that this application provides to the robot operator during retesting of the precision of the TCP of the welding robot tool motion.

1 LIN and PTP motions

Our .NET search application has only found two types of motion statements LIN and PTP in a main control program of a robot, therefore, we describe them shortly.

LIN motion (KUKA, 2002) (KUKA, 2011)

In the case of LIN motions, the controller calculates a straight line equation from the current position to the target position specified in the LIN statement. The robot is moved to the end point via auxiliary points, which are calculated and executed at intervals of one interpolation cycle.

The velocities and accelerations and the system variables $TOOL and $BASE must also be programmed for linear motions. The system variables $VEL for the path velocity and $ACC for the path acceleration are available for defining the velocities and accelerations.

PTP motion (KUKA, 2002) (KUKA, 2011)

The point-to-point motion (the PTP statement) offers the quickest way of moving the robot arm from the current position to a programmed target position. The axes are moved in a synchronized manner. The maximum velocity and maximum acceleration must be programmed separately for each axis. The system variables $VEL_AXIS[*No*] for the axis-specific velocity, and $ACC_AXIS[*No*] for the axis-specific acceleration are available. All of the values are specified as a percentage of a maximum that is defined in the machine data.

2 A .NET application searching for configuration and preset data of motion statements of a main control program of the KUKA industrial welding robot

Our .NET search application was developed in the C# language in the development environment Microsoft Visual Studio 2019 Enterprise for the Microsoft .NET Framework version 4 and for the Microsoft operating systems Windows 10. This application searching for motion statements, their types and their numbers in the *csrb325* main control program of a robot. Then this application searches for configuration and preset data in configuration files of a robot for found types of motion statements of this main control program. The KRL (KUKA Robot Language) program *csrb325* is the part of the KUKA industrial welding robot control program, together with four service KRL programs. Using the *csrb325* program, the robot carries out welding a given part of a cooling system radiator bracket of a passenger car. Immediately after the start-up our .NET search application attempts to connect through the Intranet to the robot control computer and searches for the directory with DAT and SRC files of the *csrb325* program and the directories with configuration files on its hard drive. When the

.NET search application is connected to the correct directories on the robot control computer hard drive, then it is ready to full use.

When the .NET search application was searching for the configuration and preset data of motion statements of the *csrb325* main control program of a robot in configuration files of this robot it had to search for the following system files in the directory "KRC:\R1\MADA\" (KUKA, 2003):

- **$machine.dat** - is a predefined data list with exclusively predefined system variables for adapting the controller and the robot.
- **$robcor.dat** - contains robot-specific data for the dynamic model of the robot.

and the system file and the subprogram in the directory "KRC:\R1\SYSTEM\" (KUKA, 2003):

- **$config.dat** - is a system data list with general configuration data.
- **bas.src** - a subprogram containing a basic package for motion control.

All these files are generated when the KUKA System Software (KSS) is installed and are then always available.

Our .NET search application provides for the *csrb325* main control program of a robot an output to the user that includes:

- Types and the numbers of motion statements used in the *csrb325* program
- Complete logs of all LIN and PTP statements found in the *csrb325* program
- The configuration data with default (maximum) values of the motion data set parameters for LIN and PTP statements
- The preset structure variables for LIN and PTP statements: the default motion data sets and the default sets with adjusted the tool and base (workpiece) coordinate system, the interpolation mode etc.

The .NET search application also stores the whole this output data into the disk file with the name, for example, *Config_PresetDataOf_csrb325_20210517_112209.dtx* containing the date and time when the file was created (2021-05-17 11:22:09).

Fig. 2: The first part of the file Config_PresetDataOf_csrb325_20210517_112209.dtx with found configuration and preset data of the csrb325 program motion statements (the .NET search application displays the same output in its text box)

```
The results of searching for configuration and preset data of the csrb325 program motion
statements (from 2021-05-17 11:22:09Z)

Types/the numbers of motion statements used in the csrb325 program: LIN/2, PTP/41

    The list of all LIN statements found in the csrb325 program:

 1  ;FOLD LIN P7 Vel=0.5 m/s CPDAT2 Tool[2]:Greifer Base[0];%{PE}%R 8.2.24,
    %MKUKATPBASIS, %CMOVE, %VLIN,%P 1:LIN, 2:P7, 3:, 5:0.5, 7:CPDAT2
 2  $BWDSTART=FALSE
 3  LDAT_ACT=LCPDAT2 [DECL LDAT LCPDAT2={VEL 2.0,ACC 100.0,APO_DIST 100.0,
    APO_FAC 50.0, AXIS_VEL 100.0,AXIS_ACC 100.0,ORI_TYP #VAR,CIRC_TYP
    #BASE,JERK_FAC 50.0,GEAR_JERK 50.0, EXAX_IGN 0}]
 4  FDAT_ACT=FP7 [DECL FDAT FP7={TOOL_NO 2,BASE_NO 0,IPO_FRAME #BASE,
                    POINT2[] " ",TQ_STATE FALSE}]
 5  BAS(#CP_PARAMS,0.5)
 6  LIN XP7
 7  ;ENDFOLD

 8  ;FOLD LIN P9 CONT Vel=0.3 m/s CPDAT4 Tool[2]:Greifer Base[0];%{PE}%R 8.2.24,
    %MKUKATPBASIS, %CMOVE,%VLIN,%P 1:LIN, 2:P9, 3:C_DIS, 5:0.3, 7:CPDAT4
 9  $BWDSTART=FALSE
10  LDAT_ACT=LCPDAT4 [DECL LDAT LCPDAT4={VEL 2.0,ACC 100.0,APO_DIST 100.0,
    APO_FAC 50.0, AXIS_VEL 100.0,AXIS_ACC 100.0,ORI_TYP #VAR,CIRC_TYP
    #BASE,JERK_FAC 50.0,GEAR_JERK 50.0, EXAX_IGN 0}]
11  FDAT_ACT=FP9 [DECL FDAT FP9={TOOL_NO 2,BASE_NO 0,IPO_FRAME #BASE,
    POINT2[] " ",TQ_STATE FALSE}]
12  BAS(#CP_PARAMS,0.3)
13  LIN XP9 C_DIS
14  ;ENDFOLD

    The list of all PTP statements found in the csrb325 program:

15  ;FOLD PTP HOME  Vel= 100 % DEFAULT;%{PE}%MKUKATPBASIS,%CMOVE,%VPTP,%P
    1:PTP, 2:HOME, 3:, 5:100, 7:DEFAULT
16  $BWDSTART = FALSE
17  PDAT_ACT=PDEFAULT [DECL PDAT PDEFAULT={VEL 100.0,ACC 100.0,APO_DIST
                                          100.0,GEAR_JERK 50.0}]
18  FDAT_ACT=FHOME [DECL FDAT FHOME={TOOL_NO 1,BASE_NO 0,IPO_FRAME
    #BASE,POINT2[] " ", TQ_STATE FALSE}]
19  BAS (#PTP_PARAMS,100)
20  $H_POS=XHOME
21  PTP XHOME
22  ;ENDFOLD
                                          . . .
```

Source: (author)

Fig. 2: The second part of the file Config_PresetDataOf_csrb325_20210517_112209.dtx with found configuration and preset data of the csrb325 *program motion statements (the .NET search application displays the same output in its text box)*

The config. data with default (max.) values of the motion data set parameters for LIN statements:

23 DECL LDAT LCPDAT*X*={VEL *vel_v*,ACC *acc_v*,APO_DIST *apod_v*,APO_FAC *apof_v*, AXIS_VEL *axisv_v*, AXIS_ACC *axisa_v*,ORI_TYP #VAR,CIRC_TYP #BASE,JERK_FAC *jerkf_v*,GEAR_JERK *gearj_v*, EXAX_IGN *exaxi_v*}
24 *vel_v*: not used value in this *csrb325* program. The *$VEL.CP* path velocity value (unit [m/s]) is set by the BAS subprogram (using the second parameter in the BAS(#CP_PARAMS, *vel_value*) call). The maximum path velocity is *$VEL_MA.CP = 3.0 m/s*.
25 *acc_v* = % of the *DEF_ACC_CP=2.29999995 m/sec²* (the maximum path acceleration is *$ACC_MA.CP = 10.0 m/sec²*)
26 *apod_v*: the approximation parameter value - the distance parameter value in [mm]
27 *apof_v*: not used value in this *csrb325* program.
28 *axisv_v*: not used value in this *csrb325* program.
29 *axisa_v*: not used value in this *csrb325* program.
30 *jerkf_v*: the value (unit [%]) is used to calculate the jerk limitation RET_JERK.CP (unit [m/sec³]) for the LIN motion. The RET_JERK.CP is a percentage of the *$JERK_MA.CP = 500.0 m/sec³* (the maximum value).
31 *gearj_v* = % of the *$DYN_DAT[200 ... 205] = 1000.0 deg/*sec³
32 *exaxi_v* = 0: external axis is ignored; or 1: external axis is taken into consideration.

The preset structure variables for LIN statements:

the default motion data set:
33 DECL LDAT LDEFAULT={VEL 2.0,ACC 100.0,APO_DIST 100.0,APO_FAC 50.0,AXIS_VEL 100.0,AXIS_ACC 100.0,ORI_TYP #VAR,CIRC_TYP #BASE,JERK_FAC 50.0,GEAR_JERK 50.0,EXAX_IGN 0}
the default set with adjusted the tool and base (workpiece) coordinate system, the interpolation mode etc.:
34 DECL FDAT FDEFAULT={TOOL_NO 1,BASE_NO 0,IPO_FRAME #BASE,POINT2[] " ", TQ_STATE FALSE}

The config. data with default (max.) values of the motion data set parameters for PTP statements:

35 DECL PDAT PPDAT*XX*={VEL *vel_v*, ACC *acc_v*, APO_DIST *apod_v*, APO_MODE #CDIS, GEAR_JERK *gearj_v*}
36 *vel_v* = % of the *$VEL_AXIS_MA[1 ... 6] = 4500.0 1/min*
37 *acc_v*: the value (unit [%]) is used to set the acceleration of the robot axes *$ACC_AXIS[1 ... 6]* (unit [%]) (the maximum axis acceleration *$RAISE_TIME = 1000.0 Grad/sec²*)
38 *apod_v*: the approximation parameter value - the approximation distance value in [%]
39 *gearj_v* = % of the *$DYN_DAT[200 ... 205] = 1000.0 deg/sec³*

The preset structure variables for PTP statements:

the default motion data set:
40 DECL PDAT PDEFAULT={VEL 100.0,ACC 100.0,APO_DIST 100.0,APO_MODE #CDIS, GEAR_JERK 50.0}
the default set with adjusted the tool and base (workpiece) coordinate system, the interpolation mode etc.:
41 DECL FDAT FDEFAULT={TOOL_NO 1,BASE_NO 0,IPO_FRAME #BASE,POINT2[] " ", TQ_STATE FALSE}

Source: (author)

Description of some important found configuration, preset and other data of the *csrb325* program motion statements:

Line 1 - the beginning of the FOLD section. It is interpreted as a normal comment.
Line 2 - the statement $BWDSTART=FALSE is only relevant for the SCAN method, in which the program is interpreted backwards from the current position of the program interpreter.
Line 3 - the LCPDAT2 variable of a structure type LDAT is inserted into the LDAT_ACT variable of the same structure type. Our .NET search application found the full log of this LCPDAT2 variable related to the LIN XP7 statement in the *csrb325.dat* file and wrote it into this line of the output file *Config_PresetDataOf_csrb325_20210517_112209.dtx*. Now, the LDAT_ACT variable contains the motion data set (velocity, acceleration, etc.) of the LIN XP7 statement.
Line 4 - the FP7 variable of a structure type FDAT is inserted into the FDAT_ACT variable of the same structure type. Our .NET search application found the full log of this FP7 variable related to the LIN XP7 statement in the *csrb325.dat* file and wrote it into this line of the output file *Config_PresetDataOf_csrb325_20210517_112209.dtx*. Now, the FDAT_ACT variable contains the parameters of the tool and base (workpiece) coordinate system, the interpolation mode etc. of the LIN XP7 statement.
Line 5 - the BAS(#CP_PARAMS,0.5) subprogram call. For the LIN XP7 statement this subprogram sets related system variables according to values of components of the variables LDAT_ACT and FDAT_ACT.
Line 6 - the **LIN XP7** statement that **executes** a **linear motion** to the end point **XP7**.
Line 7 - the end of the FOLD section.
Line 17 - the **PDEFAULT** variable (**the default motion data set**) of a structure type PDAT is inserted into the PDAT_ACT variable of the same structure type. Our .NET search application found the full log of this PDEFAULT variable related to the PTP XHOME statement in the *csrb325.dat* file and wrote it into this line of the output file *Config_PresetDataOf_csrb325_20210517_112209.dtx*. Now, the PDAT_ACT variable contains the motion data set (velocity, acceleration, etc.) of the PTP XHOME statement.
Line 18 - the FHOME variable of a structure type FDAT is inserted into the FDAT_ACT variable of the same structure type. Our .NET search application found the full log of this FHOME variable related to the PTP XHOME statement in the *csrb325.dat* file and wrote it into this line of the output file *Config_PresetDataOf_csrb325_20210517_112209.dtx*. Now, the FDAT_ACT variable contains the parameters of the tool and base (workpiece) coordinate system, the interpolation mode etc. of the PTP XHOME statement.
Line 19 - the BAS (#PTP_PARAMS,100) subprogram call. For the PTP XHOME statement this subprogram sets related system variables according to values of components of the variables PDAT_ACT and FDAT_ACT.
Line 21 - the **PTP XHOME** statement that **executes** a **point-to-point motion to** the end point **XHOME**. This XHOME position is used as the first and last position in the *csrb325* program because the first motion instruction in a KRL program must define an unambiguous starting position. The XHOME position, which is stored by default in the robot controller, ensures that this is the case.
Line 23 - the declaration of the LCPDATX variable of a structure type LDAT with the components VEL,ACC,APO_DIST,APO_FAC,AXIS_VEL,AXIS_ACC,ORI_TYP, CIRC_TYP,JERK_FAC,GEAR_JERK and EXAX_IGN and their formal initialization.
Line 24 - the LCPDATX.*VEL* value is not used in this *csrb325* program. The *$VEL.CP* path velocity value (unit [m/s]) is set, for example, to 0.5 m/s in the LIN XP7 statement, by the BAS subprogram (using the 2nd parameter in the BAS call).

Line 25 - the LCPDATX.*ACC* value (unit [%]) is a percentage of the default path acceleration DEF_ACC_CP=2.29999995 m/sec^2 that is defined in the *$config.dat* file. The *$ACC.CP* path acceleration value (unit [m/sec^2]) is set to *$ACC.CP* = DEF_ACC_CP * LCPDATX.*ACC*/100.0 by the BAS subprogram.

Line 26 - the LCPDATX.*APO_DIST* distance parameter value (unit [mm]) is an approximation parameter value. Approximation starts, at the earliest, when the distance to the end point falls below the value of *$APO.CDIS* (= *LDAT_ACT.APO_DIST*).

Line 27 - the LCPDATX.*APO_FAC* value is not used in this *csrb325* program. The BAS subprogram of the *csrb325* program has no reference to this value.

Line 28 and 29 - the LCPDATX.*AXIS_VEL* and LCPDATX.*AXIS_ACC* values are not used in this *csrb325* program. The velocities and accelerations of linear motions no longer refer to the motor speed of each axis (LCPDATX.*AXIS_VEL* and LCPDATX.*AXIS_ACC*) but to the motion of the TCP. Therefore, the system variables *$VEL* for the path velocity and *$ACC* for the path acceleration are available for defining them (see Line 24 and Line 25). The LCPDATX.*AXIS_VEL*, LCPDATX.*AXIS_ACC* and LCPDATX.*APO_FAC* values are only there due to compatibility with older the KUKA System Software (KSS).

Line 30 - the LCPDATX.*JERK_FAC* value (unit [%]) is used to calculate the jerk limitation RET_JERK.CP (unit [m/sec^3]) for the LIN motion by the SJERK function that is defined in the *bas.src* file. The RET_JERK.CP: Change to the path acceleration in [m/sec^3]. It is a percentage of the system variable *$JERK_MA.CP* = 500.0 m/sec^3 (the maximum value) that is defined in the *$machine.dat* file.

Line 31 - the LCPDATX.*GEAR_JERK* value (unit [%]) is used to calculate the gear jerk of the axes *$GEAR_JERK[1 ... 6]* (unit [%]) by the SJERK function that is defined in the *bas.src* file. The gear jerk is specified as a percentage value of the corresponding machine data *$DYN_DAT[200 ... 205]* = 1000.0 deg/sec^3 (the maximum values) that is defined in the *$robcor.dat* file.

Line 32 - the LCPDATX.*EXAX_IGN* value is used to ignore external axis end positions for spline motions (0: External axis is ignored; 1: External axis is taken into consideration).

Line 35 - the declaration of the PPDAT*XX* variable of a structure type PDAT with the components VEL,ACC,APO_DIST,APO_MODE and GEAR_JERK and their formal initialization.

Line 36 - the PPDAT*XX.VEL* value (unit [%]) is used to set the velocity of the robot axes *$VEL_AXIS[1 ... 6]* (unit [%]) by the BAS subprogram. This variable contains the programmed axis velocity as a percentage of the maximum axis velocity *$VEL_AXIS_MA[1 ... 6]* = 4500.0 1/min defined in the *$machine.dat* file.

Line 37 - the PPDAT*XX.ACC* value (unit [%]) is used to set the acceleration of the robot axes *$ACC_AXIS[1 ... 6]* (unit [%]) by the BAS subprogram. This variable contains the planned axis acceleration as a percentage of the maximum axis acceleration *$RAISE_TIME* = 1000.0 Grad/sec^2 defined in the *$machine.dat* file.

Line 38 - the PPDAT*XX.APO_DIST* approximation distance value (unit [%]) for PTP motions. It is the furthest distance before the end point at which approximate positioning can begin. Maximum distance 100%: half the distance between the start point and the end point relative to the contour of the PTP motion without approximate positioning. The system variable *$APO.CDIS* is set to *$APO.CDIS = PDAT_ACT.APO_DIST*.

Line 39 - the PPDAT*XX.GEAR_JERK* value (unit [%]) is used to calculate the gear jerk of the axes *$GEAR_JERK[1 ... 6]* (unit [%]) by the SGEAR_JERK function that is defined in the *bas.src* file. The gear jerk is specified as a percentage value of the corresponding machine data *$DYN_DAT[200 ... 205]* = 1000.0 deg/sec^3 (the maximum values) that is defined in the *$robcor.dat* file.

All information from the output of our .NET search application displayed in Fig. 1 and Fig. 2 and also stored in the disk file with the name, for example, *Config_Preset DataOf_csrb325_20210517_112209.dtx* can be very useful for the robot operator before retesting of a robot especially when he needs to modify some motion statements or add new statements into the *csrb325* robot program.

3 Conclusion

The most valuable outputs that our .NET search application provides to the robot operator after carrying out testing of the precision of the TCP of the KUKA industrial welding robot tool motion through particular points of the TCP programmed path in a given robot program, are as follows:

- Complete logs of all LIN and PTP statements found in the *csrb325* program with assigned the motion data sets and the sets with adjusted the tool and base (workpiece) coordinate system, the interpolation mode etc. to particular statements (Lines 3, 4, 10, 11, 17 and 18 in Fig. 1 and Fig. 2)
- The configuration data with default (maximum) values of the motion data set parameters for LIN and PTP statements (Lines 23 ... 32 and 35 … 39 in Fig. 1 and Fig. 2)
- The preset structure variables for LIN and PTP statements: the default motion data set and the default set with adjusted the tool and base (workpiece) coordinate system, the interpolation mode etc. (Lines 33, 34, 40 and 41 in Fig. 1 and Fig. 2).

In such output information, the robot operator has precise values of configuration and preset data of a given robot. He does not have to search for them in different configuration files and can immediately modify required motion statements or add new statements into the robot program. In this way, our .NET search application makes the modification and creation of new motion statements faster and makes this process more effective during testing of the precision of the TCP of a robot tool motion by the operator.

4. Resources

KUKA Roboter GmbH (2002). SOFTWARE KR C2 / KR C3 Reference Guide Release 4.1. KUKA Roboter GmbH, Augsburg, Germany.

KUKA Roboter GmbH (2003). SOFTWARE KR C2 / KR C3 Expert Programming KUKA System Software (KSS) Release 5.2. KUKA Roboter GmbH, Augsburg, Germany.

KUKA Roboter GmbH (2005). SOFTWARE KR C2 / KR C3 Configuration KUKA System Software (KSS) Release 5.2. KUKA Roboter GmbH, Augsburg, Germany.

KUKA Roboter GmbH (2010). KUKA System Software 5.5 Operating and Programming Instructions for System Integrators. KUKA Roboter GmbH, Augsburg, Germany.

KUKA Roboter GmbH (2011). KUKA System Software 8.2 Operating and Programming Instructions for End Users. KUKA Roboter GmbH, Augsburg, Germany.

A PARALLELIZATION OF INSTANCE METHODS OF A .NET APPLICATION THAT COMPUTE VALUES OF DEFINITE INTEGRALS OF FUNCTIONS USING RECTANGULAR NUMERICAL INTEGRATION RULES AND THAT USE DELEGATES

Igor Košťál[1], Pavol Sojka[2]

Abstract
In this paper we deal with the parallelization of instance methods of a .NET application, which use rectangular numerical integration rules to calculate the values of definite integrals of mathematical functions $f_1(x) = e^{2x}/x$ and $f_2(x) = sin(3x)/x$, which do not have antiderivatives composed of elementary functions. Numerical integration is one of possibilities to calculate the values of their definite integrals. The .NET application uses nine member integration methods in which three rectangular integration rules, the left-, right-, and midpoint rules are implemented. Member integration methods are created universally that means they expect a method number with an implemented mathematical function in the parameter, by which they create a delegate referring to the method identified by the numerical parameter. Member integration methods calculate the value of a definite integral of a mathematical function that is implemented in the identified method using this identified method. For each of the three rectangular integration rules, a serial method, a parallel method, and a parallel method using a partitioner are created. By comparing the execution times of particular methods, we were examining which of them is the most efficient for a given rectangular integration rule. We assume that this should be a parallel method using a partitioner for each rectangular integration rule.

Key words
parallelization of a method, rectangular integration rules, antiderivative, delegate

1. Introduction

The value of a definite integral of a given mathematical function, especially such that does not have an elementary antiderivative and therefore it cannot be calculated using known integration rules and the Newton-Leibniz formula, can be very important in some technical applications. For example, if the control software controls a mechatronic system consisting of different moving mechanical parts, then the static and dynamic parameters of these moving parts are very important for this control software. If these moving parts, solids, perform a rotary motion, the static and dynamic parameters of these rotating solids, moments of inertia and rotational kinetic energy can be very important for the control software of the mechatronic

[1] University of Economics in Bratislava, Faculty of Economic Informatics, Department of Applied Informatics, Dolnozemská cesta 1, 852 35 Bratislava, Slovakia, e-mail: igor.kostal@euba.sk

[2] University of Economics in Bratislava, Faculty of Economic Informatics, Department of Applied Informatics, Dolnozemská cesta 1, 852 35 Bratislava, Slovakia, e-mail: pavol.sojka@euba.sk

system. If the material solid of the mechatronic system was created by the revolution of the area under the curve of a function *f(x)* within an interval <*a, b*> around the *x* axis, then, the *moment of inertia of such homogeneous solid of revolution* with respect to the axis of revolution is (Smith & Longley, 1910)

$$= \int (( )) \quad (1)$$

where ρ is the density of the solid [kg/m3]. Then, we can compute the *rotational kinetic energy of this solid* (Smith & Longley, 1910):

$$= \quad (2)$$

where ω is the angular velocity of the solid [rad/s] revolving about the *x* axis.

According to the actual value of the rotational kinetic energy of the solid of revolution of the mechatronic system, the control software of this system can regulate the revolutions of this solid so that e.g. did not exceed the maximum permissible value, which could result in damage of this solid and malfunction of the entire mechatronic system subsequently. To calculate both these static and dynamic parameters of this solid of revolution, it is necessary to know the value of a definite integral of the function *f(x)* on the interval <*a, b*>. If the function *f(x)* does not have an elementary antiderivative, then one of possibilities to calculate the value of a definite integral of this function is to use some numerical integration rule, e.g. one of three rectangular integration rules. Our .NET application can calculate the value of a definite integral also of such function *f(x)* using rectangular numerical integration rules, so it can be used as a support software module of the control software of a mechatronic system. In the next chapters of the paper we briefly deal with rectangular integration rules, their implementation in the methods of our .NET application, types of parallelization of instance methods of .NET application and experiment in which we examine the execution efficiency of parallelization of instance methods with implemented rectangular integration rules.

2. Rectangular Numerical Integration Rules

The value of a definite integral of the function *f(x)* that is continuous on a given closed interval <*a, b*> corresponds to the area under the curve of this function within the interval <*a, b*> on the horizontal axis (*x*-axis). If we can use the integration rules to find the elementary antiderivative of the function *f(x)* that is composed of elementary functions, then we can calculate the value of a definite integral of the function *f(x)* using the Newton-Leibniz formula. However, there are functions that do not have elementary antiderivatives, then we can not use the Newton-Leibniz formula. One of possibilities to calculate the values of definite integrals of such functions is to use numerical integration rules to approximate the value of a definite integral. There are several numerical integration rules, e.g. rectangular rules, trapezoidal rule, Simpson's rule and others.

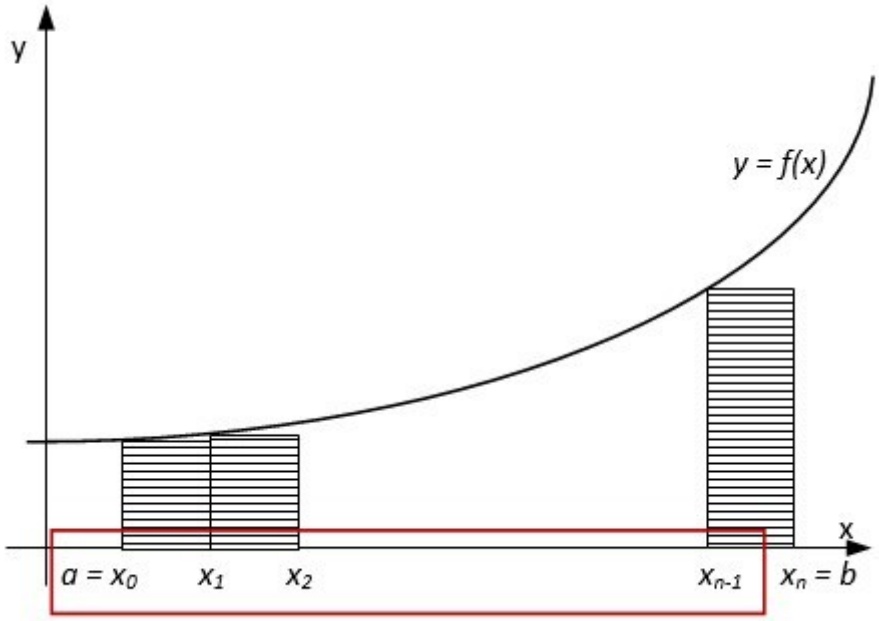
Our .NET application uses three rectangular rules, the left-, right-, and midpoint rules. Each of these rules divides the area under the curve of the function *f(x)* into several rectangles. The sum of the areas of these rectangles is then the approximate value of a definite integral of the function *f(x)* on the integration interval <*a, b*>. If there are more rectangles, then the higher the accuracy of the given approximation rectangular rule is achieved.

The *left-point rule* sets the height of each rectangle to the leftmost value of *f(x)* in each subinterval, so that the approximation is (Rouah & Vainberg, 2007)

$$\int ( ) \approx \_\_(( )+ ( )+\cdots+ ( )) \quad (3)$$

where *a* and *b* are the endpoints of the integration interval, *n* is the number of subintervals, $x_0=a$, $x_i=x_0+ih$ *(i=0, 1 ... n-1)* is the leftmost point of each subinterval, $f(x_i)$ is the value of the function evaluated at x_i and $h=(b-a)/n$ is the width of each subinterval.

Fig. 1: Left-Point Rule Numerical Integration



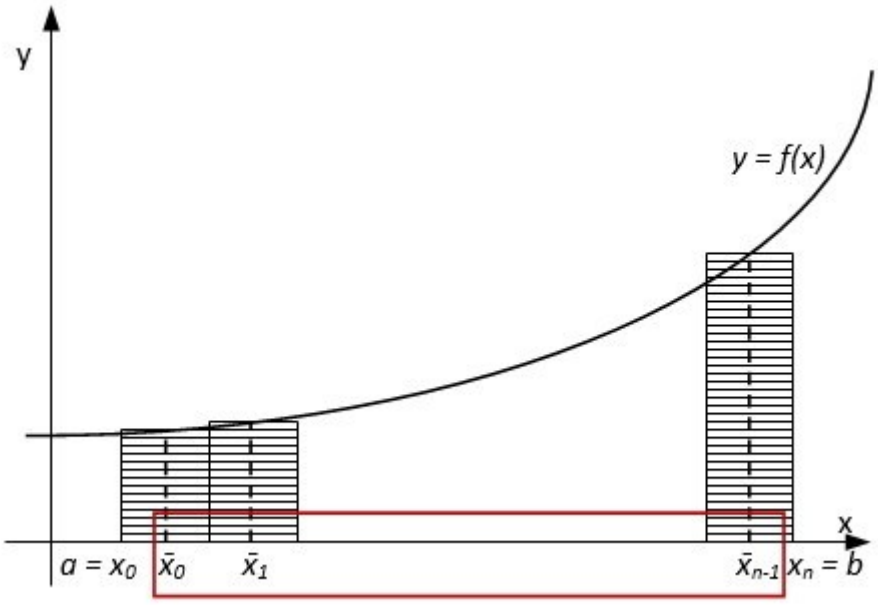
Source: (author)

The *midpoint rule* uses the value of *f(x)* obtained at the midpoint of each subinterval for the height of rectangles. Mathematically, the midpoint rule can be expressed as (Rouah & Vainberg, 2007)

$$\int ( ) \approx \_\_(( )+ ( )+\cdots+ ( )) \quad (4)$$

where *a* and *b* are the endpoints of the integration interval, *n* is the number of subintervals, $x_0=a$, $\bar{x}_i = x_0+(i+0,5)h$ $(i=0, 1 \ldots n-1)$ is the midpoint of each subinterval, $f(\bar{x}_i)$ is the value of the function evaluated at $\bar{x}_i$ and $h=(b-a)/n$ is the width of each subinterval.

Fig. 2: Midpoint Rule Numerical Integration



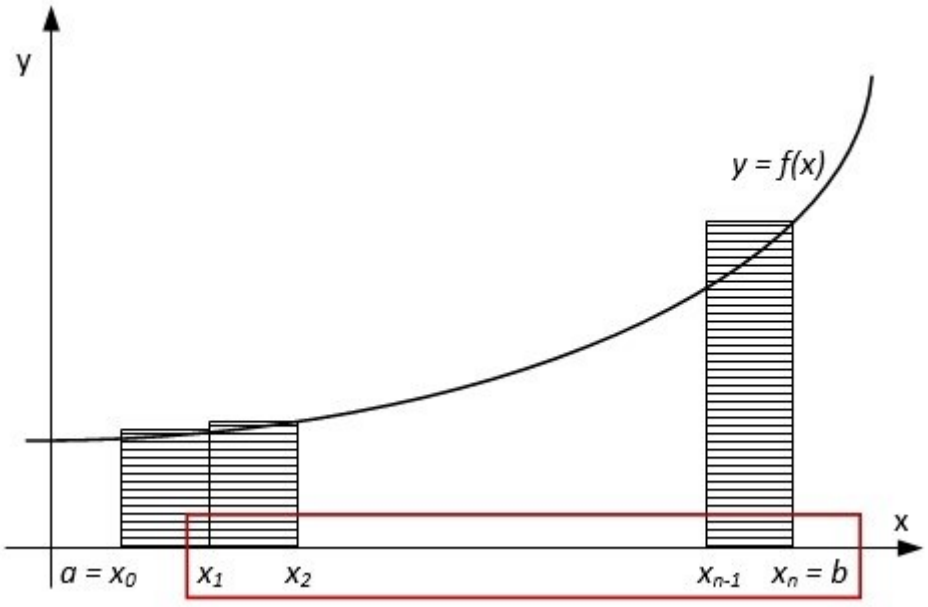
Source: (author)

The right-point rule sets the height of each rectangle to the rightmost value of f(x) in each subinterval, so that the approximation is (Rouah & Vainberg, 2007)

$$\int ( ) \qquad \approx\_\_(( )+ ( )+\cdots+ ( )) \tag{5}$$

where *a* and *b* are the endpoints of the integration interval, *n* is the number of subintervals, x0=a, xi=x0+ih (i=1, 2 … n-1) is the rightmost point of each subinterval, f(xi) is the value of the function evaluated at xi, xn=b, and h=(b-a)/n is the width of each subinterval.

Fig. 3: Right-Point Rule Numerical Integration



Source: (author)

3. Parallel Programming in the .NET Framework and Data Parallelism

Nowadays, many personal computers have four, six or eight cores (that is, CPUs) that enable multiple threads to be executed simultaneously. To take advantage of such hardware, we can parallelize our code to distribute work across multiple processors. The .NET Framework 4 enhance support for *parallel programming* by providing the Task Parallel Library.

The **Task Parallel Library** (TPL) is a set of public types and APIs (Application Programming interfaces) in the *System.Threading* and *System.Threading.Tasks* namespaces in the .NET Framework version 4. The purpose of the TPL is to make developers more productive by simplifying the process of adding parallelism and concurrency to applications. (Microsoft, 2021)

In parallel programming in the .NET Framework we know two kinds of a parallelism - data and task parallelism. We did not use task parallelism in a parallelized code of our .NET application.

Data parallelism refers to scenarios in which the same operation is performed concurrently (that is, in parallel) on elements in a source collection or array. Data parallelism with imperative syntax is supported by several overloads of the *For* and *ForEach* methods in the *System.Threading.Tasks.Parallel* class through which TPL supports this parallelism. In data parallel operations, the source collection is partitioned so that multiple threads can operate on different segments concurrently. We write the loop logic for a *ParallelFor* or *ParallelForEach* loop much as we would write a sequential loop. We do not have to create threads or queue work items. The TPL handles all the low-level work for us. When a parallel loop runs, the TPL partitions the data source so that the loop can operate on multiple parts concurrently. Behind the scenes, the *Task Scheduler* partitions the task based on system resources and workload. When possible, the scheduler redistributes work among multiple threads and processors if the workload becomes unbalanced. However, we can also supply our own custom partitioner or scheduler. (Microsoft, 2021) Data parallelism with the *Parallel.For* and *Parallel.ForEach* methods is used in a parallelized code of our .NET application.

4. Delegates in .NET Applications

A delegate is a type that represents references to methods with a particular parameter list and return type. When we instantiate a delegate, we can associate its instance with any method with a compatible signature and return type. We can invoke (or call) the method through the delegate instance. (Microsoft, 2021)

Delegates are used to pass methods as arguments to other methods. Event handlers are nothing more than methods that are invoked through delegates. We create a custom method, and a class such as a windows control can call our method when a certain event occurs. (Microsoft, 2021)

Delegates have the following properties: (Microsoft, 2021)

- Delegates are similar to C++ function pointers, but delegates are fully object-oriented, and unlike C++ pointers to member functions, delegates encapsulate both an object instance and a method.
- Delegates allow methods to be passed as parameters.
- Delegates can be used to define callback methods.
- Delegates can be chained together; for example, multiple methods can be called on a single event.

- Methods don't have to match the delegate type exactly.
- Lambda expressions are a more concise way of writing inline code blocks. Lambda expressions (in certain contexts) are compiled to delegate types.

5. A .NET Application with Parallel Methods that Use Delegates

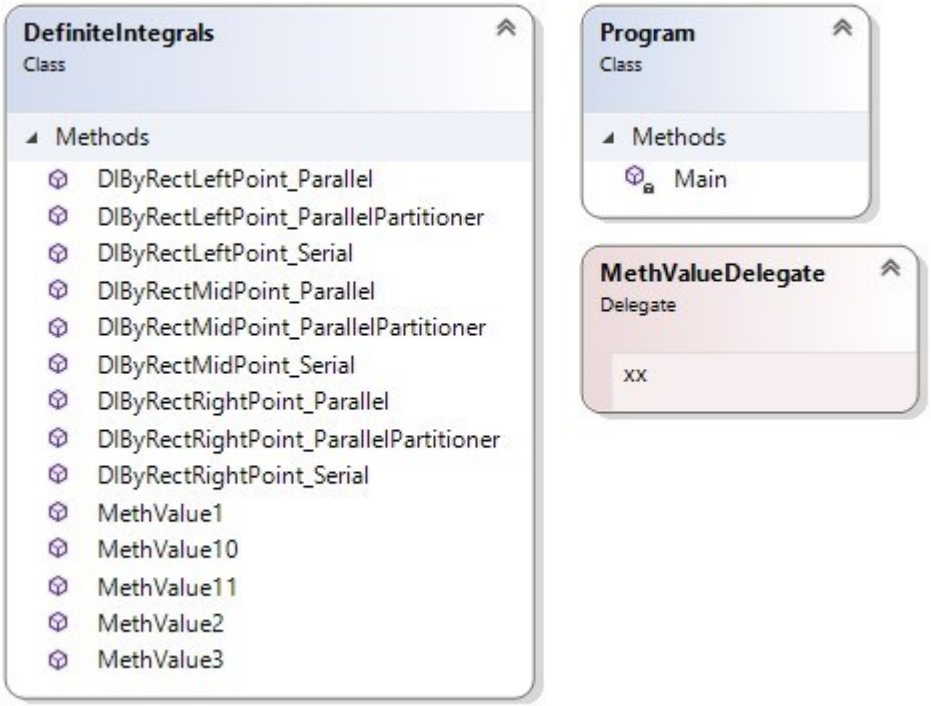
Our .NET application was created in the C# language in the Microsoft Visual Studio 2019 Enterprise development environment to the Microsoft .NET Framework version 4 and therefore it must be installed on the computer with this .NET application. Its source code is divided into two classes, *DefiniteIntegrals* and *Program*. The *Program* class contains one static *Main* method, which is the entry point for executing the entire .NET application. The *di* object of the *DefiniteIntegrals* class is created in this static method. This *di* object calls particular serial and parallel instance methods in which three rectangular integration rules, the left-, right-, and midpoint rules are implemented. All results of calculations of values of definite integrals of functions $f_1(x) = e^{2x}/x$ and $f_2(x) = sin(3x)/x$ calculated by instance methods of the *di* object, including their execution times measured by the *Main* method, are stored in the log file *LogFile.txt* by this method.

Outside of these classes, the *MethValueDelegate* delegate is declared in the application source code

public delegate double MethValueDelegate (double xx);

whose object can point to a method that has one *double* parameter and returns a *double* value.

Fig. 4: Class Diagram of our .NET Application



Source: (author)

The ***left-point rule*** is implemented in the *DIByRectLeftPoint_Serial*, *DIByRectLeftPoint_Parallel*, and *DIByRectLeftPoint_ParallelPartitioner* member methods of the *DefiniteIntegrals*

class. These methods compute the values of definite integrals in serial, in parallel, and in parallel with the partitioner.

The ***midpoint rule*** is implemented in the *DIByRectMidPoint_Serial*, *DIByRectMidPoint_Parallel*, and *DIByRectMidPoint_ParallelPartitioner* member methods of the *DefiniteIntegrals* class. These methods compute the values of definite integrals in serial, in parallel, and in parallel with the partitioner.

Fig. 5: The Source Code of the Serial Integration Member Method with implemented the Midpoint Rule

```
public double DIByRectMidPoint_Serial(double a, double b, long n, int methodNumber) {
        MethValueDelegate MethValueDlg; // declaration of the 'MethValueDlg' object of a delegate

                                        // create the 'MethValueDlg' object that point to the 'MethValue1' method
        if (methodNumber == 1) MethValueDlg = new MethValueDelegate(MethValue1);
        else if (methodNumber == 2) MethValueDlg = new MethValueDelegate(MethValue2);
        else if (methodNumber == 3) MethValueDlg = new MethValueDelegate(MethValue3);
        else if (methodNumber == 10) MethValueDlg = new MethValueDelegate(MethValue10);
        else if (methodNumber == 11) MethValueDlg = new MethValueDelegate(MethValue11);
        else MethValueDlg = new MethValueDelegate(MethValue10);

        double sum = 0.0;
        double step = (b - a) / (double)n;
        for (int i = 0; i < n; i++)
        {  double x = a + (i + 0.5) * step;   sum = sum + MethValueDlg(x);  }
        return step * sum;  }
```

Source: (author)

Fig. 6: The Source Codes of the Parallel Integration Member Methods with implemented the Midpoint Rule

```
public double DIByRectMidPoint_Parallel(double a, double b, long n, int methodNumber) {
        MethValueDelegate MethValueDlg;

        if (methodNumber == 1) MethValueDlg = new MethValueDelegate(MethValue1);
        else if (methodNumber == 2) MethValueDlg = new MethValueDelegate(MethValue2);
        else if (methodNumber == 3) MethValueDlg = new MethValueDelegate(MethValue3);
        else if (methodNumber == 10) MethValueDlg = new MethValueDelegate(MethValue10);
        else if (methodNumber == 11) MethValueDlg = new MethValueDelegate(MethValue11);
        else MethValueDlg = new MethValueDelegate(MethValue10);

        double sum = 0.0;
        double step = (b - a) / (double)n;
        object monitor = new object();
        Parallel.For(0, n, () => 0.0, (i, state, localSum) =>
        {  double x = a + (i + 0.5) * step;  return localSum + MethValueDlg(x);  },
        localSum => { lock (monitor) sum += localSum; });
        return step * sum;  }
```

```
public double DIByRectMidPoint_ParallelPartitioner(double a, double b, long n, int methodNumber)
{       MethValueDelegate MethValueDlg;

        if (methodNumber == 1) MethValueDlg = new MethValueDelegate(MethValue1);
        else if (methodNumber == 2) MethValueDlg = new MethValueDelegate(MethValue2);
        else if (methodNumber == 3) MethValueDlg = new MethValueDelegate(MethValue3);
        else if (methodNumber == 10) MethValueDlg = new MethValueDelegate(MethValue10);
        else if (methodNumber == 11) MethValueDlg = new MethValueDelegate(MethValue11);
        else MethValueDlg = new MethValueDelegate(MethValue10);

        double sum = 0.0;
        double step = (b - a) / (double)n;
        object monitor = new object();
        Parallel.ForEach(Partitioner.Create(0, n), () => 0.0, (range, state, localSum) =>
        {   for (long i = range.Item1; i < range.Item2; i++)
            {  double x = a + (i + 0.5) * step;  localSum += MethValueDlg(x);  }
            return localSum;
        }, localSum => { lock (monitor) sum += localSum; });
        return step * sum;  }
```

Source: (author)

The ***right-point rule*** is implemented in the *DIByRectRightPoint_Serial*, *DIByRectRightPoint_Parallel*, and *DIByRectRightPoint_ParallelPartitioner* member methods of the *DefiniteIntegrals* class. These methods compute the values of definite integrals in serial, in parallel, and in parallel with the partitioner.

All member methods mentioned in the previous three paragraphs are created universally. This means they can calculate a definite integral of arbitrary mathematical function, which is identified by a number passed into a parameter of the given integration method. According to this number, the method creates a delegate object that points to a member method that calculates and returns a function value at a given point *x* of the mathematical function implemented in this member method. Using this returned function value and using the given integration rule, the integration method calculates the value of a definite integral by a numerical parameter the identified mathematical function.

If the integration methods would not be created universally, we would have to create one integration method for each mathematical function and the given integration rule, which would work without using the delegate object, but it would be able to calculate the value of a definite integral of only one mathematical function implemented in this integration method. In our .NET application, all we have to do is to create a "stack" of the *MethValueX* member methods with implemented appropriate mathematical functions and thus to expand the number of these functions for which the .NET application can calculate the value of a definite integral.

6. Experiment and Its Results

We performed the experiment in order to measure the execution efficiency of parallelized integration instance methods of the object of our .NET application compared to serial integration instance methods of the same object.

Execution times of particular serial, parallel and parallel instance integration methods using a partitioner of the object provides us the .NET application itself, it also measures these times in its *Main* method and saves them into the log file *LogFile.txt* on the hard disc and displays them into the console. We used the measured execution times for calculations of speedups of both types of parallel instance methods in opposite to serial instance methods.

Our .NET application was running on the computer with the following basic hardware configuration: CPU: Intel Core i7-4700 MQ Processor (6MB Cache, 2.40 GHz, 5 GT/s, 4 Cores, 8 Threads), RAM: 8 GB.

Our .NET application can calculate definite integrals of the functions $f_1(x) = e^{2x}/x$, $f_2(x) = sin(3x)/x$ and others using the left-point, right-point and midpoint rules. For each from these three rules has our .NET application created three integration instance methods of its object - serial, parallel and parallel with a partitioner (Fig. 4).

For the measurement of execution times of particular integration instance methods we used two sets of inputs entered successively into the .NET application input.

The first set of inputs for the function $f_1(x) = e^{2x}/x$ *:*

the endpoints of the integration interval <a, b> : **0.18 1.19**

the numbers of integration subintervals : **105000000**, **150000000**, **200000000**, **300000000**, **400000000**, **500000000**, **600000000**, **700000000**, **800000000**, **900000000**, **1200000000**

The second set of inputs for the function $f_2(x) = sin(3x)/x$ *:*

the endpoints of the integration interval <a, b> : **0.01 0.98**

the numbers of integration subintervals : **105000000**, **150000000**, **200000000**, **300000000**, **400000000**, **500000000**, **600000000**, **700000000**, **800000000**, **900000000**, **1200000000**

The definite integrals values of the f_1 function and the execution times values of all the instance methods that calculated these definite integrals values were computed and measured for *the first set of inputs*. The definite integrals values of the f_2 function and the execution times values of all the instance methods that calculated these definite integrals values were computed and measured for *the second set of inputs*. The measured values of the execution times of calculations of three instance methods (serial, parallel and parallel with a partitioner) computing the definite integrals values of one function by a single rule for one set of inputs are always displayed in one graph.

We computed the speedup of every parallel method using p computing elements (CPUs) by the following formula

$$Speedup(p) = T_{serial} / T(p)$$

The term T_{serial} refers to the execution time of the serial version of the given method.

The accuracy of the computations of definite integrals values by particular instance methods of the .NET application is sufficient. We can see it, for example from the computations output of definite integrals values computed by the midpoint rule (the rectangular rule with the smallest error) for the first set of inputs for the selected function $f_1(x) = e^{2x}/x$ (Fig. 7). This output was displayed into the console output of the .NET application.

Fig. 7: The Output of our .NET Application

```
Insert the method number for an integration with using delegates
[e^(2x)/x  : 1, cos(x^2) : 2, sin(3x)/x : 3, 4/(1+x^2) : 10, (e^x)/x : 11] : 1

Insert the endpoints of the integration interval <a, b> with a space: 0.18 1.19

Insert the number of integration subintervals of the functions integration (max. 4000 000 000):
105000000
_____

The definite integral of the f(x) = e^(2x)/x function over the <0.18, 1.19> interval computed with the
number of integration subintervals 105000000

   by the Midpoint Rectangle Rule:

 [6.55860120381387 (computed by the 'int' MATLAB function)]

       Serial computed:
 6.5586012038130 4  [the execution time: 00:00:02.9324582 (2932.4582 ms)]
       PARALLEL computed:
 6.5586012038138 9  [the execution time: 00:00:01.1327733 (1132.7733 ms)]
       PARALLEL with the PARTITIONER computed:
 6.5586012038137 5  [the execution time: 00:00:00.7833545 (783.3545 ms)]

   by the Left-Point Rectangle Rule:

       Serial computed:
 6.5586011984428  [the execution time: 00:00:02.8434350 (2843.435 ms)]
       PARALLEL computed:
 6.55860119844246  [the execution time: 00:00:01.1019994 (1101.9994 ms)]
       PARALLEL with the PARTITIONER computed:
 6.55860119844249  [the execution time: 00:00:00.7460132 (746.0132 ms)]

   by the Right-Point Rectangle Rule:

       Serial computed:
 6.55860120918544  [the execution time: 00:00:02.8334502 (2833.4502 ms)]
       PARALLEL computed:
 6.55860120918521  [the execution time: 00:00:01.0982331 (1098.2331 ms)]
       PARALLEL with the PARTITIONER computed:
 6.55860120918513  [the execution time: 00:00:00.7590929 (759.0929 ms)]
```

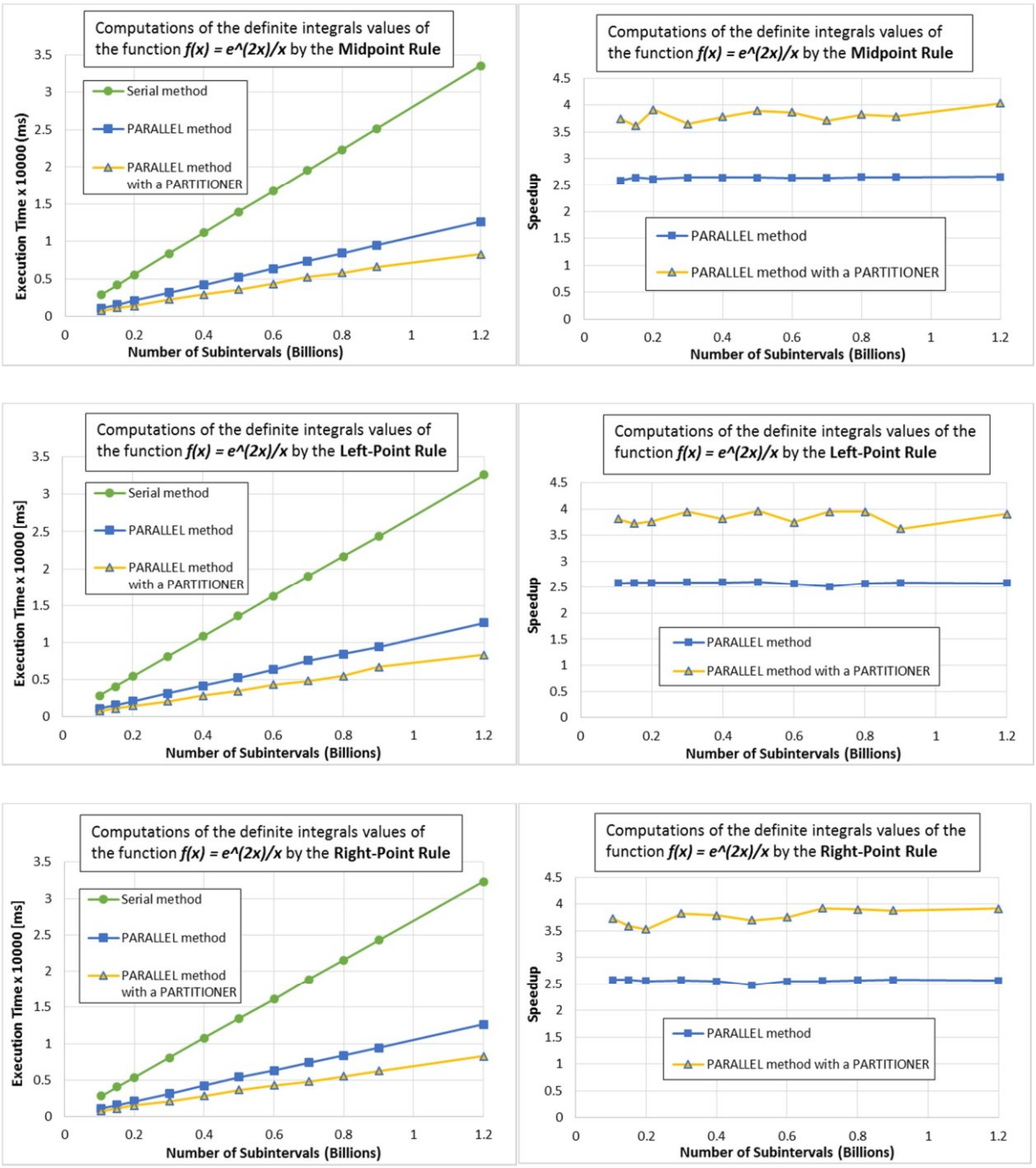
Source: (author)

In this output there are 3 calculated definite integrals values of the function $f_1(x) = e^{2x}/x$ identical with the value computed by the MATLAB function *int*.

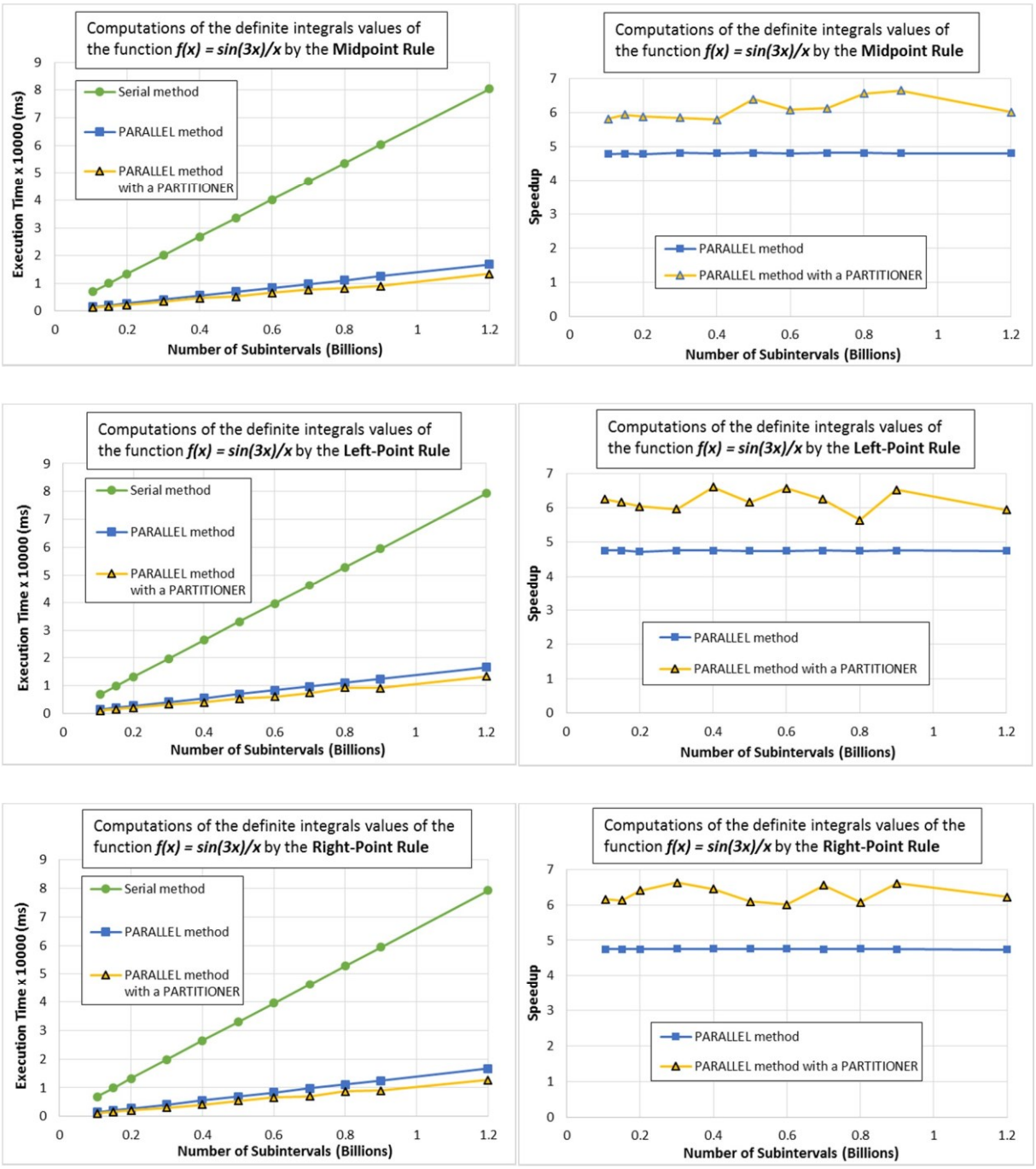
Execution times of all the integration instance methods and speedups of all parallel and parallel integration instance methods with a partitioner of our.NET application calculating definite integrals values of the particular functions are shown in the following graphs.

Fig. 8: Execution times of all instance methods of the .NET application and speedups (opposite serial methods) of all parallel and parallel instance methods with a partitioners of the .NET application for both integrated functions ((a) $f_1(x) = e^{2x}/x$, (b) $f_2(x) = \sin(3x)/x$)

(a)


Computations of the definite integrals values of the function f(x) = e^(2x)/x by the Midpoint Rule
Execution Time x 10000 (ms)
Number of Subintervals (Billions)
Serial method
PARALLEL method
PARALLEL method with a PARTITIONER
Computations of the definite integrals values of the function f(x) = e^(2x)/x by the Midpoint Rule
Speedup
Number of Subintervals (Billions)
PARALLEL method
PARALLEL method with a PARTITIONER
Computations of the definite integrals values of the function f(x) = e^(2x)/x by the Left-Point Rule
Execution Time x 10000 [ms]
Number of Subintervals (Billions)
Serial method
PARALLEL method
PARALLEL method with a PARTITIONER
Computations of the definite integrals values of the function f(x) = e^(2x)/x by the Left-Point Rule
Speedup
Number of Subintervals (Billions)
PARALLEL method
PARALLEL method with a PARTITIONER
Computations of the definite integrals values of the function f(x) = e^(2x)/x by the Right-Point Rule
Execution Time x 10000 [ms]
Number of Subintervals (Billions)
Serial method
PARALLEL method
PARALLEL method with a PARTITIONER
Computations of the definite integrals values of the function f(x) = e^(2x)/x by the Right-Point Rule
Speedup
Number of Subintervals (Billions)
PARALLEL method
PARALLEL method with a PARTITIONER

(b)



Source: (author)

Brief results analysis. From the presented graphs showing execution times of all integration instance methods and speedups of all parallel and parallel methods with a partitioner of our .NET application in dependency on the numbers of integration subintervals it is obvious that all parallel methods have significantly shorter execution times compared to serial methods. Parallel methods with a partitioner reach even shorter execution times and better speedups than the parallel method without the partitioner.

7. Conclusion

Parallel integration instance methods without a partitioner and with the partitioner of the object of our .NET application are able to calculate definite integrals values with large numbers of integration subintervals (up to 1,2 billion) with significantly shorter execution times than serial integration instance methods of the same object of this .NET application.

Parallel instance methods with the partitioner of this object show a relatively large speedups, up to **6.63** for the function $f_2(x) = sin(3x)/x$, compared to serial instance methods of the same object. These speedups are generally larger than the speedups of parallel instance methods without the partitioner of this object. This can be explained by the fact that in parallel instance methods with the partitioner range partitioning is used. The only overhead involved in range partitioning is the initial work of creating the ranges; no additional synchronization is required after that. Therefore, it can provide good performance as long as the workload is divided evenly.

Our .NET application using parallel integration instance methods can calculate definite integrals values of the functions $f_1(x) = e^{2x}/x$ and $f_2(x) = sin(3x)/x$, which do not have elementary antiderivates, quite effectively and with high accuracy. In this case for the computation of definite integrals values of these functions we can not use the Newton-Leibniz formula, but we can use numerical approximation methods. Thus, our .NET application calculates the values of definite integrals of the functions $f_1(x)$ a $f_2(x)$ by the left-, right- and midpoint rules. Using such calculated definite integrals values can be computed, for example moments of inertia and a rotational kinetic energy of a homogeneous solids of revolution created by the revolution of curves of just such or similar functions around the x axis. These may be important static and dynamic parameters of mechatronic system components, which the control software of the system needs to determine the most effectively. Therefore our .NET application can be a useful part of this control software.

In addition, the use of universal integration instance methods, simple the *MethValueX* instance methods that compute the values of given mathematical functions at the point x, and delegates allows easy extension of our .NET application with additional mathematical functions, which then can its instance methods to integrate.

8. Resources

Microsoft Corporation. (2021, June 1). Documentation. Retrieved from https://docs.microsoft.com

ROUAH, F. D., VAINBERG, G. (2007). Option Pricing Models and Volatility Using ExcelVBA. John Wiley & Sons, Inc., Hoboken, New Jersey, 40-50.

SMITH, P. F., LONGLEY, W. R. (1910). Theoretical Mechanics. Ginn and Company, Boston - New York - Chicago - London, 7-30, 231-233.

CRITICAL THINKING AND TRUST: THE ART OF DECISION MAKING

Tamás Kun[1]

Abstract
In this article, we discuss the circumstances of misinformation and its effects on socio-economic matters. Many nations in the world use global communication policies and devices, which are stamping their marks on politics, trade, and social behavior. Trust in an institution, a news agency and other kinds of sources determines our choices in life, such as in voting, investing, purchasing goods and services and in decision making. In forward, many areas in our operations will be connected on such higher levels than it is done today, so in rapid succession, the chain of events would happen in couple of seconds all over the world. Although, information technology is topping any field of science, however, society and legal instruments cannot follow its development rate. Generally, in other sectors, measures may have often taken decades until they are coming into force, such as laws, medical treatments, but in the communication sector, or more likely info-communication these are less time-consuming things. If we want reliable information that is closer to reality, we must filter these data from their misleading elements, we must know when, where and why they fit into the great puzzle.

Keywords
critical thinking, trust, decision making, information filtering

1. Introduction

The purpose of this study is, to frame the nature of trust, and to understand the reasons how processes like news making are working. There are no coincidences, the behavior of the actors in this issue are much likely reasonable than to be otherwise. Like in chess, the moves on the table brings us the effect of a decision, which can be a part of a strategy, a sacrifice or even worse, a fatal choice. There is a new phenomenon in mass communication, because there is a huge inventory of tools for the mass population to express their ideas and opinions worldwide, in most cases live, causes the turn of mass communication's direction from one-sided to the many sides. There are less likely such things as "status quo" or "correct narrative", because the number of context editors has been expanded enormously. However, there is a huge problem with that opportunity. The phrase of "common sense" is fading away with these technological backgrounds. There are so many views and opinions and most of them are completely fake, they have no connection with reality, there is no action that could ratify them for their existence, to put it simple: they are lies. Facts does not need explanation for those who see and understand why they are happening. The world is still based on the laws of physics (moreover mathematics) today as well, it was the same as 333 years ago, where according to Newton's third law of motion (Newton, 1687) *"for every action there is an equal and opposite reaction",* this means for us if there is a fake event or lie in the world what never happened, curiosity and doubt arise to discover it why and how it happened, regardless of its justification. The difference between data and information is, that in the first case we know only numbers and measures, on the other side we know the connections between them. For the future, the

[1] Óbuda University Doctoral School of Safety and Security Sciences, 1081 Budapest, Népszínház str. 8., e-mail: kun.tamas@uni-obuda.hu; https://orcid.org/0000-0002-6620-7157;

main problem will be, that the historically working filters are fading, who helped us to choose between “good and bad” replaced with simulated and created events, what never existed in the same concept in the physical reality.

2. Theoretical background

Ennis says that the issue of critical thinking has been neglected in the past century. Although he instructed definition for this matter, which is having wide acceptance in the field, he says *“the picture I paint of critical thinking assessment is not all rosy, though there are some bright spots.”* His idea is based on Blooms’ taxonomy of educational objectives: analysis, synthesis, and evaluation. He defines critical thinking as: *“is the correct assessing of statements.”* (Ennis, 1993) Fisher writes in his book, that critical thinking is a basic competence, like reading and writing should be taught to future generations. His three main factors are the ability to interpreting, analyzing, and evaluating. With his co-author Scriven they have a definition for critical thinking as: *“is skilled and active interpretation and evaluation of observations and communications, information and argumentation”* (Fisher, 2011) According to Lipman, critical thinking is self-corrective, has criteria, and sensitive to context. (Lipman, 1987)

Seligman cites many examples, from the minutiae of everyday life to the central problems of political and economic life, showing what new "external" constraints are forcing civilization and trust in today's life to develop trust. The situation is confusing, but Seligman shows that trust loses its unifying power precisely because the individual, long regarded as the primary repository of rights and values, is reduced to the sum of an abstract matrix of group identities and rules. Seligman's irony is that, as we become postmodern, we kind of return to a premodern state in which group life is based on group sanctions, not on trust. (Seligman, 2000)

Catlette writes, that in an era of insecurity, global turmoil, divorce, and confused family life, it is increasingly difficult to remain vulnerable, optimistic, and open to life. Fortunately, many of us have friends and family we can rely on, or a partner we can turn to as a haven where we can relax, unwind, and be ourselves. But sometimes it’s hard even here. When everyday stress creeps into our protected space or unexpected relationship problems disrupt our peace of mind, we may feel insecure. We may also begin to question our partners love, devotion, and reliability. Without realizing it, we can answer these doubts by gently moving away from the loved one. Why is trust on such a shaky foundation? On the other hand, isn’t it risky to be too naive and confident? What trust issues are couples facing today that have barely existed for just a decade? How can we best deal with events or situations that may threaten our trust and confidence? Here comes the question, what do we mean under trust? The Oxford Dictionary defines trust as “a firm belief in the reliability, truth, ability, or strength of someone or something.” For example, we trust people who are kind to us, innocent (does not mean harm to us), and whose actions are consistent with their words. We are confident that we can expect someone to consistently do what is “right” from our perspective. We trust our partner in intimate relationships if they are predictable, reliable, and honest. Trust can also be defined as a verb: actions based on self-confidence. At the action level, trust includes the ability to do something without fear. (Catlette, 2016)

Vroom and Yetton writes in their book, that decision -making is a central point in many scientific studies, but they have different approaches. While the most of human behavior is usually a reflection of the decisions that people make, and those processes that are for to regulate and maintain control over these choices and decisions are key factors to understand and predict this behavior. In economics, statistics, and operation research a normative standpoint is the fundamental interest, whereupon the choices will be made. On the other hand, in psychology, sociology and political science, are interested in more likely in the outcomes of these decisions which led people to their behave as they did. (Vroom & Yetton, 1973)

Concluding the above we can say that critical thinking is not a widespread thing even today if nearly a century was not enough to make strong roots for it. Either way we are talking about critical thinking as a mindset or as a skill, we can agree upon that with the advance of technology and information culture, in the future it will be more and more crucial, to have the ability to choose correctly. The issue of trust building is the same also, throughout human history we are depending on our relatives which is mostly family and friends. In the case of making friendships, the idea of choice is leaking in the picture as an option (but also as a problem), just as the way it is with romantic relationships. For the case of decision-making there is a living conflict even these days, where scientists arguing with each other which factor is the more valuable: a nominal number or the predicted behavior. The "research problem" will be only one thing: the *choice* itself. A problem always carries the solution with it, which is the *decision,* that has to be made.

3. The legitimacy and the types of sources

In general, the best source of information is a *primary source* (or original source), which is the closest thing, what describes the most precisely what happened in connection with the research subject, and it has been made in with equal time of the study.

Primary sources are direct, first-hand messages about a topic from people who have been in direct contact with it. Primary sources can be:

- Legislation texts and other original documents,
- Newspaper reports from journalists who witnessed the event, or quote those who saw it,
- Speeches, diaries, letters, and interviews - what the participants said or wrote,
- Original research,
- Datasets, survey data such as censuses or economic statistics,
- Photos, videos, or sound recordings of the event.

At the case of *secondary sources*, things are becoming more difficult, because there is always an opinion embedded in them. The research method, the interpretation of the findings, intuition and many other factors can bend and/or alter these, in some cases unwillingly. This comes first-hand from the human nature, we are experiencing actions in different ways, that is why there are so many *certain points of view*, because we evaluate *truth* differently.

4. Model for information filtering

In standard, we are using mainly three keywords as *when*, *where,* and *why*. With these, we are localizing time and space, furthermore, searching for a causality between data, this results in links that connects and gives meaning for them. The first problem occurs when someone does not know where to look for information. There are good *indicators* in nowadays, like the *https* protocol, when we are browsing online, because the standard requires well set system time to show the page that we are looking for. Another good factor can be the *digital certification* of the domain that we are browsing.

In the case of noise reduction, there is a mathematical model called Kalman filter (Kalman, 1960) which would be a good starting point in clearing false elements from the broadcasted information. The model is an algorithm, which supposed to give an optimal

estimate for dynamic systems in motion. There are several fields, from space science to automotive (Bhargava, Verma, & Chaurasia, 2015), where this sampling method has been tested and in use today with several updates of the base concept.

News media is just like that, it has the noise of opposing parties, we are less likely friendly with outsiders or strangers, because it is a potential harm (that is the risk) to us. What we do not know, we fear it. Fellow scientists, who thinking about this issue, usually comes up with a solution searching. Trust is a two-sided connection, which means you can build it up or play it away. There is a race between content creators in modern times news broadcasting.

5. Models for building trust

Mayer et al.'s model (Mayer, Davis, & Schoorman, 1995) has three factors ***ability***, ***benevolence*** and ***integrity***. These factors significance based on a trustor's propensity which defines ultimately a trust. According to that, we can vote trust for anyone, however, that is not yet a relationship. To do that, the factors mentioned before must be confirmed, because this is only the Trustor's opinion about the Trustee, therefore this does not mean any action or commitment. As it goes in decision making, there is an inherited risk, in that case a perceived one. The authors underline the fact in their study, that the key concept is, when it comes to the willingness to become vulnerable, it is important to analyze the ***behavior*** itself, in that case what is necessary ***risk taking***, that leads us to the ***willingness*** to become ***vulnerable***. The model is dynamic, it shows only a snapshot of a cycle, through our role in a relationship we are making decisions over and over, which ***outcomes*** pointing back to the starting three factors.

Maister et al. describes a model (Maister, Green, & Galford, 2000) called "The Trust Equation" in their book, where they are not calculating with risk. The core principles of trust are relied on a two-sided connection. When this occurs, one party have confidence in the other. There is an equation model, where four elements are the following: creditability, reliability, intimacy, and self-orientation. Trust covers more than only one thing. When we speak about it, we are defining several actions and behavior that connected to the word.

When we share information with others, this state of mind or emotional status describes that we are in comfort with the channel and/or methods we are using, in other words, we are considering it as a secure channel, regardless is it actually true or not.

$$T_{rustworthiness} = \frac{C_{redibility} + R_{eliability} + I_{ntimacy}}{S_{elf-Orientation}}$$

Model of "The Trust Equation" developed by Maister et al. (2000)

Credibility has a function with ***words***. In general, we say that we have trust in something or someone, eventually, what the source states, because it is credible. Opposing that, reliability connected with ***actions***, where we can say that our supplier is reliable, because it will concern itself to the deadline that we have agreed before. In that case, we trust in the supplier's actions. Intimacy links to the security of our communication channel, where we can say that we can trust in our source, because it won't sell us out, so we can feel ***emotionally*** our information is in a safe place. This equals in information science with the term integrity.

6. Defining the reliable source and building trust

The first question that arises from the start is: *How can I trust you?* In past experiences, we called sources reliable or trusted, when they have been met with the following simple criteria:

- No need for to “fact-check” the statements called “news”, because they are just “right”,
- There was minimal backslash of negative feedback according to misleading and scandals,
- Too big in measures as an organization to be a liar

In these times, to check the validation for statements was ineffective with costs and time, so these “investigations” were rare, otherwise, their competitors done the job for us. But in that case, we had to be sure or at least marking the competitor with the “reliable” tag, which multiplies the standard problem in the process when we are building trust with a source.

News networks coverage

The online news has a different approach than traditional broadcasting like TV or radio because the content can be easily copied and transferred to another audience. This fact is shaping the social media communication trend nowadays, where now the actors can comment under posts, share them, or even make reactions through with emoticons, stickers, and other tools. In the past, on editorial duties there were many teams who had to cooperate their work to make the daily progress: photo gathering, making the physical form, choosing formats, making paper and context. In opposition today, a few people can do that work virtually with a computer. There are several editing programs for audio-video content and infographic solutions.

Trust in “mainstream media” news network valued up again under the pandemic and the so called “fake news” became a good reflection point to them. Research made by the United States Studies Centre, La Trobe University and the University of Melbourne revealed that trust in professional journalism still has value in the case of choosing the source of information. A sample of a thousand both in Australia and the U.S. showed, that under the COVID-19 pandemic, when it came to gathering news about the pandemic, they prefer the old way. Preference tended to quality urban media brands, in the United States, trust heavily affected by party preferences, 68 percent of Australians trusted professional journalists opposing with the 57% among Americans. (Carson, Ruppanner, & Ratcliff, 2020)

According to Kantar’s global “Trust in News” research found that the reputation of the “mainstream media” was largely unaffected, while social media and digital-only news platforms suffered significant reputational damage in recent elections as a result of “fake” narrative loops. In a study, where 8,000 people from Brazil, France, the United Kingdom, and the United States were asked about their attitudes toward political and electoral news appeared that it was largely just a flop. The reputation of traditional print and broadcast media has proven to be more resilient than social media platforms and online news, primarily due to the depth of coverage. People strongly believe that quality journalism remains the key to a healthy democracy, but they are skeptical about what they read and about the effectiveness of journalism that is holding them accountable. Audiences are becoming more informed and sophisticated in their commitment and appreciation of news content. In Brazil and the United States, large percentages of the population believe that “fake news” influenced the results of their last elections. Media brands have clear growth potential in developing subscriber models targeting those under 35 who are more willing to pay for non-print news at the right price. Who do we trust? Fake news became popular all over the world in 2016 and 2017. Kantar found that these attacks on mainstream media did not largely tarnish its reputation. In the United States, Britain, France and Brazil, there is still a strong belief (73% agree) that quality journalism is the key to a healthy democracy, in the UK 77% of respondents agreed with this. However, just over half believe that what they read is “mostly” correct. Likewise, nearly two-thirds (61%) worldwide fear that politicians and business leaders are not adequately addressed by the media.

The “reputation crash” in 2017 hit social media companies, while the “traditional media” reputation was more flexible both globally and in the UK. Print magazines are the most trusted source of news, while social networking sites and messaging applications are the least trusted in the world. Traditional broadcasting channels (television and radio) rank second and third in terms of reliability, respectively, followed by newspapers. Online-only newsletters are significantly less trusted than their print and broadcast counterparts. (Cooke, 2020)

7. US Presidential elections in 2020 and past experiences

For more than a generation, research has tried to identify the personality traits that draw people’s attention to news. The paper concludes that people generally want journalism to be fair, balanced, accurate and complete. However, it is unclear what these common factors mean and how news organizations might achieve them. Even more complex, these traditional beliefs were formulated before the advent of the Internet and did not anticipate what kind of news consumers would face today and how publishers would deliver it. A study by Media Insight Project shows, that news trust can be divided into factors that publishers can consider, and consumers can recognize. It also shows that in the digital age, several new factors that have largely not yet been explored, such as ad penetration, navigability, download and update times, are also critical in deciding whether consumers consider them to be competent and worthy the publisher can trust. The study finds that the specific factors that lead people to trust and rely on a news source also vary from topic to topic. For example, how much consumers appreciate a certain component of trust depends on whether they want to search for news about politics, transportation, and weather, not just about lifestyle. In some subjects, consumers appreciate more detailed reports and expert resources. Ease of use is of great value, for others, fun is more important. And on social media, consumers are skeptical of content and want to use trust marks such as clearly identifying the original source of the message. Finally, it sheds new light on why trust should matter for today’s publishers: it is not only a journalistic activity, but also a business necessity. People who pay higher fees for trust factors are more likely to use news, are more likely to pay for it, install news apps, or share and advertise news with their friends. (Young, 2016)

Americans are using the Internet for source of news now, more than five times since 1998. They trust news on the Internet much more than they did in the past, when there were only a small percentage of Americans received such news in that way. Around now, four from ten Americans now say they trust online news, up from 25% in 1998, when Gallup last surveyed it. The confidence in radio and television conversations also rose to double digits over the same period. Meanwhile, from the news sources tested by Gallup in 1998 and 2019, CNN and network nightly news programs alone were the only sources of trust that were significantly worth less than before. From the overall news sources measured in the May 15-30 Gallup polls, three local sources - television, newspapers, and radio - are more credible than dozens of other sources. Local sources were among the most reliable in the 1998 survey, but not all three are at the top of the list currently. Picking as a source of information local television (74%), local newspapers (67%) and local radio (65%), several other news sources are trusted by most of the population. These include conversations with friends or family (62%), public television news (59%), National Public Radio (55%), nighttime television news programs (54%) and news and interviews on morning national television (51%). Less than half of Americans say they trust national newspapers (49%), CNN (48%), Fox News (43%), online news (40%), radio conversations (38%), in television conversations (33%). and half-hour entertainment news TV programs (29%). (Jones & Saad, 2020)

In this remarkable year, the fallout from the 2020 presidential election was just as unusual - President Donald Trump has so far refused to allow Joe Biden to run for president (since his December 15 speech). However, a new Pew Research Center poll reveals well-known biased

patterns of how Republicans and Democrats has reacted to the latest news, including the ongoing political drama and the rise of the coronavirus across the country. Party members largely approve of their party's candidate's post-assembly message, though about a third of Republicans say Trump was wrong with the Americans in the party split, and after the polls ended on November 3rd, there were dramatically different views on Trump and Biden's public statements, approvals of the post-election messages sent by the candidate of his party, and does not accept the messages of the opponent. These views are especially prevalent among Democrats and independents leaning towards the Democratic Party all of them believed, that Biden is right, and Trump is giving the wrong message. But a poll conducted by the Center American News Pathways project between November 18 and 29 also found that Republicans are more divided than Democrats. About a third of Republicans and independents with Republican inclinations say Trump sent the wrong message since the election - and almost as many say Biden sent the right one. And while Trump has consistently argued that widespread electoral fraud was responsible for his defeat, Republicans say roughly one in ten electoral fraud allegations deserves more attention. And while Trump has consistently argued that widespread electoral fraud was the reason for his defeat, about ten Republicans say the allegations of electoral fraud do not deserve more attention. Half of American adults know that there has never been a losing presidential candidate in modern history who publicly refused to give his consent until inauguration day. Again, there is a big difference between parties: roughly two-thirds of Democrats (64%) answer this question correctly, compared to 36% of Republicans. On the other hand, Republicans prefer to say that there was a losing candidate who did not give up in modern history (23% versus 7% of Democrats), or that he was not sure if this happened (40% versus 29%). Most of both parties say their most frequently used news sources covered the election extensively, but Democrats generally praise the media and the reactions of each candidate. This election attracted the attention of most Americans, most people report positive news to their main news sources in many areas of post-election coverage, such as the voting process and individual candidate reactions. However, Democrats are far more likely than Republicans to say their news sources have done a good job. For example, 50% of Democrats say the news sources they look for the most explained the counting process “very well” after the polls ended, compared to 16% of Republicans who say the same (Mitchell et al., 2020).

8. Conclusion

Trust is an irreplaceable element of decision making, and when we are talking about its function, we barely analyze its components beneath the obvious approaches. Between theories there are different ways of explanation of common factors like risk, which is in “The Trust Equation” embedded in the “Self-Orientation” factor, however it is not mentioned separately. Point is, that this need for security comes from the uncertainty of growing technological systems. With automatization, systems are becoming more and more interconnected with each other, and beginners losing path to follow the development. This means uncertainty for these individuals, therefore rises the fear from the unknown. The other lesson is, that voting trust does not mean security, unless we are willing to take the risk to become vulnerable. We can trust blindly in a software or company, product, information etc., however, we will be the ones, who must bear the consequences of our choices. According to the data, trust is stronger on closer relations like family, friends, local news media etc., while on the other side, towards national or transnational media organizations trust is on lower level. This has nothing to do with “professionalism” that is coming from the human nature. That is why we are constantly looking for handrails to grab, analyzing, interpreting, and evaluating in order to evade the uncertainty, in other words: to evade risk.

9. Resources

BHARGAVA, A., VERMA, S., & CHAURASIA, B. K. (2015). Kalman filter for trust estimation in VANETs. *2015 IEEE UP Section Conference on Electrical Computer and Electronics (UPCON)* (old.: 1-6). Allahabad, India: IEEE.

CARSON, A., RUPPANNER , L., & RATCLIFF , S. (2021, June 23). Trust in quality news outlets strong during coronavirus pandemic. The Conversation. https://theconversation.com/trust-in-quality-news-outlets-strong-during-coronavirus-pandemic-138410.

CATLETT, J. (2016, August 2). Trust issues: Why is it so hard for some people to trust? PsychAlive. https://www.psychalive.org/trust-issues/.

COOKE, K. (n.d.). 'Fake news' reinforces trust in mainstream news brands. Kantar. Understand People, Inspire Growth. https://www.kantar.com/inspiration/advertising-media/fake-news-reinforces-trust-in-mainstream-news-brands.

ENNIS, R. H. (1993). Critical thinking assessment. *Theory into practice 32(3)*, 179-186.

FISHER, A. (2011). *Critical Thinking: An Introduction.* Cambridge: University Press.

JONES, J., & SAAD, L. (2020. 11 November 23). *In U.S., 40% Trust Internet News Accuracy, Up 15 Points.* https://news.gallup.com/file/poll/265808/190822Media.pdf

KALMAN, R. E. (1960). A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering 82(1)*, 35–45.

LIPMAN, M. (1987). Critical thinking: What can it be? *Analytic Teaching, 8(1)*, 5-12.

MAISTER, D., GREEN, C., & GALFORD, R. M. (2000). *The Trusted Advisor.* New York, NY: Simon and Schuster.

MAYER, R. C., DAVIS, J. H., & SCHOORMAN, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review, 20(3)*, 709-734.

NEWTON, I. (1687). *Philosophiæ naturalis principia mathematica (Mathematical principles of natural philosophy).* London, England: Benjamin Motte.

MITCHELL, A., JURKOWITZ, M., OLIPHANT, J. B., & SHEARER, E. (2020, December 15). Most Republicans Approve of Trump's Post-election Messaging, but About a Third Say It Has Been Wrong. Washington, DC; PEW RESEARCH CENTER.

SELIGMAN, A. B. (2000). *The Problem of Trust.* Princeton, New Jersey: Princeton University Press.

VROOM, V. H., & YETTON, P. W. (1973). *Leadership and Decision-Making.* London: University of Pittsburgh Press.

YOUNG, E. (2016, April 17). A New Understanding: What Makes People Trust and Rely on News. Chicago; American Press Institute.

SEARCH OF THE RATIONAL COST OF INFORMATION PROTECTION FROM INTERNAL THREATS FOR THE COMPANY'S BUSINESS PROCESS

Igor Mandritsa[1], Fabrizio d'Amore[2], Anna Fensel[3],
Alexander Zhuk[4], Vyacheslav Petrenko[5], Olga Mandritsa[6]

Abstract
The article is devoting to the search for cost criteria when choosing among the information protection options of the business process of a commercial firm from internal threats. Business process scheduled in stages of business operations allows you to get a criterion for determining the cost of protecting business information and not to spend above the norm.

Keywords
Cost criteria of protecting business information, Information security, Business process

1. Introduction

Information security is currently a critical need for all companies, worldwide. It is well known that information security is basing on the fulfilment of the CIA requirements, where the three-letters acronymous stands for confidentiality, integrity and availability: other requirements, like accountability, non-repudiation etc., are not considered here, because they are not always requested for information security (Bissell K., Ponemon L., Dal Cin P., 2019).

However here will focus only the Russian case. For our purpose it is particularly important to address confidentiality, as a data breach is a successful attack to this. Attacks to confidentiality are the most frequent attacks and address unauthorized access to information. There are two main paradigms for enforcing confidentiality: data encryption and access control, one not excluding the other. However, investments for implementing such measures are significant. Russian companies are actively investing in protection against hacker attacks. In 2018, investment in information security in IT budgets increased to 22%. The average business IT budget in the Russian Federation was $1.1 million in the next three years, there will be an increase of another 18%, due to the fact that the information technology infrastructure in companies is developing, and they need professional knowledge on cybersecurity, according to a study conducted by Kaspersky Lab (Infowatch, 2018).

[1] North-Caucasus Federal University, Institute of Information Technologies and Telecommunications, 1, Pushkin Street, Stavropol, 355009, Russia, imandritsa@ncfu.ru.

[2] SAPIENZA University, DIAG, formerly DIS - Department of Computer, Control and Management Engineering, Via Ariosto 25, I-00184 Rome, Italy, damore@diag.uniroma1.it.

[3] University of Innsbruck, Department of Computer Science, Technikerstr. 21a, A-6020 Innsbruck, Austria, anna.fensel@uibk.ac.at.

[4] North-Caucasus Federal University, Institute of Information Technologies and Telecommunications, 1, Pushkin Street, Stavropol, 355009, Russia, azhuk@ncfu.ru.

[5] North-Caucasus Federal University, Institute of Information Technologies and Telecommunications, 1, Pushkin Street, Stavropol, 355009, Russia, vipetrenko@ncfu.ru.

[6] Russian Technological University - MIREA, Branch office, Department of regional Economics, 8, Kulakov street, Stavropol, 355035, Russia, man_olga@mail.ru.

Financial damage to Russian companies from data leaks has increased over the past six months. For large businesses, it was approximately $246 thousand, 2.5% more than last year, according to a study by Kaspersky Lab. For the average – it has tripled to $74 thousands; the damage caused by cyber-attacks is becoming larger, broader and more serious and includes financial and strategic losses. Most likely, some cyber-attacks are part of national or state campaigns interests. In addition, some self-serving firms may intentionally limit their investment in cybersecurity and rely on information provided by other organizations to protect themselves. This can lead to insufficient investment in cybersecurity if all participants adopt the same strategy. The urgency of the problem causes an objective need to find a rational strategy and economic model for estimating the cost of protecting business processes from penetration and threats.

The subject of this research is the current business process in a commercial organization and its current economic strategy for protecting information while executing the organization's business process. The purpose of this article is to present the theoretical economy fundamentals to choose one of possible variants of economic strategy in information security process of business organizations, through a combinatorial selection of factors of the mathematical model of minimally adequate level of protection of the information assets of the organization to counter threats to the business process in the framework of the strategy of multi-level proposals for the recovery and to counter the information threats to the organization.

In turn, we will give a brief description of the role, place, and importance of current information security for the organization's business process. An important point is to determine the stages of economic operations (processes) within the framework of the client's order execution. Today, commercial organizations in the course of their production and economic activities widely use CRM systems, in which the determining factor in describing and optimizing the current stages and tasks of business processes are correctly defined goals at each stage of customer order fulfilment, which each business has defined for itself. Companies have different approaches to the formulation of these goals and objectives.

For example, in the economic practice of business processes of enterprises, its management identifies two distinct formulations of business goals, which determine which methodology for describing business processes in the CRM environment (Microsoft Project) will be the most acceptable and most effective. Accordingly, there are two models, the content of which is as follows. In the first model, the purpose and objectives of the organization of the entire business process of the enterprise are defined functionally, namely, it is provided that «...to solve the tasks, it is necessary to create a functional (process) model of the company that displays the structure, relationships and functions of the system, as well as the flows of information and material objects that link these functions» (Salnitri, M., Dalpiaz, F. & Giorgini, P., 2014). This approach implements an economic model such as «budget localization» or «cost budgeting» for the stages of production and sale of products and services ordered by the client.

For our study of this model, the future amount of possible damages from information security will depend on the measures and their cost, which will ensure the network infrastructure of the «data movement» between business units from the position of possible cyber-attacks by competitors in order to stop or disrupt the entire business process, create «temporary damage» to stop the business process. The second model describes the setting of goals and tasks in different way and proceeds from the fact that «... descriptions of algorithms (scenarios) for the execution of processes in the form of its participants are necessary. In which, first of all, employees and cause-and-effect relationships, time sequences of their performance of economic actions, are defined as an ordered combination of events and functions of employees» (Ahmed, N., 2014)

In this case, the emphasis is on describing the sequence of actions of employees participating in the business process, determining the timing of initial and final events, identifying participants, performers, material and documentary flows and the cost of the business process by its stages. For the purpose of ensuring information security of the second model of conducting a step-by-step business process, it will be important to determine the localization of information flows between employees of the organization by stages of work when executing a client's order in the existing software.

This model of business process organization is more susceptible to cyber fraud. Cyber fraudsters (hackers) made it as difficult as possible and even reset the information between the stages of the business process of production and sale of products and services by introducing chaos, disorienting the company's employees, and violating the integrity of information on client contracts in the software that elaborates the business process of the enterprise. The total amount of possible damage will depend on the time spent by the specialist on restoring the document flow system between employees, as well as the time spent on restoring financial flows between the client and the order execution departments. Due to the existing problems of ensuring information security in the organization of a business process based on software, it is necessary to create an economic model for evaluating the protection of the business process, which is the purpose of further research.

2. Results and discussion

Existing approaches to describing business processes, as well as existing software, with rare exceptions, are specialized and poorly suited for solving those tasks for which they were not originally intended. Figure 1 below provides a detailed description of the business task by business process entities, using the example of a conditional commercial organization, in the form of a customer order project (business task) in the Microsoft Project environment. The information security specialist must, first, secure the business process itself in the form of forming a plan of protective measures as an estimate of protection measures within the framework of the current information security standards in the Russian Federation adopted by the FSTEC of the Russian Federation:

- GOST R ISO / IEC 15408-1-2008 «Information technology-security Methods - criteria for evaluating it security-Part 1: Introduction and General model (IDT)» [2];
- GOST R ISO / IEC 27007-2014 identical to the international standard ISO / IEC 27007:2011 «Information technology. Security method. Guidelines for audit of information security management systems»;
- ISO/IEC 27002:2013 (E) «Information technology. Method of protection. A set of recommended rules for managing information security».

Secondly, the information security specialist will need to classify the threats to the business process and their probabilities, and then develop a modernized version of the threat model within the framework of the order of the FSTEC of the Russian Federation dated February 11, 2013 N 17, in which the protection of the client's personal data and the business process we are considering have the same features as when drawing up a threat model in relation to the business process. These features (in correlation with the international standard ISO / IEC 27002:2013) include:

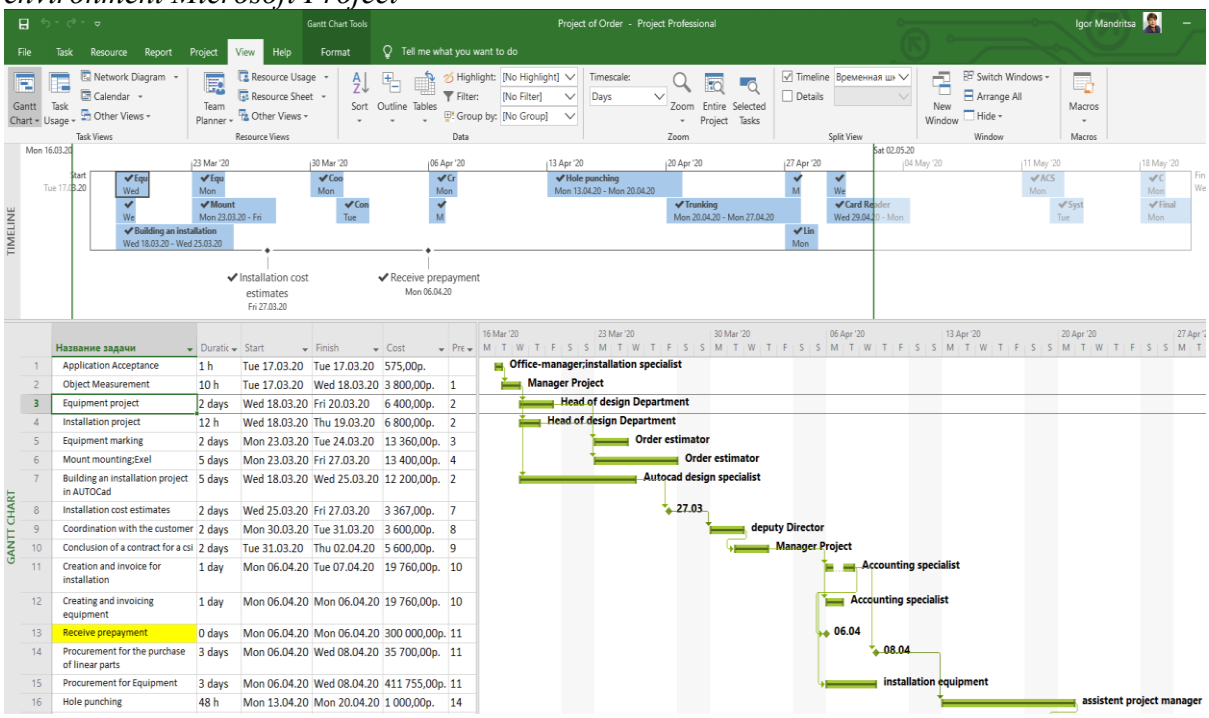
- area of probable threats by stages of the business process;

- the type of information asset that will be threatened;
- the view of this factor and threats;
- indicators of the organization's business process that will be exposed to the threat.

According to this standard (which has no Russian counterpart), organizations of all types and sizes (including the public and private sector, commercial and non-commercial) accumulate, process, store and transmit information in various forms, including electronic, physical and oral (for example, interviews and presentations) information.

The value of information in the business process is not only in documented terms, numbers, and images – it is knowledge, concepts, ideas, technologies, and brands as examples of intangible forms.

Fig.1: Stages of the business process "Customer-ordered project implementation" in the environment Microsoft Project



	Название задачи	Duratic	Start	Finish	Cost	Pre
1	Application Acceptance	1 h	Tue 17.03.20	Tue 17.03.20	575,00р.	
2	Object Measurement	10 h	Tue 17.03.20	Wed 18.03.20	3 800,00р.	1
3	Equipment project	2 days	Wed 18.03.20	Fri 20.03.20	6 400,00р.	2
4	Installation project	12 h	Wed 18.03.20	Thu 19.03.20	6 800,00р.	2
5	Equipment marking	2 days	Mon 23.03.20	Tue 24.03.20	13 360,00р.	3
6	Mount mounting;Exel	5 days	Mon 23.03.20	Fri 27.03.20	13 400,00р.	4
7	Building an installation project in AUTOCad	5 days	Wed 18.03.20	Wed 25.03.20	12 200,00р.	2
8	Installation cost estimates	2 days	Wed 25.03.20	Fri 27.03.20	3 367,00р.	7
9	Coordination with the customer	2 days	Mon 30.03.20	Tue 31.03.20	3 600,00р.	8
10	Conclusion of a contract for a csi	2 days	Tue 31.03.20	Thu 02.04.20	5 600,00р.	9
11	Creation and invoice for installation	1 day	Mon 06.04.20	Tue 07.04.20	19 760,00р.	10
12	Creating and invoicing equipment	1 day	Mon 06.04.20	Mon 06.04.20	19 760,00р.	10
13	Receive prepayment	0 days	Mon 06.04.20	Mon 06.04.20	300 000,00р.	11
14	Procurement for the purchase of linear parts	3 days	Mon 06.04.20	Wed 08.04.20	35 700,00р.	11
15	Procurement for Equipment	3 days	Mon 06.04.20	Wed 08.04.20	411 755,00р.	11
16	Hole punching	48 h	Mon 13.04.20	Mon 20.04.20	1 000,00р.	14

In a real world where everything is interconnected, information and related business processes, systems that serve the business process, information networks that combine everything into a single process, and personnel who functionally carry out their operation, processing and protection – all these are information assets that, like other important business (filled with value) financial and material assets.

In table 1, we present the cost of business process stages for a single customer order and the probabilistic amounts of damage from information threats (*condition example of enterprise that implements security information systems (SIS) for customers*)).

Tab.1: Distribution of the cost of the order according to the stages of the business process (condition example).

Np	Title	Start	Ending	Duration	Expenses	Summ of step
1	2	3	4	5	6	7
1	Application Acceptance	Tue 17.03.20	Tue 17.03.20	1 hour	400,00 ₽	400,00 ₽
2	Object Measurement	Tue 17.03.20	Wed 18.03.20	10 h	3 800,00 ₽	4 200,00 ₽
3	Equipment project SIS	Wed 18.03.20	Fri 20.03.20	2 days	6 400,00 ₽	10 600,00 ₽
4	Installation project	Wed 18.03.20	Thu 19.03.20	12 h	6 800,00 ₽	17 400,00 ₽
5	Equipment marking SIS	Mon 23.03.20	Tue 24.03.20	2 days	13 360,00 ₽	30 760,00 ₽
6	Mount mounting; Exel	Mon 23.03.20	Fri 27.03.20	5 days	13 400,00 ₽	44 160,00 ₽
7	Building an installation project SIS in AUTOCAD	Wed 18.03.20	Wed 25.03.20	5 days	12 200,00 ₽	56 360,00 ₽
8	Installation cost estimates	Wed 25.03.20	Fri 27.03.20	2 days	3 367,00 ₽	59 727,00 ₽
9	Coordination with the customer	Mon 30.03.20	Tue 31.03.20	2 days	3 600,00 ₽	63 327,00 ₽
10	Conclusion of a contract for a SIS	Tue 31.03.20	Thu 02.04.20	2 days	5 600,00 ₽	68 927,00 ₽
11	Creation and invoice for installation	Mon 06.04.20	Tue 07.04.20	1 day	19 760,00 ₽	88 687,00 ₽
12	Creating and invoicing equipment	Mon 06.04.20	Mon 06.04.20	1 day	19 760,00 ₽	108 447,0 ₽
13	Receive prepayment	Mon 06.04.20	Mon 06.04.20	3 days	300 000,00 ₽	
14	Procurement for the purchase of linear parts	Mon 06.04.20	Wed 08.04.20	3 days	35 700,00 ₽	144 147,0 ₽
15	Procurement for Equipment	Mon 06.04.20	Wed 08.04.20	3 days	411 755,00 ₽	555 902,0 ₽
16	Hole punching	Mon 13.04.20	Mon 20.04.20	48 h	1 000,00 ₽	556 902,0 ₽
17	Trunking	Mon 20.04.20	Mon 27.04.20	48 h	500,00 ₽	557 402,0 ₽
18	tightening wires into pipes	Mon 27.04.20	Mon 27.04.20	1 day	2 840,00 ₽	560 242,0 ₽
19	Linear Testing	Mon 27.04.20	Tue 28.04.20	2 days	4 000,00 ₽	564 242,0 ₽
20	Surrender under the Act of the linear part	Wed 29.04.20	Fri 01.05.20	2 days	3 520,00 ₽	567 762,0 ₽
21	Turnstile mounting	Wed 29.04.20	Thu 30.04.20	12 h	2 860,00 ₽	570 622,0 ₽
22	Card Reader Installation	Wed 29.04.20	Mon 04.05.20	3 days	9 720,00 ₽	580 342,0 ₽
23	Installation of CPU and other equipment	Mon 04.05.20	Thu 07.05.20	4 days	12 000,00 ₽	592 342,0 ₽

24	ACS settings	Mon 11.05.20	Wed 13.05.20	20 h	9 100,00 ₽	601 442,0 ₽
25	System testing SIS	Tue 12.05.20	Thu 14.05.20	2 days	6 480,00 ₽	607 922,0 ₽
26	Commissioning facility	Mon 18.05.20	Tue 19.05.20	12 h	7 800,00 ₽	615 722,0 ₽
27	Final settlement for ACS SIS	Mon 18.05.20	Wed 20.05.20	3 days	₽ 615,722.00	

Next table 2 shows the listed threats and their amounts of probable damages to the business process of a commercial organization.

The formula for calculating possible damages (column 6) is equal to the total amount of the client's order *615,722 thousand rubles* minus the cost of the business process stage.

From table 2, we can see that the amount of damage falls to zero by the end of the business process itself, which suggests that as the client's order is completed, the firm can also reduce future information security costs. The client did not refuse our contractor, and the order will bring the company a margin (profit) from its execution.

The cost of completing the order will be paid by the client, and the revenue will be received by the company.

However, the most important issue is the rational amount of costs at the first stages – when the client appears at the company, receiving his personal data, which is especially important for competitors (in case of possible «poaching»), and accordingly all stages of pre – «processing» the order, when our company will spend its resources (assets) and competence (visiting the customer, measurements, and development of a sketch of the future order, cost calculation, order approval and cost coordination, etc. types of work) that will be performed before the first receipt of an advance amount equal to half the cost of the order (Mandritsa I., Peleshenko V., Mandritsa O., Fensel A., Petrenko V., Solovieva I., Mecella M., 2018).

We believe that the rationality of information security will be calculated as a certain threshold for reducing the probability of losing a client and, accordingly, not receiving the amounts of income (revenue) and compensating our company for the costs incurred at the first stages of the business process. In table 2, these amounts are shown in positions 1 through 12.

In the position 13 we consider the information security threshold for the given example of a business process, when the probability of subsequent damages is reduced by half.

Tab 2 : Amounts of probable damages to the organization's business process by process stage

Npp	Title	Duration	Expenses	Summ Damage of step	Summ Damage for full process
1	2	3	4	5	6
1	Application Acceptance SIS	1 hour	400,00 ₽	400,00 ₽	615722
2	Object Measurement	10 h	3 800,00 ₽	4 200,00 ₽	615322
3	Equipment project SIS	2 days	6 400,00 ₽	10 600,00 ₽	611122
4	Installation project	12 h	6 800,00 ₽	17 400,00 ₽	600522
5	Equipment marking SIS	2 days	13 360,00 ₽	30 760,00 ₽	583122
6	Mount mounting; Exel	5 days	13 400,00 ₽	44 160,00 ₽	552362
7	Building an installation project SIS in AUTOCAD	5 days	12 200,00 ₽	56 360,00 ₽	508202

8	Installation cost estimates	2 days	3 367,00 ₽	59 727,00 ₽	451842
9	Coordination with the customer	2 days	3 600,00 ₽	63 327,00 ₽	392115
10	Conclusion of a contract for a security SI	2 days	5 600,00 ₽	68 927,00 ₽	328788
11	Creation and invoice for installation	1 day	19 760,00 ₽	88 687,00 ₽	259861
12	Creating and invoicing equipment	1 day	19 760,00 ₽	108 447,00 ₽	171174
13	Receive prepayment SIS	3 days	300 000,0 ₽		
14	Procurement for the purchase of linear parts	3 days	35 700,00 ₽	144 147,00 ₽	59820
15	Procurement for Equipment	3 days	411 755,0 ₽	555 902,00 ₽	58820
16	Hole punching	48 h	1 000,00 ₽	556 902,00 ₽	58320
17	Trunking	48 h	500,00 ₽	557 402,00 ₽	55480
18	tightening wires into pipes	1 day	2 840,00 ₽	560 242,00 ₽	51480
19	Linear Testing SIS	2 days	4 000,00 ₽	564 242,00 ₽	47960
20	Surrender under the Act of the linear part	2 days	3 520,00 ₽	567 762,00 ₽	45100
21	Turnstile mounting	12 h	2 860,00 ₽	570 622,00 ₽	35380
22	Card Reader Installation	3 days	9 720,00 ₽	580 342,00 ₽	23380
23	Installation of CPU and other equipment	4 days	12 000,00 ₽	592 342,00 ₽	14280
24	ACS settings	20 h	9 100,00 ₽	601 442,00 ₽	7800
25	System testing SIS	2 days	6 480,00 ₽	607 922,00 ₽	6600
26	Commissioning facility	12 h	7 800,00 ₽	615 722,00 ₽	0
27	Final settlement for ACS SIS	3 days	615,722.00		615722

Table 3 and 4 present calculations of the required amounts for information security for 3 levels of information security when a firm hires specialists of various competencies who will provide these levels of security in accordance with the firm's own understanding of the probability that their firm will be attacked by an information attack to poach a client or other probabilistic information problems for the firm.

Tab.3: Parameters of labour costs for information security specialists (when hiring them for 1 month)

Cost of defence approach	Strong level	Middle level	Ligth level
1	2	3	4
Salary of an information security specialist according to the choice of security level (Rouble/Hour) * 300 hours for full complex defence work for task	1500	940	220
Salary of an information security specialist according to the choice of security level (Rouble/Month) 160 hours a month	240000	150000	35000
Coefficient of competence for task of defence work (at light level)	0.3	0.6	1.0

At the same time, we assume (as of the current date of writing this article) that specialists differ not only in the level of remuneration for their work, but also in the time they need for a standard set of information and support actions to ensure a decent level of security.

A particularly important part of the business process for each organization is a block-a link known as «settlement and cash services» and its possible «damage», such as the downtime of receiving an advance or the total income from completing a business task (Ahmed, N., Matulevičius, R, 2015).

From the point of view of information cyber threats, at the moment 01-03-2020, any target virus that has penetrated the company's systems at the entry points, namely, any company's equipment can be potentially vulnerable, can cause both denial of service of cash systems, and illegitimate transfer of funds to the «fake accounts» of the attacker (Cjaputa K., 2016).

At the same time, the stage known in the business process of any organization as the approval of the design layout of the customer's project and any target virus that has penetrated the company's systems at the entry points will lead to downtime for calculating and approving the start of installation work under this client agreement.

Tab. 4: Calculations of the company's expenses for the types of information work of an information security specialist (when hiring him for 1 month), taking into account his competence coefficient

№	Cost of defence approach	Strong level	Middle level	Ligth level
1	2	3	4	5
All departure	Salary of an information security specialist according to the choice of security level (Rouble/Hour) * 300 hours	1500	940	220
	Microsoft Windows Server 2016 Standard 64-bit Russian 1pk DSP OEI DVD 16 Core ~70000 roubles.)	70000	70000	70000
Office-manager	Installing access control (Rouble/Hour) * 30 hours	13500	16920	6600
	Adequate security policy (Rouble/Hour) * 20 hours	9000	11280	4400
	Creating a threat model (Rouble/Hour) * 20 hours	9000	11280	4400
Accounting office	Installing access control (Rouble/Hour) * 150 hours	67500	84600	33000
	Adequate security policy (Rouble/Hour) * 100 hours	45000	56400	22000
	Creating a threat model (Rouble/Hour) * 50 hours	22500	28200	11000
Clients Data	Installing access control (Rouble/Hour) * 50 hours	22500	28200	11000
	Adequate security policy (Rouble/Hour) * 30 hours	13500	16920	6600
	Creating a threat model (Rouble/Hour) * 20 hours	9000	11280	4400
Auto Cad	Installing access control (Rouble/Hour) * 100 hours	45000	56400	22000

	Adequate security policy (Rouble/Hour) * 50 hours	22500	28200	11000
	Creating a threat model (Rouble/Hour) * 25 hours	11250	14100	5500
Stock (manager)	Installing access control (Rouble/Hour) * 100 hours	45000	56400	22000
	Adequate security policy (Rouble/Hour) * 50 hours	22500	28200	11000
	Creating a threat model (Rouble/Hour) * 20 hours	9000	11280	4400
Production	Installing access control (Rouble/Hour) * 50 hours	22500	28200	11000
	Adequate security policy (Rouble/Hour) * 30 hours	13500	16920	6600
	Creating a threat model (Rouble/Hour) * 20 hours	9000	11280	4400
Service of firm	Installing access control (Rouble/Hour) * 20 hours	9000	11280	4400
	Adequate security policy (Rouble/Hour) * 10 hours	4500	5640	2200
	Creating a threat model (Rouble/Hour) * 5 hours	2250	2820	1100

The main reason will be a denial of service after a local target attack on the company's information system, which allows replacing a legitimate project (accepted for execution in the business process of the organization) with a false one.

Also known as the «calculation» stage of the business process, after a local target attack on the company's information system, the customer will fail to fulfil the contract for the order, in the form of downtime at the stages of installation, installation of equipment, for calculating and approving the start of installation work (Leitner, M., Miller, M., Rinderle-Ma, 2013).

Here, information security is no less important from attacks on intermediary companies or suppliers, through the introduction of «fake-contracts» into the organization's information system and into the business process of the organization.

It is also necessary to highlight possible forced production downtime from third parties (power supply, lack of components, etc.) when commissioning a customer's order, through local DDoS attacks on IP telephony (Rodriguez A., Fernandez-Medina E,. Piattim M, 2007).

Next table 5 shows the results of the cost of information security specialists according to the list of works to protect the company's business process.

Tab.5: Calculations of the company's costs for the types of information work of an information security specialist by links (participants) of the business process

Departure of enterprises	Strong level	Middle level	Ligth level
1	2	3	4
Office-manager	31500	39480	15400
Accounting office	135000	169200	66000
Clients Data	45000	56400	22000
AutoCad desiner	78750	98700	38500
Stock (manager)	76500	95880	37400

Production	45000	56400	22000
Service of firm	15750	19740	7700
All departure	70000	70000	70000
Summ	497500	605800	279000

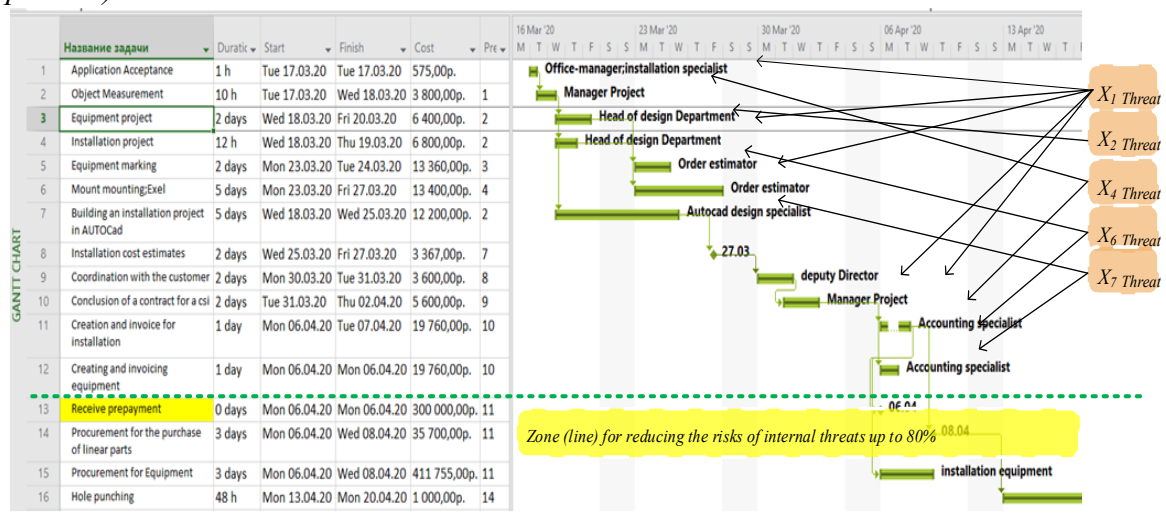
It is also likely that there will be no access to the component database, changes to its contents, or a sudden failure to provide components, and as a result, a delay in the start of installation or commissioning of the order.

Tab. 6: Probability of leakage risks for threat model of the business process

№	Business process phase	Likelihood of threats through technical channels	Likelihood of threats from staff	Total probability
1	2	3	4	5
1	Accepting the buyer's application SIS	0	0,3	0,3
2	Froze the order object SIS	0,01	0,04	0,05
3	Order equipment project SIS	0,01	0,29	0,3
4	Order mounting project SIS	0,01	0	0,01
5	Making estimates for equipment SIS	0,01	0,8	0,81
6	Making estimates for installation; MS Excel	0,01	0	0,01
7	Building a montage project in AUTOCAD	0,01	0,19	0,2
8	Building an estimate of the installation	0,01	0,04	0,05
9	Reconciling with the customer	0,01	0,29	0,3
10	Conclusion of a contract on security SIS	0,01	0,09	0,1
11	Create and bill for installation SIS	0,01	0,09	0,1
12	Create and invoice for equipment SIS	0,01	0	0,01
13	Receiving prepayment from the buyer SIS	0	0	0

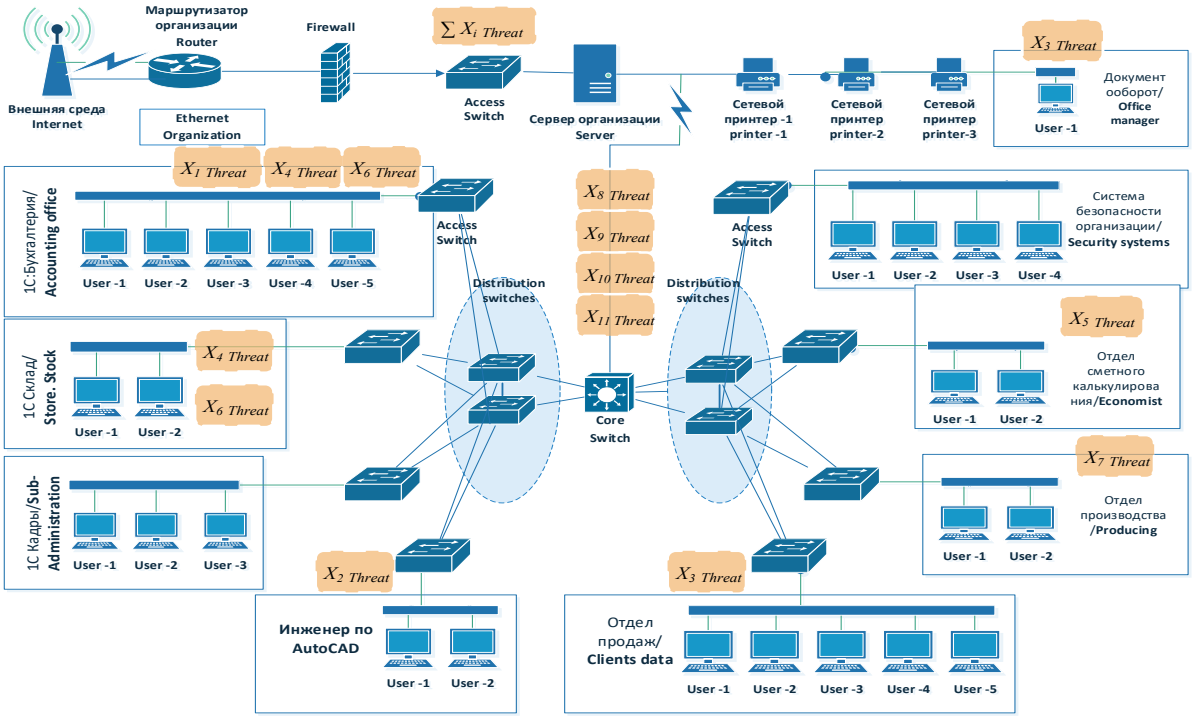
Next, figure 2 and 3 shows the localization of information threats to the business process of organizations that perform the task «Implementing a project on the client's order» with the indication of the risk zones of malicious penetration.

Fig 3 : Information security internal risk Zones of a business task (as part of a business process)



Also, well-known threats to the business process of the organization are-disruption of the delivery of components for the execution of the order. We will also mention a possible technical failure (failure) of the customer's order installed after installation (dismantling time, subsequent alterations and waiting for delivery of new components) (Belov et al., 2018).

Fig. 4: Information security risk Zones of a business task (in correlation with a business process fig. 3)

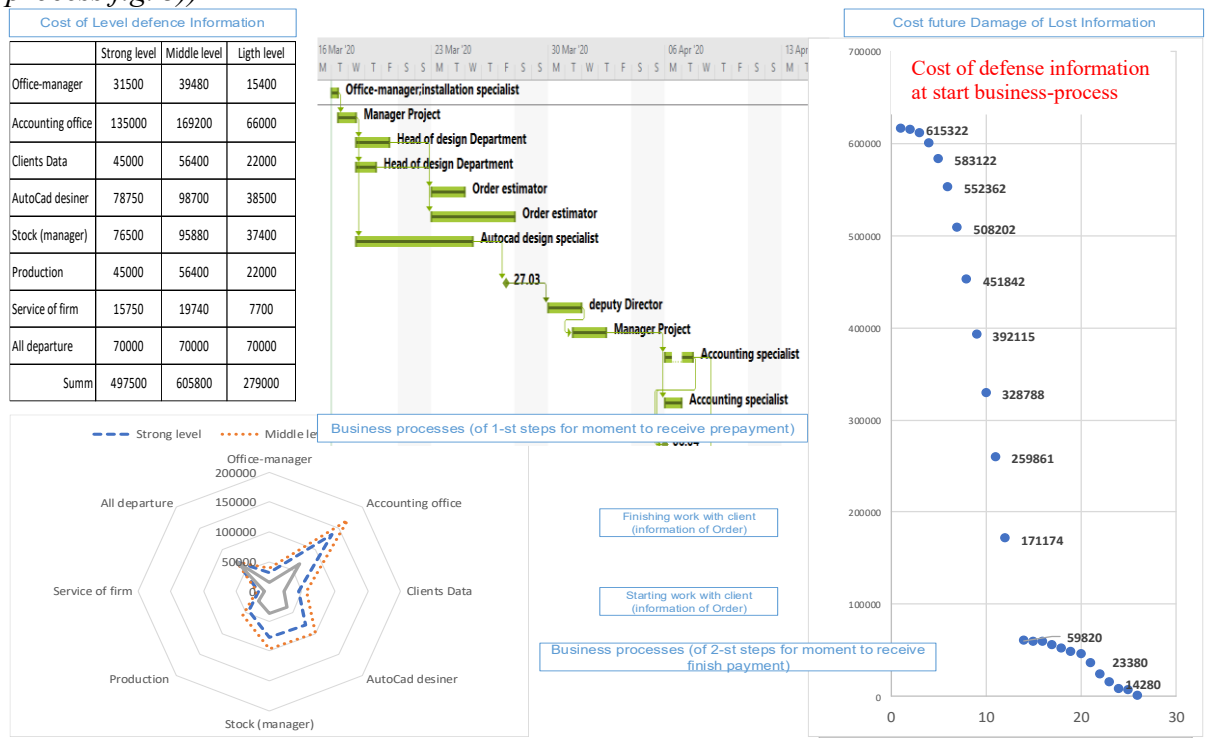


The most likely threats include infection of all information assets of the company and destruction of data (including all existing projects, documents, reports and other stages, and key factors of the business process) (Gordon et al., 2020).

The least likely information threats, but quite feasible in the conditions of competition for a niche of the organization's market will be a complete violation of the information integrity of the software OR hardware of the organization, according to the well-known example of Tailored Access Operations, with the possibility of interception of management or destruction of data [11]. The creation of a «fake-counterpart of the customer» (or a targeted attack for a long time with the subsequent destruction of the company) should be attributed to completely «unrealistic», but still considered information threats (Florêncio, D., Herley, C, 2013).

We will summarize the presented classifications of information threats and reflect these factors in the modernized threat model in figure 5.

Fig. 5 : Implementation of costs for information measures to protect the organization's business process (according to ISO/IEC 27001:2013 (E) and in correlation with a business process fig. 3))



3. Conclusions

An organization's information assets are subject to both deliberate and accidental threats, and their associated processes, systems, networks, and people have inherent vulnerabilities.

Changes to business processes and systems or other external changes (such as new laws and regulations) may create new risks to information security. Therefore, given the many ways in which threats of cybercriminals, using vulnerabilities of «stages» and «blocks» of the business process, can harm the organization, we can confidently say that information security risks are always present. The information security management system (isms for short), as defined in ISO/IEC 27001, provides a holistic, consistent view of the organization's information security risks in order to implement a comprehensive set of measures to ensure information security within an integrated management system.

Many information systems were developed without taking into account security requirements in the context of ISO / IEC 27001 and this standard. Security provided only by technical means is limited and must be supplemented by appropriate management and

procedures. Determining which tools to use in a particular case requires careful planning and attention to detail. The successful functioning of the ISMS requires the support of all employees of the organization.

This may also require the participation of shareholders, suppliers, or other external parties. You may also need advice from outside experts. Thus, we can begin to develop a sufficient information security strategy within the framework of the proposed threat model for the object of research.

4. Resources

AHMED, N. (2014). Deriving security requirements from business process models. Tartu Estonia; University of Tartu Press. Ph.D. thesis, ISBN 978-9949-32-717-1

AHMED, N., & MATULEVIČIUS, R. (2015). Presentation and validation of method for security requirements elicitation from business processes. *Lecture Notes in Business Information Processing*, *204*(CAiSE Forum 2014), 20–35. https://doi.org/10.1007/978-3-319-19270-3_2

BELOV, V., PESTUNOV, A., & PESTUNOVA, T. (2018). On the issue of information security Risks assessment of business processes. *2018 XIV International Scientific-Technical Conference on Actual Problems of Electronics Instrument Engineering (APEIE)*, *8*, 136–139. https://doi.org/10.1109/apeie.2018.8545576

BISSELL, K., & PONEMON, L. (2020). THE COST OF CYBERCRIME NINTH ANNUAL COST OF CYBERCRIME STUDY. Traverse City; Ponemon Institute LLC.

CJAPUTA, K. (2016, January 15). *Business processes based introduction of security aspects in enterprise architecture*. nda.rtu.lv. https://nda.rtu.lv/en/view/14843.

FLORÊNCIO, D. & HERLEY, C. (2012). Sex, lies and cyber-crime surveys. *Economics of Information Security and Privacy III*, *3*, 35–53. https://doi.org/10.1007/978-1-4614-1981-5_3 Springer, New York

GORDON, L. A., LOEB, M. P., & ZHOU, L. (2020). Integrating cost–benefit analysis into the NIST CYBERSECURITY framework via the GORDON–LOEB MODEL. *Journal of Cybersecurity*, *6*(1). https://doi.org/10.1093/cybsec/tyaa005

INFOWATCH COMPANY, I. W. A. C. (2018). (rep.). *Data Breach Report: A Study on Global Data Leaks in H1 2018* (Vol. 1, p. 22). Moscow, Russian Federation: InfoWatch Company.

ISO/IEC JTC 1/SC 27 Information security, cybersecurity and privacy protection. (2018, February 7). *Iso/iec 27000:2018*. ISO. https://www.iso.org/standard/73906.html.

LEITNER, M., MILLER, M., & RINDERLE-MA, S. (2013). An analysis and evaluation of security aspects in the business process model and notation. *2013 International Conference on Availability, Reliability and Security*, *In: Proceedings of the Eighth International Conference on Availability, Reliability and Security*, 262–267. https://doi.org/10.1109/ares.2013.34

MANDRITSA, I., PELESHENKO, V., FENSEL, A., MECELLA, M., PETRENKO, V., SOLOVIEVA, I., MANDRITSA, O., & TEBUEVA, F. (2018, October). Defining a cybersecurity strategy of an organization: criteria, objectives and functions. Stavropol-Dombay; North-Caucasian Federal University. International Multidisciplinary Symposium ICT Research in Russian Federation and Europe. 2018

SALNITRI, M., DALPIAZ, F., & GIORGINI, P. (2014). Modeling and Verifying security policies in business processes. *Enterprise, Business-Process and Information Systems Modeling*, *175*, 200–214. https://doi.org/10.1007/978-3-662-43745-2_14

RODRIGUEZ, A., FERNANDEZ-MEDINA, E., & PIATTINI, M. (2007). A bpmn extension for the modeling of security requirements in business processes. *IEICE Transactions on Information and Systems*, *E90-D*(4), 745–752. https://doi.org/10.1093/ietisy/e90-d.4.745

CLASSIFICATION AND CLUSTER ANALYSIS OF BIG DATA ECOSYSTEM IMPLEMENTATION

Martin Misut[1]

Abstract
A significant trend in the introduction of big data is currently associated with building the necessary technological infrastructure. There are several popular solutions available on the market. The paper presents an analysis of the acceptance of individual solutions depending on the selected characteristics of the company. Two types of analysis were used: cluster and classification analysis in the form of decision trees. The results obtained from both methods differ slightly, which is probably due to a small sample of data. The results take the form of a model that can predict/recommend selecting a specific solution for a given company depending on its characteristics.

Keywords
big data, technology, infrastructure, big data ecosystem, classification, cluster analysis

1. Introduction

The flood of new data goes hand in hand with continuous technological progress. There is so much data, data storage is becoming more accessible and reaching such a capacity that data can be collected by anyone who can and about everything he has access to and sees potential. The data and information hidden in them are crucial to many industries, especially the telecommunications, financial sector, online sales and services. These data often come from different sources and come in different forms, often unstructured data obtained from monitoring people's activity on social networks or elsewhere in the online Space, or vast amounts of data coming in rapid succession.

The usual goal of companies is to improve financial results, either directly - by reducing costs or indirectly - by creating a strategy or platform for data management and analysis to increase their competitiveness or the quality of services provided. The prerequisite for a successful future thus becomes the implementation of big data. (Kościelniak & Puto, 2015) It is not an easy task, as it requires extensive investments in technical infrastructure, human resources development and business processes.

"Big data" solutions for companies are often very complex and costly, and managerial decisions to adopt "big data" solutions are mostly strategic decisions with a view to the future, rather than short-term solutions to temporary problems. (Jeble, Kumari, & Patil, 2018) Therefore, companies make great efforts and costs in the research and decision-making phase to make the right decision built on solid foundations and can be implemented and achieve the required level of profitability. From the perspective of investing in new solutions, there are not many worse scenarios than wasting a considerable amount of funds for a solution that will later prove to be unfeasible or unprofitable. (Nasrollahi, Ramezani, & Sadraei, 2020)

[1] University of Economics in Bratislava, Faculty of Economic Informatics, Department of Applied Informatics, martin.misut@euba.sk, ORCID: 0000-0002-5545-2624

For these reasons, implementing big data is slower and is accompanied by several unsuccessful attempts. Managers to reduce the risk with such a significant investment as big data undoubtedly are, trying to use all possibilities to refine decisions about big data implementation. One of the mentioned possibilities is to use information about already implemented big data projects. Therefore, the research subject described here was the analysis of available information on successfully implemented big data projects so that the information obtained can be used to design the decision models to facilitate managers' decisions in new big data projects. Specifically, to find out what solutions are typical for individual categories of companies. Also, to propose a method enabling the use of experience from the successful implementation of big data projects.

The article is further organized as follows: In the Background section, the basic concepts and characteristics of the big data ecosystem are presented. The following section briefly describes the essential big data solutions used in practice. The method section describes the procedure used to analyze the obtained data and the design of decision models. The results section discusses the obtained results.

2. Background

While the term "big data" has been defined in various literary sources in various ways, there is still no generally accepted definition. (Beyer & Laney, 2012) Intuitively, according to the authors (Athamena & Houhamdi, 2018), we understand that the term "big data" refers to a set of new information that must be available to many users in near real-time and that comes from vast amounts of data to accelerate critical decision-making processes.

In order to use the potential of Big Data, it is necessary to have suitable technology for the analysis of large amounts of data. Because Big Data is a continuation of the "traditional" data analysis, Big Data technologies should fit into the existing enterprise IT environment. For this reason, it is helpful to have a common structure that explains how Big Data complements and differs from existing systems, Business Intelligence, databases, and other solutions. Such a common structure is called reference architecture (RA). An RA is an abstract software architecture based on one or more domains and with no implementation features. (Moreno,

Fernández, Serrano, & Fernández-Medina, 2019)

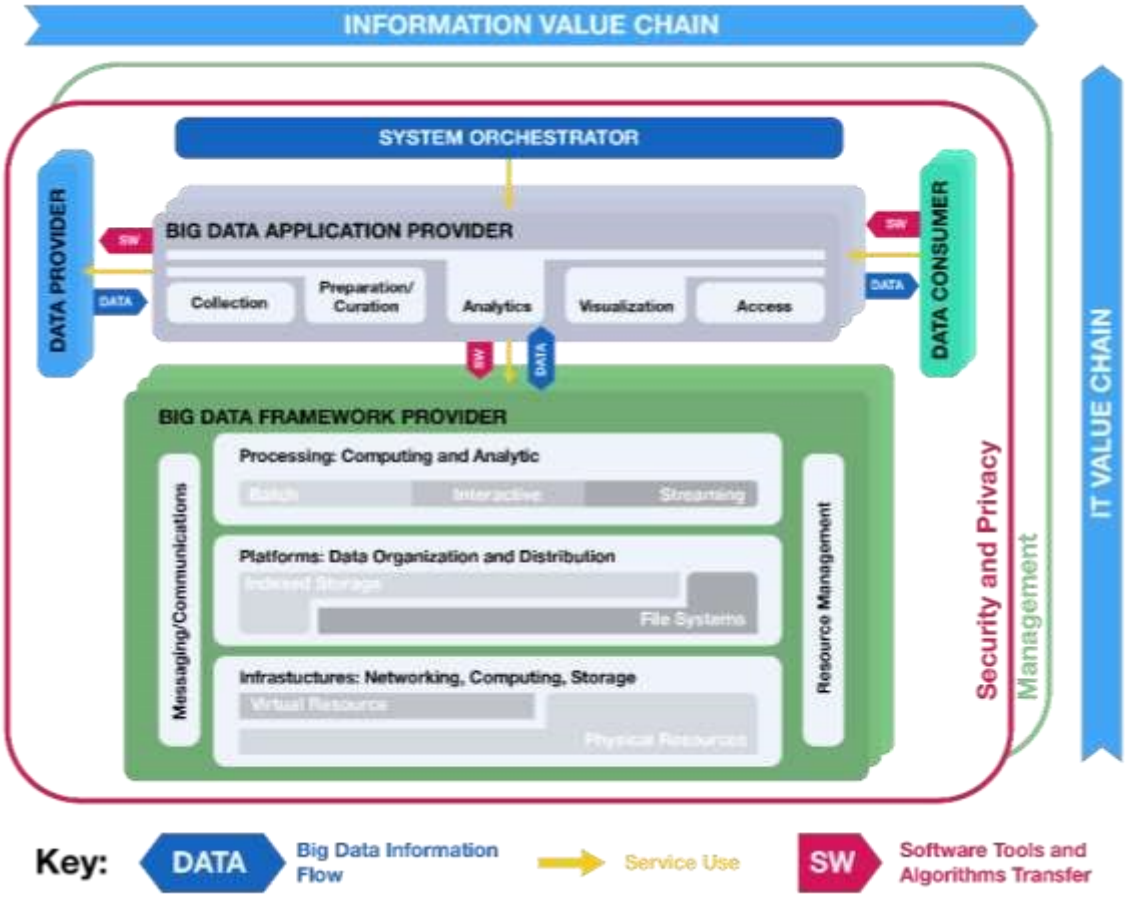
For the last several years, the NIST has defined an RA for Big Data which has received the consensus of the industry and scientific community. (NIST, 2021) This big data reference architecture is shown in Figure 1. It represents a big data system composed of five logical functional components or roles connected by interoperability interfaces (i.e. services) representing the different categories of technical activities and functional components within a Big Data system. In order to apply the NBDRA to a particular system, it is necessary to construct architecture views of these activities and the functional components that implement them. (NIST, 2021)

3. Big data ecosystem definition

Big Data is not just a question of data storage, although a database or Hadoop forms the essential technologies and components for extensive data processing and analysis. Big data includes various components for storing, processing, visualizing and delivering results to data consumers. This whole complex of interrelated components can be defined as the Big Data Ecosystem (BDE), which deals with evolving data, models and support infrastructure throughout the big data lifecycle. (Demchenko, Laat, & Membrey, 2014) There are several structural models of the big data ecosystem, for example (Bashari Rad & Ataei, 2017), (Hamid,

2020), (Singh, Behera, & Mantri, 2018). However, the key is to know that infrastructure technologies are at the heart of the big data ecosystem and process, store and often analyze data. For decades, companies have relied on relational databases to process structured data. However, the volume, speed, and variety of current data means that relational databases often cannot provide the performance and latency needed to process large and complex data. In particular, the increase in unstructured data meant that data collection and storage had to be transformed. This has created new infrastructure technologies that are capable of transmitting vast amounts of data and allow applications to run on systems with thousands of nodes, potentially encompassing thousands of terabytes of data.

Fig. 1: NIST Big Data Architecture



Source: NIST 2021

4. Technologies and solutions for processing big data

Data must be first received from sources, translated and stored, then analyzed before final presentation in an understandable format. It is a long and demanding process, which can take months or even years to implement. (Hamid, 2020) Figure 2 shows aggregated activities expressing a dynamic view of the big data ecosystem. These activities are provided by individual components of the big data ecosystem.

A brief description of some components of the Hadoop ecosystem is given below:(Prajapati & Patel, 2021)

Hadoop - A whole ecosystem of technologies designed for data storage, processing and analysis. Hadoop's core technologies work on the principle of dividing and distributing data into parts and their simultaneous analysis, rather than solving one monolithic block of data at once.

Hadoop is a set of programs and procedures - Hadoop is not a stand-alone product. It is a group of computer programs and systems that form a whole, cooperate and in many cases will not work fully without each other.

Hadoop has open-source code - a considerable advantage of Hadoop is that it is equally accessible to everyone thanks to open source code. There is no need to purchase licenses to use its basic version without additional add-ons from other manufacturers or service providers.

Hadoop can be a crucial element of high-volume operations - there are many different tools, programs, approaches and add-ons to Hadoop, but Hadoop itself is a vital part of the solution. At the same time, it is possible to build extensive and scalable solutions on Hadoop, the skeleton of which will not need to be replaced for a long time.

Hadoop usually includes the following five modules:

1. Hadoop Distributed File System (HDFS)
2. MapReduce - a highly efficient methodology for parallel data processing
3. YARN - assists in planning and monitoring tasks and managing available resources
4. Hadoop Common/libraries
5. Hadoop Ozone

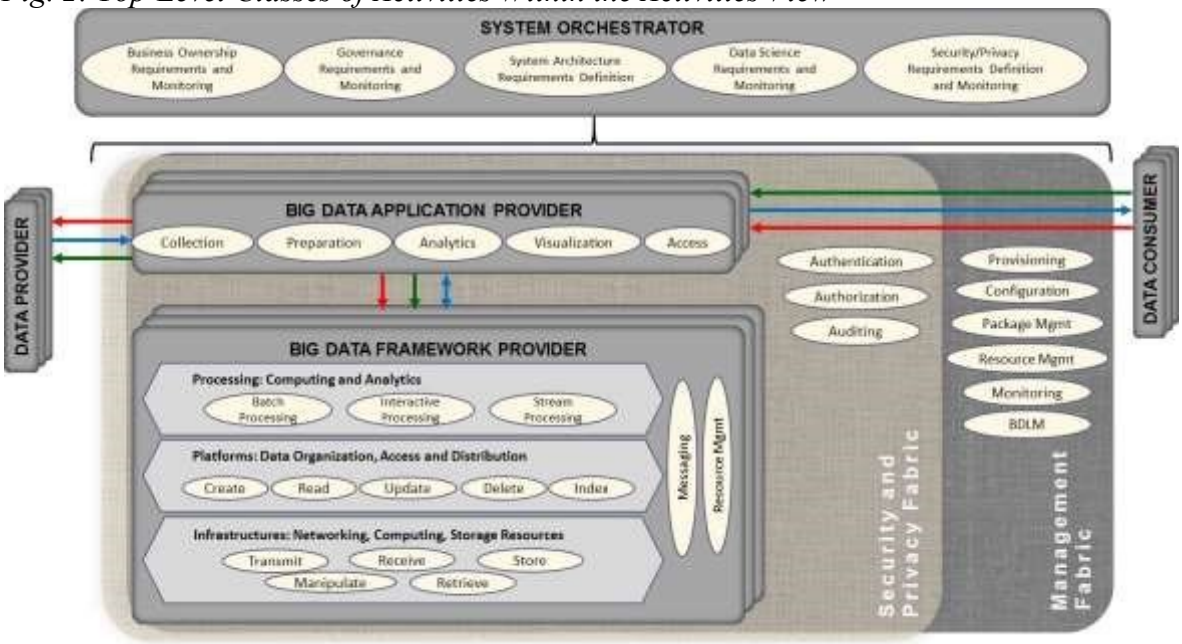
The key modules of Hadoop are the first two modules from the list - HDFS and MapReduce. However, Hadoop is far from forming only the mentioned five modules. A whole so-called Hadoop ecosystem includes many other tools and elements that can be used as needed. An essential capability of the Hadoop architecture is the ability to ensure continuous operation. Hadoop can ensure that operations are quickly moved to other free nodes if any of the nodes fail during operation or is temporarily unavailable.

Traditional approaches are still highly effective when storing data in a structured form and process transactions. The MapReduce / Hadoop architecture offers strong support for more complex and less structured data types and structures and is open source and often a more costeffective solution thanks to its open-source code.

NoSQL and MPP databases are also used to store unstructured data. NoSQL - means Not Only SQL; it is also involved in processing large amounts of multi-structured data. Most NoSQL databases are most competent in handling discrete data stored between structured data.

Some NoSQL databases, such as HBase, can work concurrently with Hadoop. (Oussous, Benjelloun, Ait Lahcen, & Belfkih, 2015) Massive Parallel Processing (MPP) Databases - MPP databases work by segmenting data between multiple nodes and processing these data segments in parallel and using SQL. While Hadoop is usually run on cheaper clusters of commodity servers, most MPP databases run on expensive specialized hardware.

Fig. 2: Top-Level Classes of Activities Within the Activities View



Source: NIST 2021

During our research, we found that there are a large number of tools and solutions for processing large volumes of data, and also that there are many vendors of various tools and solutions on the market, ranging from large systems that can meet many needs of big companies to specific solutions and products designed for specific problems. The most common solutions and largest resellers include Hadoop, Teradata, Spark, Cassandra, Amazon_ Web_Services, MapR_Data_Platform, Wherescape, Hitachi, etc.

5. Method

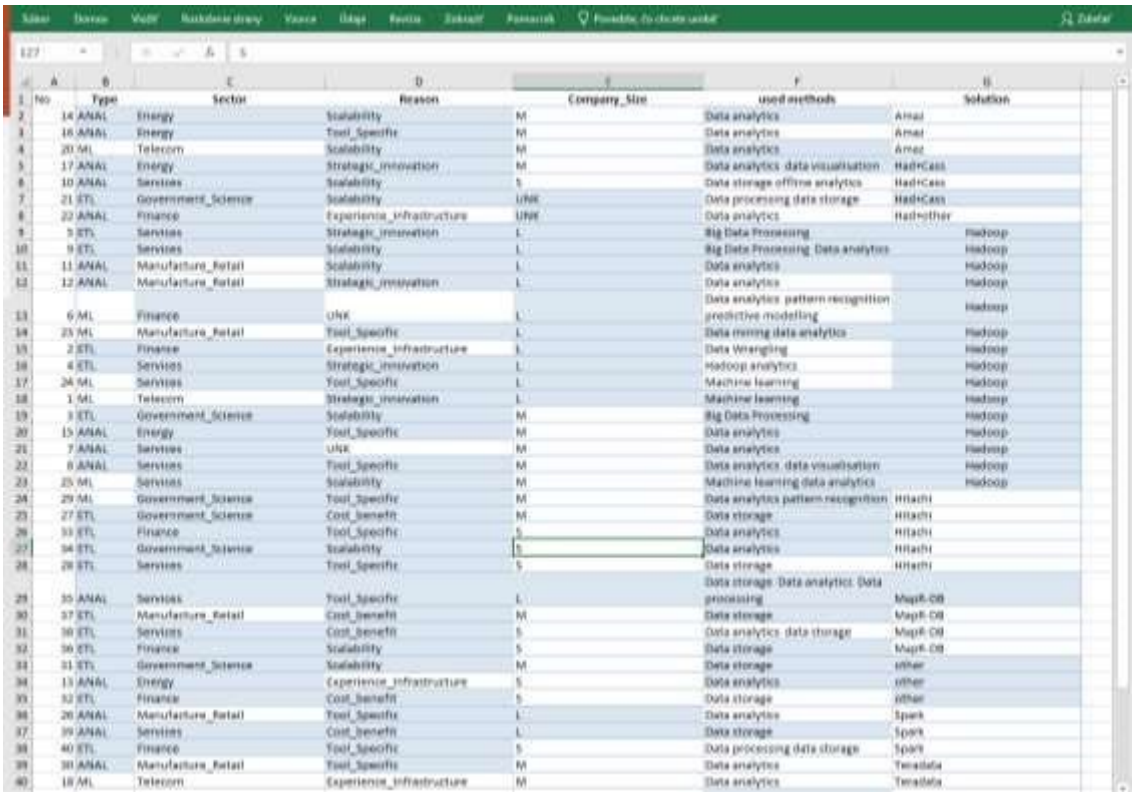
Because the goal of the research is based on the analysis of published information about the implemented projects of big data deployment, we used the following procedure to solve the research problem, respecting the logical sequence of individual steps:

1. Finding a relevant description of big data implementation
2. Datafication of individual cases
3. Determination of attributes characterizing individual cases
4. Cluster analysis
5. Creating a decision tree and its transformation into rules

6. Raw data collection

We extracted the information found from the written text from many sources using reading comprehension and transformed it into a structured form in the form of a table (Figure 3 shows the resulting table after the analysis of attributes). The result is a matrix (table), whose rows contain the values of individual attributes defining a specific case of big data implementation. The columns contain the values of the given attribute for different cases. The values of all attributes are coded as categorical so that they can be classified. Some attributes initially contained three numeric values, but we recoded them into categorical ones.

Fig. 3: Source data table



No	Type	Sector	Reason	Company_Size	used methods	Solution
2	14 ANAL	Energy	Scalability	M	Data analytics	Amaz
3	16 ANAL	Energy	Tool_Specific	M	Data analytics	Amaz
4	20 ML	Telecom	Scalability	M	Data analytics	Amaz
5	17 ANAL	Energy	Strategic_innovation	M	Data analytics data visualisation	HadrCass
6	10 ANAL	Services	Scalability	S	Data storage offline analytics	HadrCass
7	21 ETL	Government_Science	Scalability	UNK	Data processing data storage	HadrCass
8	22 ANAL	Finance	Experience_infrastructure	UNK	Data analytics	Hadrother
9	5 ETL	Services	Strategic_innovation	L	Big Data Processing	Hadoop
10	9 ETL	Services	Scalability	L	Big Data Processing Data analytics	Hadoop
11	11 ANAL	Manufacture_Retail	Scalability	L	Data analytics	Hadoop
12	12 ANAL	Manufacture_Retail	Strategic_innovation	L	Data analytics	Hadoop
13	6 ML	Finance	UNK	L	Data analytics pattern recognition predictive modelling	Hadoop
14	23 ML	Manufacture_Retail	Tool_Specific	L	Data mining data analytics	Hadoop
15	2 ETL	Finance	Experience_infrastructure	L	Data Wrangling	Hadoop
16	4 ETL	Services	Strategic_innovation	L	Hadoop analytics	Hadoop
17	24 ML	Services	Tool_Specific	L	Machine learning	Hadoop
18	1 ML	Telecom	Strategic_innovation	L	Machine learning	Hadoop
19	3 ETL	Government_Science	Scalability	M	Big Data Processing	Hadoop
20	15 ANAL	Energy	Tool_Specific	M	Data analytics	Hadoop
21	7 ANAL	Services	UNK	M	Data analytics	Hadoop
22	8 ANAL	Services	Tool_Specific	M	Data analytics data visualisation	Hadoop
23	25 ML	Services	Scalability	M	Machine learning data analytics	Hadoop
24	29 ML	Government_Science	Tool_Specific	M	Data analytics pattern recognition	Hitachi
25	27 ETL	Government_Science	Cost_benefit	M	Data storage	Hitachi
26	33 ETL	Finance	Tool_Specific	S	Data analytics	Hitachi
27	34 ETL	Government_Science	Scalability	S	Data analytics	Hitachi
28	28 ETL	Services	Tool_Specific	S	Data storage	Hitachi
29	35 ANAL	Services	Tool_Specific	L	Data storage Data analytics Data processing	MapR-DB
30	37 ETL	Manufacture_Retail	Cost_benefit	M	Data storage	MapR-DB
31	38 ETL	Services	Cost_benefit	S	Data analytics data storage	MapR-DB
32	36 ETL	Finance	Scalability	S	Data storage	MapR-DB
33	31 ETL	Government_Science	Scalability	M	Data storage	other
34	13 ANAL	Energy	Experience_infrastructure	S	Data analytics	other
35	32 ETL	Finance	Cost_benefit	S	Data storage	other
36	26 ANAL	Manufacture_Retail	Tool_Specific	L	Data analytics	Spark
37	39 ANAL	Services	Cost_benefit	L	Data storage	Spark
38	40 ETL	Finance	Tool_Specific	S	Data processing data storage	Spark
39	30 ANAL	Manufacture_Retail	Tool_Specific	M	Data analytics	Teradata
40	18 ML	Telecom	Experience_infrastructure	M	Data analytics	Teradata

Source: author

In searching for suitable historical records of large data processing solutions, we came across many ways and forms of expressing the information we were searching for. They were primarily expressed in verbal form within a coherent text, e.g. in published articles or case studies, customer reports, customer case studies, etc.

The collected data were published in the time interval 2014 - 2019. We managed to find more than 70 documented cases, but not all descriptions contained information that interested us. Excluding those cases that contained insufficient information, there are 44 documented descriptions in the final file.

7. Datafication

For datafication and attributes analysis, we used the method proposed in (Misut & Jurik, 2021). By analyzing the initial conditions of the assignment, we concluded that the ecosystem of big data would be affected mainly by two phenomena: the characteristics of the company and the characteristics of the tasks for which big data is planned to be used. This consideration was based on the assumption that implementing a specific solution (big data ecosystem) results from rational decisions of managers respecting the available resources of the company and taking into account the company's business goals. In terms of business resources, financial resources and human potential are critical to a big data ecosystem. Naturally, smaller companies (measured by the number of employees) cannot make significant investments due to their ability to raise sufficient capital resources. In comparison, multinational corporations are more likely to make large-scale investments. Therefore, we assumed that the company's size would play a role in choosing the big data ecosystem.

It should be noted that there is no equality between the size of the company measured by the number of employees and its capital adequacy. There is only a correlation in this respect. We consciously use this simplification here, as data on the number of employees were available in published case studies, but data on the capital adequacy of enterprises were not. Individual companies differ from each other by business processes that define their essence. In practice, it

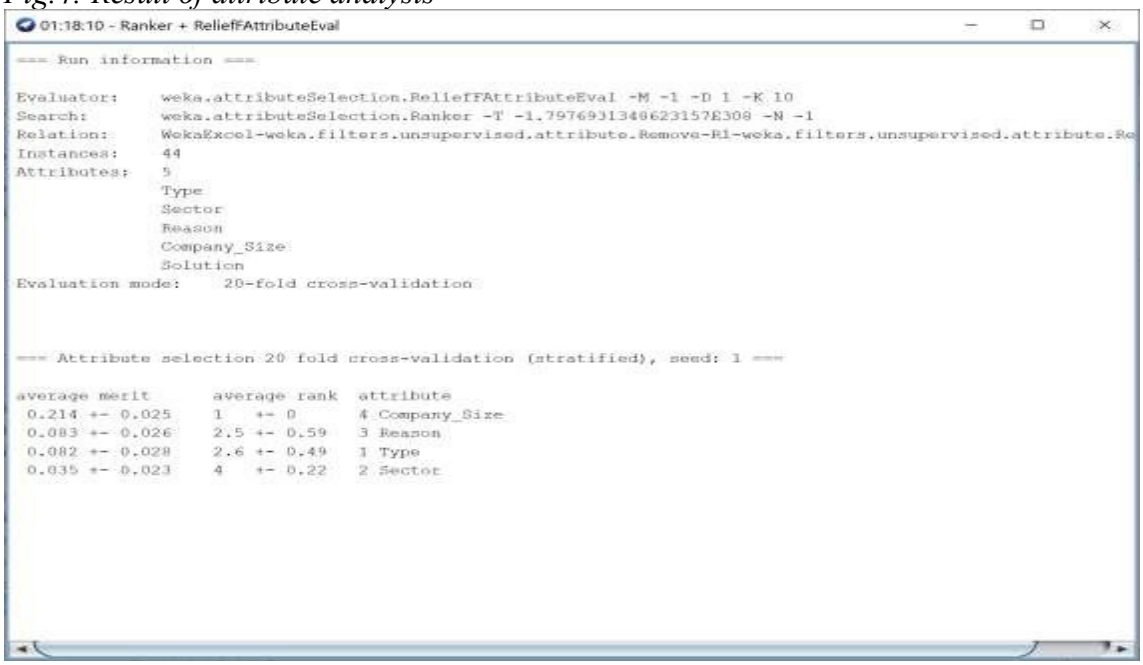
can be observed that companies belonging to one sector have similar characteristics of their business processes, and therefore it can be reasonably assumed that the tasks that cover these business processes are similar. For this reason, we have selected this attribute of the company (i.e. to which sector it belongs) as another attribute of the classification.

8. Attribute analysis

Due to the low number of documented cases, we had to narrow the research space significantly. We had to generalize the number of attributes and the values they can acquire in order to get several classes in the final that was appropriate for our needs and (so that each class is represented by a sufficiently relevant number of classified elements - for our needs, it would not make sense to have classes represented, for example, only 1-2 elements). For a small data set, it would be irrelevant if each of the examined elements were unique and if each attribute could acquire unique values at which it would be difficult to find a usable degree of correlation between the elements of the examined set of records.

We used Incremental Wrapper Subset Selection IWSS to reduce the number of attributes. This attribute selector is a part of Weka software in version 3.8.4. It first creates a ranking of attributes based on the selected metric. Then it runs an Incremental Wrapper Subset Selection over the ranking (linear complexity) by selecting attributes (using the WrapperSubsetEval class) which improve the performance for a given minimum number of folds out of the folds of the wrapper cross-validation. (Bermejo, 2020) As Attribute evaluation method, we chose RelieffAttributeEval, which Evaluates the worth of an attribute by repeatedly sampling an instance and considering the given attribute's value for the nearest instance of the same and different class. As a Search method, we used Ranker, which Ranks attributes by their individual evaluations.

Fig.4: Result of attribute analysis



Source: author

After analysis, from the original number of nine attributes in the first phase of data collection, we finally selected the following five attributes according to their ranking in the final model. The following are the attributes with allowed values. The number of described instances with the given attribute value is given in parentheses.

Company_Size - the size of a company/organization by the number of employees.:

- L - Large (companies/organizations with more than 10,000 employees) (15)
- M - Medium (companies / organizations with 1 001 to 10 000 employees) (18)
- S - Small (companies/organizations with up to 1,000 employees) (9)
- UNK (companies/organizations where the number of employees was unknown) (2)

Reason - the main reason for choosing the provider and/or instrument and reflects, in particular, the reason for choosing the provider and the instrument.

- Cost_benefit (cost-effective solution / cost reduction) (8)
- Experience_infrastructure (existing expertise and infrastructure in the company) (5)
- Scalability (scalability of the solution, long-term complex solution) (12)
- Strategic_innovation (position of the company's strategy for data, interest in innovation) (5)
- Tool_Specific (tool or solution suitable for the specific needs of the company) (12)

Type - generalization of the primary/priority nature of the problem/task, resp. customer requirements. The attribute can take the following values:

- ANAL - analysis of large volumes of data of any kind (15)
- ETL - Extract / Transform / Load operations, resp. data organization management
- (20)
- ML - use of machine learning algorithms to improve services (9)

Sector - industrial sector of the company. Big data is particularly widely used in telecommunications, advertising, Finance and insurance, healthcare, retail (especially online shops) and online services and entertainment sectors. The attribute can take the following values:

- Energy - energy sector (5)
- Finance - financial and banking sector, insurance and related industries (8)
- Government_Science - government sector, science and research (6)
- Manufacture_Retail - production and sales (7)
- Services - services of all kinds (including transport, online services, entertainment) (13)
- Telecom - Telecommunications (from the point of view of Big Data specific sector)
- (5)

Solution - the attribute describes the implemented solution/solution supplier and represents a dependent variable in the assembled models. The attribute values are as follows:

- Amaz (3)
- Had + Cass (3)
- Had + other (1)
- Hadoop (15)
- Hitachi (5)
- MapR-DB (4)
- Spark (3)

- Teradata (3)
- Wherescape RED (4)
- other (3)

These attributes, therefore, became a test sign when creating the decision tree. Leaf nodes represent implemented solutions suitable for the company, specified by attribute values.

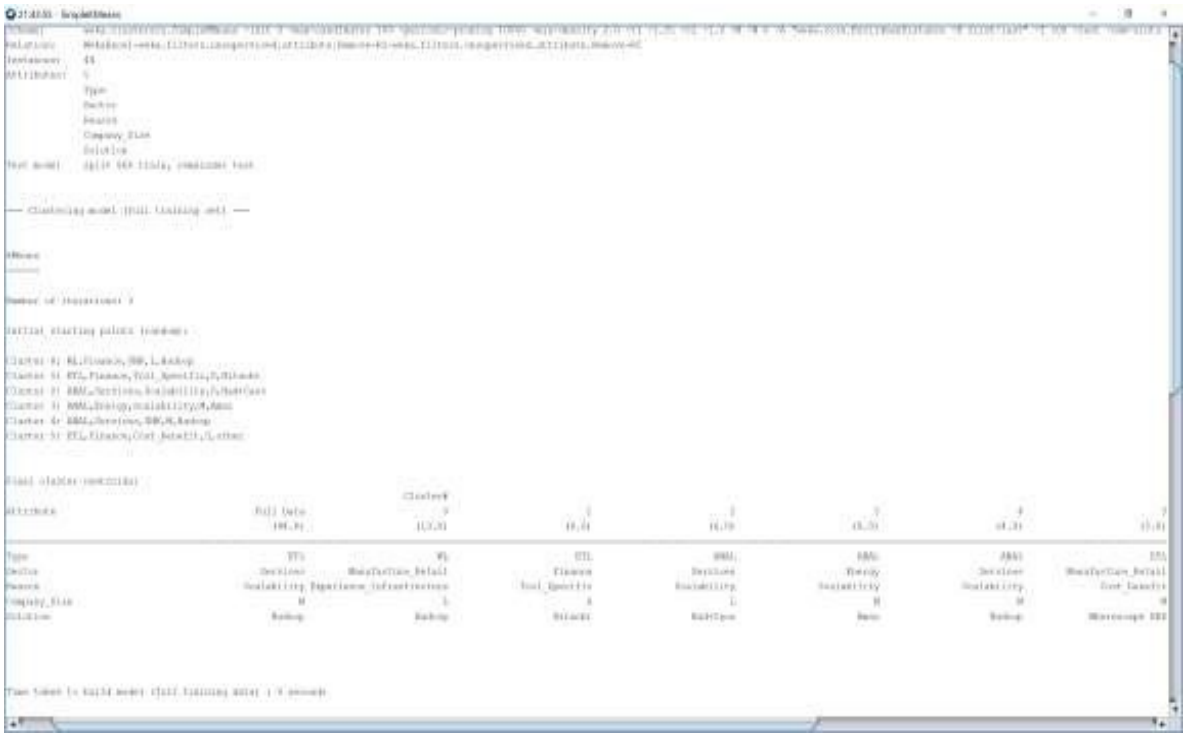
9. Cluster analysis

Cluster analysis is a type of data classification carried out by separating the data into groups. The aim of cluster analysis is to categorize n objects in k (k>1) groups, called clusters, by using p (p>0) variables. (Romesburg, 2004)

There are two main sub-divisions of clustering procedures. In the first procedure, the number of clusters is predefined. This is known as the K-Means Clustering method. When the number of clusters is not predefined, we use Hierarchical Cluster analysis. (Kettenring, 2006)

In our case, we used the k-means method with a percentage split of 66% as a training set and the remainder for a test. We chose the number of clusters so that the resulting clusters give a clear picture of the relationship between the values of the attributes and are not formed by a small number of elements. The basic idea of this method is to group individual companies into clusters based on the similarity of attribute values. The more similar the attribute values within the cluster, the more homogeneous the cluster. When considering the attribute representing the dependent variable, the cluster analysis will show which combination of individual attributes the resulting value of the attribute representing the big data solutions implemented in our case is typical. The result of the cluster analysis is shown in Figure 5.

Fig.5: Result of cluster analysis



Source: author

10. Classification analysis

The decision tree is a representation of the decision procedure for classifying cases into the appropriate classes. (Li, Xu, & Deng, 2019)

It is a graph structure in a tree containing a root node, non-leafy and leaf nodes. Nodes represent a class or testing sign. The edges represent the testing sign values. If the output variable is categorical, each leaf node represents one of the categories of the output variable classes. Then we talk about classification trees. (Song & Ying, 2015). An essential criterion in the decision tree algorithm is selecting the sign (attribute) to be tested in each tree's decision node. (Li et al., 2019) The aim is to select the attribute that best classifies the examples. A statistical property provides an excellent quantitative measure of an attribute's suitability called information gain, which indicates the extent to which an attribute divides training examples into its target classification. This measurement is performed by selecting from the candidate attributes at each step of the tree growth. (Myles, Feudale, Liu, Woody, & Brown, 2004) The finished tree structure can be rewritten into a set of decision (classification, production) rules.

Each classification rule contains a description of one path from the root node to a leaf node. These classification rules can then be used in a decision support system.

Statistical software packages (as SPSS, SAS, ...) can be used to create classification trees. We have decided to use the popular and, for this purpose, suitable tool Weka in version 3.8.4. Weka (Waikato Environment for knowledge analysis) is a suitable tool for performing many data mining tasks. According to Srivastava (2014), these include preprocessing data, selecting attributes, classification, clustering, and improving knowledge discovery using various meta classifiers. There are four steps involved in Weka for classification (Singhal & Jena, 2013), which we followed:

- Preparing the data
- Choose classify and apply an algorithm
- Generate trees
- The result or output.

Because we want to use the classification results in decision making, as Classifier, we used Decision_table with the parameter creation of decision rules. The result is then directly applicable in decision support systems. The Best_first search method was used to find out the good attribute combinations for the decision table. This search method searches the Space of attribute subsets by greedy hillclimbing augmented with a backtracking facility. A forwardlooking training and test set was used for classification. Eighteen rules are defined in the resulting decision table (see Fig.6). The *Reason* attribute was excluded during their construction, so the resulting rules have three attributes in the assumptions.

Fig.6: Result of classification analysis

```
22:20:11 - rules.DecisionTable

Number of training instances: 29
Number of Rules : 18
Non matches covered by IB1.
        Best first.
        Start set: no attributes
        Search direction: forward
        Stale search after 10 node expansions
        Total number of subsets evaluated: 14
        Merit of best subset found:   31.034
Evaluation (for feature selection): CV (leave one out)
Feature set: 1,2,4,5

Rules:
=============================================================================
Type                Sector              Company_Size       Solution
=============================================================================
ML                  Services            M                  Amaz
ETL                 Government_Science  S                  Amaz
ANAL                Services            M                  Hadoop
ML                  Telecom             L                  Amaz
ANAL                Manufacture_Retail  M                  Amaz
ETL                 Manufacture_Retail  M                  Amaz
ETL                 Finance             S                  Hitachi
ML                  Government_Science  M                  Amaz
ETL                 Services            L                  Hadoop
ETL                 Government_Science  M                  Hadoop
ANAL                Services            L                  Amaz
ETL                 Manufacture_Retail  L                  Amaz
ANAL                Manufacture_Retail  L                  Spark
ANAL                Energy              M                  Had+Cass
ANAL                Services            S                  Amaz
ETL                 Government_Science  UNK                Amaz
ANAL                Finance             UNK                Amaz
ML                  Telecom             M                  Teradata
=============================================================================


=== Predictions on test split ===

    inst#     actual  predicted error prediction
        1   10:other 2:Had+Cass    +    0.608
```

Source: author

11. Results

We used two different methods (cluster analysis and decision trees) to obtain relevant results. Both methods were to show us whether it is possible to observe typical features of implemented solutions for environments/companies characterized by similar properties. These results might be accepted conditionally, given the relatively small number of instances. If the number of instances increases, the validity of the results will also increase. However, the mentioned results already enable the basic orientation of managers when deciding on the choice of big data ecosystem for implementation in their company.

The results of the attribute analysis showed that the most robust distinguishing feature in the set of attributes is the company's size (see Fig. 4). Such a result seems logical. As mentioned above, the company's size correlates with the number of available resources that can be used to implement big data projects. Therefore, it is natural that larger companies implement more extensive and more expensive solutions compared to small companies. For this reason, even when interpreting the results, we will stick to the division of companies according to size.

Cluster analysis has shown that a solution using Hadoop is typical for large companies (L). This solution can also be recommended for medium-sized companies operating in the service sector, focusing primarily on big data analysis. In comparison, the proposed classification rules recommend using not only Hadoop as a solution for large companies but also an Amazon solution for the telecommunications sector using machine learning methods to improve its services.

Clusters 0 and 2, where is the dominance of large companies, together contain 48% of all elements. A typical sector of operation is Manufacture_Retail and Service sector, while in Manufature_Retail are typical uses of Machine learning methods, for the sector of services big data analysis. Medium-sized companies (M) are dominant in three clusters with a total percentage of 38% of elements. Typically, solutions focus on big data analysis and data organization management in the Energy, Services, and Manufacture_retail sectors. Recommended solutions are Amazon, Hadoop and Wherescape-RED. Instead of

Wherescape_RED, the decision rules contain recommendations for Teradata telecoms, and they are also characterized by machine learning to improve services, analogous to large telecom companies. The difference is in the recommended solution. Both methods show that there is the most significant diversification of solutions for medium-sized companies (3 clusters, eight rules), namely four solutions, while the source data contain 18 cases of medium-sized enterprises.

In comparison, solutions for large companies (2 clusters, five rules) are described in 15 cases, and three solutions are recommended (Hadoop, Amazon and Spark). Interestingly, in the described cases, Amazon's solution for large companies is not documented. From this point of view, the cluster analysis results appear to reflect the actual deployment cases better. We assume that this situation is due to the low number of documented cases and the related inaccuracy of the model, which was trained in only 29 instances.

For small companies (1 cluster, three rules, described in 9 instances), the recommended solution is Hitachi and Amazon, through the rules. A typical sector is Finance, and the task type is Extract / Transform / Load, resp. data organization management.

Tab. 1: Clusters

Cluster	**0**	**1**	**2**	**3**	**4**	**5**
no. of instances	13 (30%)	8 (18%)	6 (14%)	8 (18%)	4 (9%)	5 (11%)
Type	ML	ETL	ANAL	ANAL	ANAL	ETL
Sector	Manufacture_Retail	Finance	Services	Energy	Services	Manufacture_Retail
Reason	Experience_Infrastructure	Tool_Specific	Scalability	Scalability	Scalability	Cost_benefit
Company_Size	L	S	L	M	M	M
Solution	Hadoop	Hitachi	Had+Cass	Amaz	Hadoop	Wherescape RED

Source: author

12. Conclusion

By analyzing the obtained descriptions of big data implementation through two methods of analysis (cluster analysis and classification analysis), we obtained a basic idea of the implementation of different solutions depending on the specific situation and conditions, expressed through the values of four attributes. The results obtained for both methods differ

slightly due to the low number of cases described. Therefore, it is necessary to take into account the lower explanatory power of the results. Nevertheless, the results obtained in this form can form a starting point in managers' decisions about implementing big data in their company. Naturally, it is necessary to increase the accuracy and validity of the model, and in the future, it will be necessary to obtain a description of several successful implementations of big data.

13. Acknowledgements.

This article could have been created only thanks to the perseverance and patience of Ing. Daniel Talaj, who, within his diploma thesis under my guidance, collected and processed big data descriptions, which formed the basis for the analyzes described here.

This work was supported by a project VEGA No. 1/0373/18 entitled "Big data analytics as a tool for increasing the competitiveness of enterprises and supporting informed decisions" by the Ministry of Education, Science, Research and Sport of the Slovak Republic.

14. Resources

Athamena, B., & Houhamdi, Z. (2018). Model for decision-making process with big data. *Journal of Theoretical and Applied Information Technology, 96*, 5951-5961.

Bashari Rad, B., & Ataei, P. (2017). The big data Ecosystem and its Environs. *IJCSNS International Journal of Computer Science and Network Security, 17*(3), 38-42.

Bermejo, P. (2020). IWSS: Incremental Wrapper Subset Selection (Version 1.0.0): WEKA.

Beyer, M. A., & Laney, D. (2012). The importance of 'big data': a definition. *Stamford, CT: Gartner*, 2014-2018.

Demchenko, Y., Laat, C., & Membrey, P. (2014). *Defining architecture components of the Big Data Ecosystem*. Paper presented at the International Conference on Collaboration Technologies and Systems (CTS).

Hamid, A. (2020). A Functional View of Big Data Ecosystem. *International Journal of Computer Trends and Technology, 68*. doi:10.14445/22312803/IJCTT-V68I4P135

Jeble, S., Kumari, S., & Patil, Y. (2018). Role of Big Data in Decision Making. *Operations and Supply Chain Management: An International Journal, 11*, 36. doi:10.31387/oscm0300198

Kettenring, J. R. (2006). The practice of cluster analysis. *Journal of classification, 23*(1), 3-30.

Kościelniak, H., & Puto, A. (2015). BIG DATA in Decision Making Processes of Enterprises. *Procedia Computer Science, 65*, 1052-1058. doi:10.1016/j.procs.2015.09.053

Li, M., Xu, H., & Deng, Y. (2019). Evidential Decision Tree Based on Belief Entropy. *Entropy, 21*(9), 897.

Misut, M., & Jurik, P. (2021). *Datafication as a necessary step in the processing of big data in decision-making tasks of business*. Paper presented at the International Conference on Innovations in Science and Education, Prague, Czech Republic.

Moreno, J., Fernández, E., Serrano, M., & Fernández-Medina, E. (2019). Secure Development of Big Data Ecosystems. *IEEE Access, XX*, 1-18. doi:10.1109/ACCESS.2019.2929330

Myles, A. J., Feudale, R. N., Liu, Y., Woody, N. A., & Brown, S. D. (2004). An introduction to decision tree modeling. *Journal of Chemometrics: A Journal of the Chemometrics Society, 18*(6), 275-285.

Nasrollahi, M., Ramezani, J., & Sadraei, M. (2020). The Impact of Big Data Adoption on SMEs Performance. doi:10.21203/rs.3.rs-66047/v1

NIST, B. D. P. W. G. (2021). NIST Big Data Interoperability Framework:Volume 6, Reference Architecture,

Version 3, (October 2019). Retrieved from https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1500-6r2.pdf

Oussous, A., Benjelloun, F.-Z., Ait Lahcen, A., & Belfkih, S. (2015). *Comparison and Classification of NoSQL Databases for Big Data*. Paper presented at the International conference on Big Data, Cloud and Applications, Tetuan, Morocco.

Prajapati, M., & Patel, S. (2021). *A Review on Big Data with Data Mining*, Singapore.

Romesburg, C. (2004). *Cluster analysis for researchers*: Lulu. com.

Singh, K., Behera, R., & Mantri, J. (2018). *Big Data Ecosystem – Review on Architectural Evolution*. Paper presented at the IEMIS 2018, Kolkata, West Bengal, India.

Singhal, S., & Jena, M. (2013). A study on WEKA tool for data preprocessing, classification and clustering. *International Journal of Innovative technology and exploring engineering (IJItee), 2*(6), 250-253.

Song, Y.-Y., & Ying, L. (2015). Decision tree methods: applications for classification and prediction. *Shanghai archives of psychiatry, 27*(2), 130.

Srivastava, S. (2014). Weka: a tool for data preprocessing, classification, ensemble, clustering and association rule mining. *International Journal of Computer Applications, 88*(10), 26 - 29.

FEATURES OF ADAPTIVE APPLICATIONS - DEVELOPMENT AND IMPLEMENTATION

Meiramgul Mukhambetova[1], Zulfiya Zhanuzakova[2], Nurgul Baitemirova[3]

Abstract
This article discusses the features of adaptive software applications, innovative capabilities of the Bootstrap framework in creating applications. This paper describes the configuration of the framework, the features of the grid mode, the division of content into blocks, the implementation of the application, reading of the apps from the mobile version or from other devices. The use of this approach in the development of applications helps to avoid problems in programming and to carry out work with quality as a result of the use of optimal methods. The introduction of the Bootstrap framework in the educational process was carried out in the educational programs "Computer Science" and "Physics and Computer Science" at the University of Atyrau in Kazakhstan in the discipline "Web-programming".

Key words
bootstrap, framework, adaptive application, web-site, educational programs.

1. Introduction

The digital revolution has led to a complete overhaul of the entire system, the creation of computer applications in the formation of a modern employment ecosystem, characterized by the ability of people to quickly adapt to new conditions, flexibility (Rudakova & Perebatov, 2016). With the popularity and popularity of mobile devices and gadgets, developers have begun to look for solutions to problems in this area and to adapt to new situations. Adaptability is a feature that allows software applications to be easily accessed from any device and automatically adapted to be read on screens of different sizes. In this regard, the problem of performing tasks such as writing a mobile version of the application or a creating application is solved, and simple and convenient innovative methods help to increase the speed of software development (Moreto, 2016). The Bootstrap framework, developed by Twitter specialists, is a new approach to solving problems such as a set of programs for different extensions, simplification of work with the user and increase productivity, as well as browser compatibility and the absence of unique programming templates (Shibanov, Mezenkov & Makarychev, 2010). The problem of adaptability of the user interface and ways to solve of using an interface the automatically changes as a result of the full use of system capabilities with an interconnected set of software and hardware (Ageev, 2018).

The purpose of the study. Clarify the features of adaptive applications and ways to implement them. Provide as effective solutions for use in the learning process.

[1] Kh. Dosmukhamedov Atyrau University, Faculty of Physics, Mathematics and Information Technology, Department of Informatics, Kazakhstan, Atyrau,
mukhambetovamj@gmail.com.

[2] Kh. Dosmukhamedov Atyrau University, Faculty of Physics, Mathematics and Information Technology, Department of Informatics, Kazakhstan, Atyrau, zh.zulfiya@mail.ru.

[3] Kh. Dosmukhamedov Atyrau University, Faculty of Physics, Mathematics and Information Technology, Department of Software Engineering, Kazakhstan, Atyrau, n.baytemirova@asu.edu.kz

2. Advantages of adaptive applications

In programming technology mainly were used object-oriented programming languages and are still in use today. Libraries were used as the main tool for program developers to get rid of large amounts of one type of code and avoid errors. A library is a collection of various subroutines or objects used to create a software product. Libraries are still used as an integral part of any programming language. However, recently there has been a typical structure of a program that is applied to a specific context, called programming templates or patterns. After that, the world of programming began to get acquainted with the services of frameworks. Framework defines the structure of the program at all stages of program development and simplifies the work, adds standard modules, prevents errors in writing program text and integrates different components into one software product (Gamma, Helm, Johnson & Vlissides, 2015).

In adaptive web design we use a single codebase for different devices. If in full screen version we can display various content in same time, for small screen need to know for sure what part of content you should arrange in main part.

For creation of comfortable interface for all users need to consider various interests and aims of them. Adaptive web design give more effect and time will be invested in developing the necessary user interface for the target audience, eliminating the development of different versions of the same application and duplication of code.

The most important responsive designs are:

- responsive table grid that needs to be responsive to the client environment. There are two different approaches to responsive layout: fixed and floating layouts. Second approach differs in that the floating grid is always proportional to the size screen;
- floating images must maintain their aspect ratio. To maintain proportions, you can also use two approach: set the maximum height, or the maximum width 100 percent. This causes the image to stretch in any direction. Another way of processing is to crop a part adaptable image;
- using media queries, the simplest filters in the format CSS3 that allow browsers to track the resolution used device and load the appropriate CSS at runtime.
- So, responsive design is made possible by using three main components: adaptive table grid, floating images and multimedia, as well as media queries in CSS3 format.

The most important benefits of a responsive web application design are as follows:

- nice user interface on every type of device;
- reduced manufacturing costs relative to development and release of subsequent releases of different versions of the application for each device type;
- reduced advertising and marketing costs versus support application campaigns for each device;
- better search engine indexing and better optimization search engines (SEO, Search Engine Optimization);

- high conversion rates and the generation of leading positions in as SEO (visibility of a web application in search engines) increases.

3. Implementation of Bootstrap

Using the Bootstrap framework helps prevent many of the problems encountered in the traditional web programming industry. In this way, programmers help to find the optimal solution to the problems of organization and control of the creation of websites.

Our study aims to clarify and implement in the learning process ways to use the adaptive capabilities of the Bootstrap framework as a method of creating a web-client application.

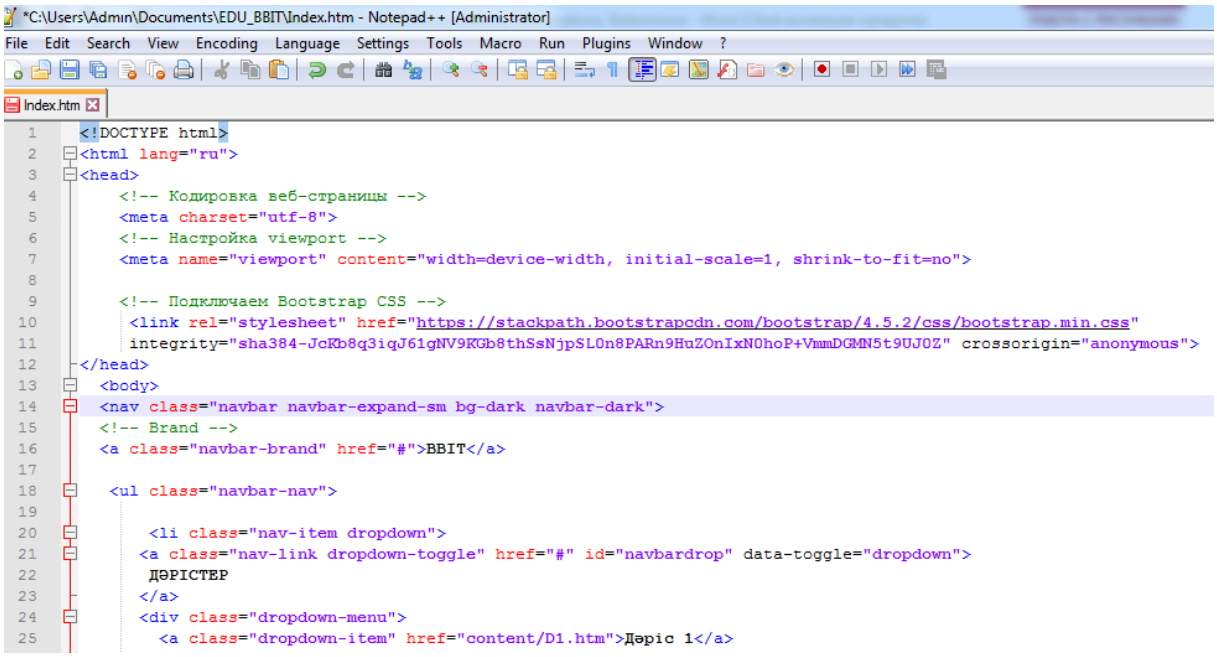
Bootstrap functions are available in three different modules:

- CSS. This module has a number of standard classes for usage and easily expands for settings.
- Component. this module has all the many uses built-in components, for example, an extensive set of pictograms.
- JavaScript: The framework includes a Bootstrap-style jQuery plugin.

Today, this framework can be implemented in different interpretations:

1) install the framework separately from the special site https://getbootstrap.com/;
2) by inserting in the HTML-code the link of necessary settings connected with CDN content;
3) Using of integrated environments, such as ASP.NET MVC 5, developed by Visual Studio 2019.

Figure 1: Example of using HTML code with connection of CDN content



Depending on the preferences of the developers, they can choose which of the above methods to use Bootstrap. Once the program's working environment is ready, the task is to master the components and commands of the framework. We can say that Bootstrap has its own grid system as a foundation. The program includes components that ensure optimal placement of web-sites in the structure of an adaptive network system.

The design of web-pages here can be created in any structure according to the required dimensions.

The main elements of the Bootstrap environment for creating a grid:

- containers - elements of the class container or .container-fluid;
- rows - an element of the row class;
- adaptive blocks are elements of one or more col classes.

A container is the first element in which a web page layout is created. Its primary purpose is to determine the width of the layout to be developed. Bootstrap 3 and 4 have two types. The first (container) is adaptive - fixed, and the second is used to create an adaptive - container (fluid) layout. Rows are also containers, but they are used for adaptive blocks of the Bootstrap grid. For Bootstrap 3, its main function is to create margins from 15px to the left and right. Boostrap 4 performs not only negative indentations, but also flex-container function. That is, if this element is not installed, their actions on the adaptive blocks will not be performed. The principle of using a row element is simple, it should always act as a parent for responsive boxes. That is, if you want to mark any element (container or block) with adaptive blocks, you must first set the row before creating them, and then place the blocks in these rows. An adaptive block is an element with an adaptive width. That is, its width in the viewport range can have one value and second in another. Assignment of an adaptive block action is performed by one or more col classes. Boostrap's adaptive blocks are arranged in a line. There are no more than 12 blocks in a row. Blocks that do not fit in one row are placed in the next row (Shah, 2015).

There are special settings that are assigned to read adaptive applications created by Boostrap from different devices. It allows you to set settings to suit any size of phone, tablet and desktop computer displays. The following table lists the extensions according to the devices (Spurlock, 2013).

Table 1 - Extensions according to class prefixes

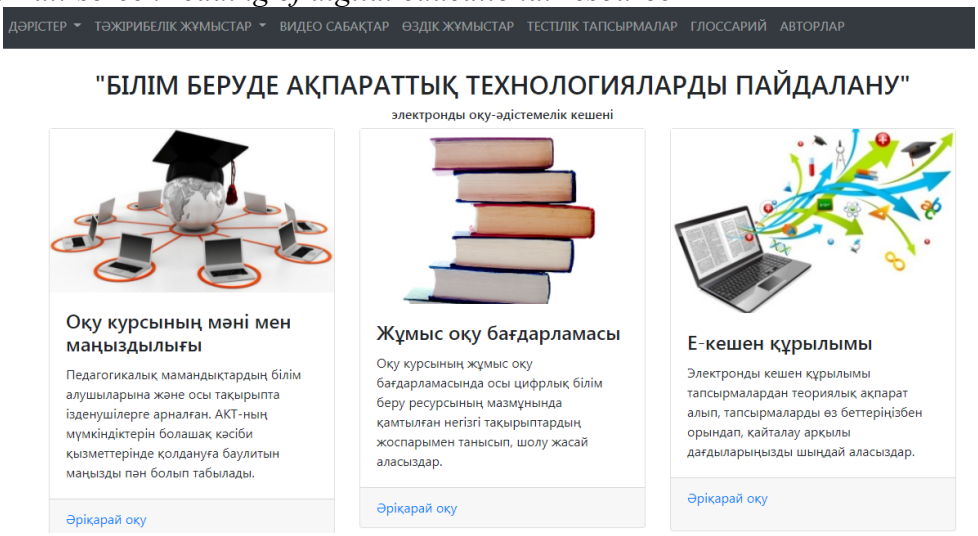
	Very small devices (phones <544 px)	**Small devices (tablets>= 544 px and <768 px)**	**Medium devices (desktops>= 768 px and <900 px)**	**Large devices (desktops> = 900 px and <1,200 px)**	**Very large devices (desktops> = 1,200 px)**
Features of the grids	Only horizontal tracks	Adaptive according to column grids			
Fixed container width	Auto	544 px	750 px	970 px	1170 px
Class prefix	.col-xs-*	.col-sm-*	.col-md-*	.col-lg-*	.col-xl-*
Number of columns	12				
Fixed column width	Auto	~ 44 px	~ 62 px	~ 81 px	~ 97 px

The steps described above are the basic steps to follow at the initial stage of building an adaptive application. Then the design content for programming the client part of the software application is developed, which is tailored to the tastes and needs of each user.

4. Results

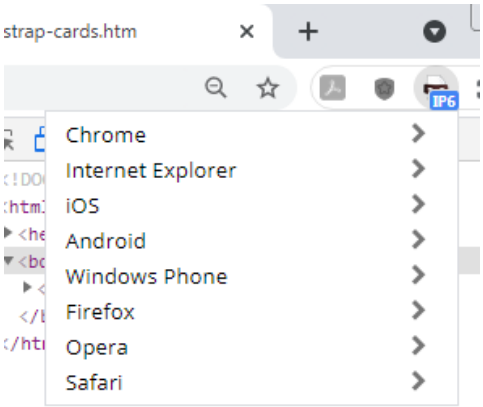
Adaptive software applications are in great demand for projects in academic, business processes, marketing and other areas. In our case, the adaptive software developed with Boostrap is intended for use in the learning process. The digital educational resource "Use of information technology in education" can be easily used by students through computers and mobile phones.

Figure 2: Full-screen reading of digital educational resource



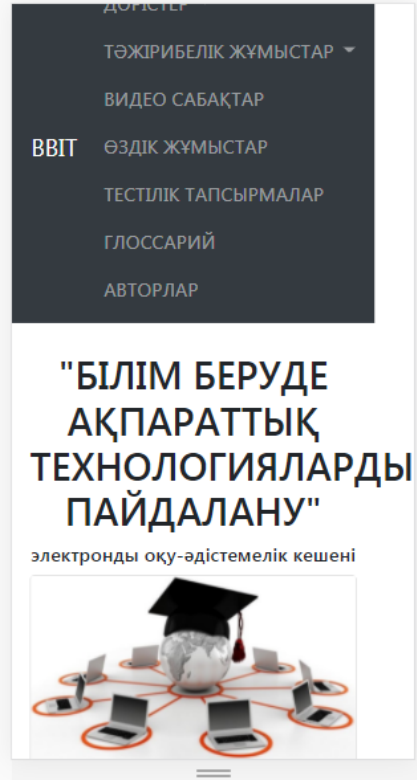
You can use the User-Agent Switcher for Chrome extension for Google Chrome to test the adaptability of the mobile version. Using the settings of this command, we can select the desired environment from the following list and check the application. In our case, the readability of the applications was checked in as iOS Android applications.

Figure 3: Selection a mobile application test environment using User-Agent Switcher for Chrome



For creation of comfortable interface for all users need to consider various interests and aims of them. Adaptive web design give more effect and time will be invested in developing the necessary user interface for the target audience, eliminating the development of different versions of the same application and duplication of code.

Figure 4: Reading a web application in a mobile version



This method offers an innovative design style, as well as the ability to innovate in programming the client part of software applications. In our case, the MS SQL server was first used in the course "Client-Server Technologies" as an integrated method for the ASP.NET MVC 5 pattern developed in the Visual Studio 2019 integrated environment. The MySQL server was used when the link was inserted into the web-site code. In both cases, the adaptability of the client part of the software has increased its importance and value.

Students used the Bootstrap framework to create dynamic web-site projects on individual topics.

5. Conclusion

Bootstap framework in the creation of adaptive software applications were implemented in the educational content of the educational programs of Atyrau University "Computer Science" and "Physics and Computer Science".

Bootstrap framework is a convenient way to create software applications, improve the design of web-sites, offers a modern structure, and its adaptability increases its importance in terms of hardware and software. Improvement of educational programs for future specialists in the field of information technology in accordance with the needs of employers, the introduction of innovative software methods and optimal solutions in the educational process will be possible as a result of constant study and research of ICT.

6. Resources

RUDAKOVA G. M, PEREBATOV K. S. (2016). Using the Bootstrap development environment to create modern adaptive sites. Information technologies in science, education and management. 279-285.

MORETO S. Bootstrap By Example. (2016). Packt Publishing. 45-48.

SHIBANOV S. V, MEZENKOV A. A, MAKARYCHEV P.P. (2010). Application of metadata in adaptive information systems of client-server architecture. Proceedings of the International Symposium "Reliability and Quality". T. 1. 120-125.

AGEEV E. L. Using frameworks in software development. (2018). Bulletin of modern research. No. 11.1. 210-212.

GAMMA E., HELM R., JOHNSON R., VLISSIDES J. (2015). Openings of object-oriented programming. Design patterns. 297-301.

SPURLOCK J. (2013). Bootstrap: responsive web development. " O'Reilly Media, Inc.". 105-108.

SHAH M. (2015). Responsive web development using the Twitter Bootstrap framework. 21-23

Bootstrap - Grid Elements and How to Use Them to Create a Layout. https://itchief.ru/bootstrap/grid-elements. 03/21/2021.

CHARACTERIZING DESK-BOUNDED JOB HAZARDS VIA BIOMETRIC SCREENINGS: A CASE STUDY

Lourdes Cecilia Ruiz Salvador[1], Richard Perez[2]

Abstract
Occupational Safety and Health is a cross-cutting topic among businesses. Organizations are continuously looking for ways to provide healthier and safer workplaces. Biometric screenings are an effective tool to recognize safety and health issues. The following work describes occupational problems encountered in office workers in Ecuador thanks to biometric screenings. It is a case study in which the primary data of 409 employees were collected, organized, and analyzed using descriptive statistics. As a result, health problems such as overweight, high cholesterol, eye problems, and musculoskeletal disorders were detected. These health issues are related to office-bounded professions. Timely intervention, management, and treatment of these biometric indicators aid in the identification of potential safety hazards that can develop in long-term diseases and accidents in the workplace.

Keywords
Biometric, Screenings, Occupational, Safety, Health

1. Introduction

Occupational Safety and Health (OSH) pertains to all the elements concerning health and safety in the workplace. It comprises the legal framework, policies, programs, regulations, and guidelines toward preserving the welfare of the employees and preventing hazards that can provoke work-related injuries or accidents(Goetsch, 2011)(Ruiz Salvador & Thinh, 2016). OSH is continuously evolving and adapting to new workplace environments. For example, office or home office work had become prevalent nowadays. Hence, this type of job leads to specific health conditions that can affect the workers' health and the organization's safety.

Small and large businesses are constantly seeking new approaches to contribute to safer and healthier workspaces (Breaux-shropshire et al., 2012). Hence, wellness programs, including biometric screenings, are becoming pervasive in organizations (Rameswarapu et al., 2014). According to the 2019's survey by Kaiser Family Foundation(KFF, 2019), among the companies that offer health programs, 26% of the small companies and 52% of the large companies supply biometric screenings to their employees. Moreover, 58% of the large companies offer incentives upon screening completion and 14% compensate or penalize the worker based on their biometric outcomes.

A biometric screening is defined as the measurement of physical characteristics such as height, weight, body mass index (BMI), blood sugar, blood pressure, blood cholesterol, waist circumference, and more (Loeppke, 2013). Its main objectives are to offer a snapshot of the employees' health, identify health risks, initiate preventive care, reduce healthcare costs and enhance performance and productivity(Hecht, 2020)(Maeng et al., 2017). Furthermore,

[1] Óbuda University, Budapest, Hungary, lourdes.ruiz@bgk.uni-obuda.hu

[2] Universidad Técnica de Cotopaxi, Ecuador, richard.perez@utc.edu.ec

biometric screenings led to increased drug prescriptions to treat blood pressure and high cholesterol, according to the Employee Benefit Research Institute (Fronstin & Roebuck, 2015).

A study conducted in New Mexico(Brown-Connolly et al., 2014) found that biometric screenings are cost-effective because they aid in preventing chronic diseases.

The present research describes a case study executed in a higher education institution in Ecuador. It consists of the data collection of biometric characteristics via biometric screenings, the analysis and evaluation of these physical traits, and identifying health and safety hazards among the workforce. Descriptive statistics were used to display the collected data and provide a summary of the main findings.

2. Methodology

Primary data of 409 employees at a higher education institution were acquired as part of the annual biometric screening event held in the institution. The screening consisted of four parts:

1. Primary biometric characteristic collection (gender, age, height, weight)
2. Blood and urine sample collection
3. Physical examination
4. Drug consumption screening

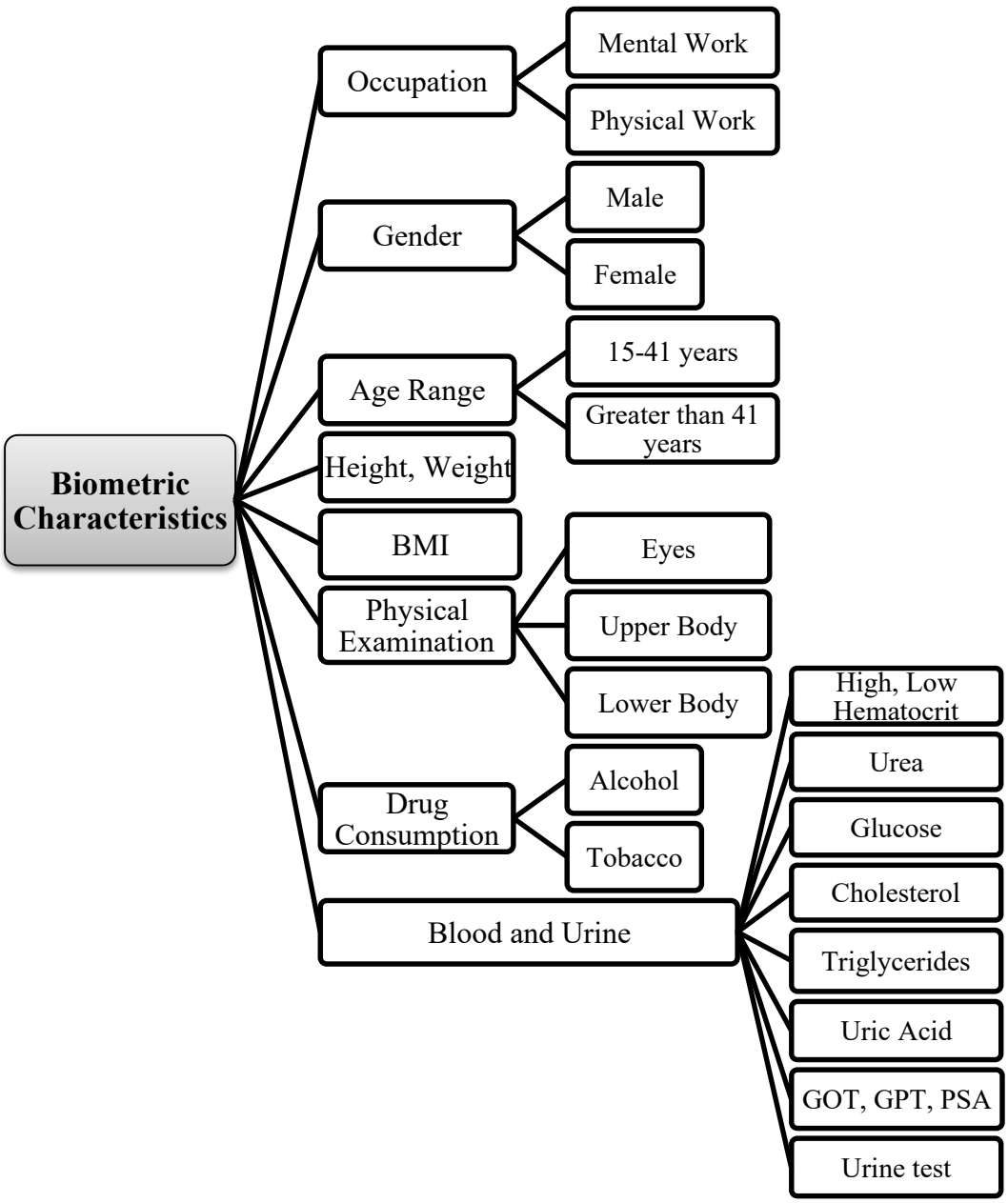
The laboratory test and the other screenings' results were tabulated and saved in an excel database to analyze and later report the workforce's whole health status.

Figure 1 enumerates the biometric characteristics collected and their subcategories. In the physical examination, the problems detected in the upper body included: nose, abdomen, neck, shoulders, arms, elbows, forearm, wrists, and hands. Likewise, the problems detected were in hips, thighs, knees, legs, feet, and back for the lower body. For blood and urine samples, the laboratory tests performed included Glutaminic-Oxaloacetic Transaminase (GOT), Glutamate Pyruvate Transaminase (GPT), and Prostatic Specific Antigen (PSA).

The occupational doctor established a diagnosis stating if the worker is fit, fit with limitations, or unfit to work. This diagnosis is a clinical evaluation that contrasts the workers' health with the working conditions, environment, and specific work duties to assure that the workers do not constitute a hazard for themselves, their colleagues, and the organization(CCOHS, 2016; Cowell, 1986).

The outcome of the whole biometric screening process was presented and discussed with each of the employees. In addition, health advice regarding eating habits, physical activity, intervention, and prevention procedures was offered and explained in case of abnormal results. Biometric screening outcomes were compiled and categorized using descriptive statistics. Microsoft Excel and SPSS were used to calculate means, standard deviation, frequencies, and percentages.

Figure 1: Biometric characteristics acquired in the case study



3. Results

Figure 2 shows in percentages the biometric screening results in terms of gender, age, occupation, alcohol, tobacco consumption, and occupational diagnosis. Males constitute a larger percentage of employees compared with females. Concerning the age group, participants aged 15 to 41 years represent a higher percentage than their colleagues over 41 years old. A relatively small number of participants (12%) perform physical work than mental work (88%). Alcohol and tobacco consumption presented minimal values of 8% and 3%. The occupational analysis revealed that nearly 84% of the employees are apt to work in their current job position,

more than 16% are apt to work but with specific restrictions. Consequently, there are no cases of employees that are classified as unfit to work.

Figure 2: Results biometric screening in percentages

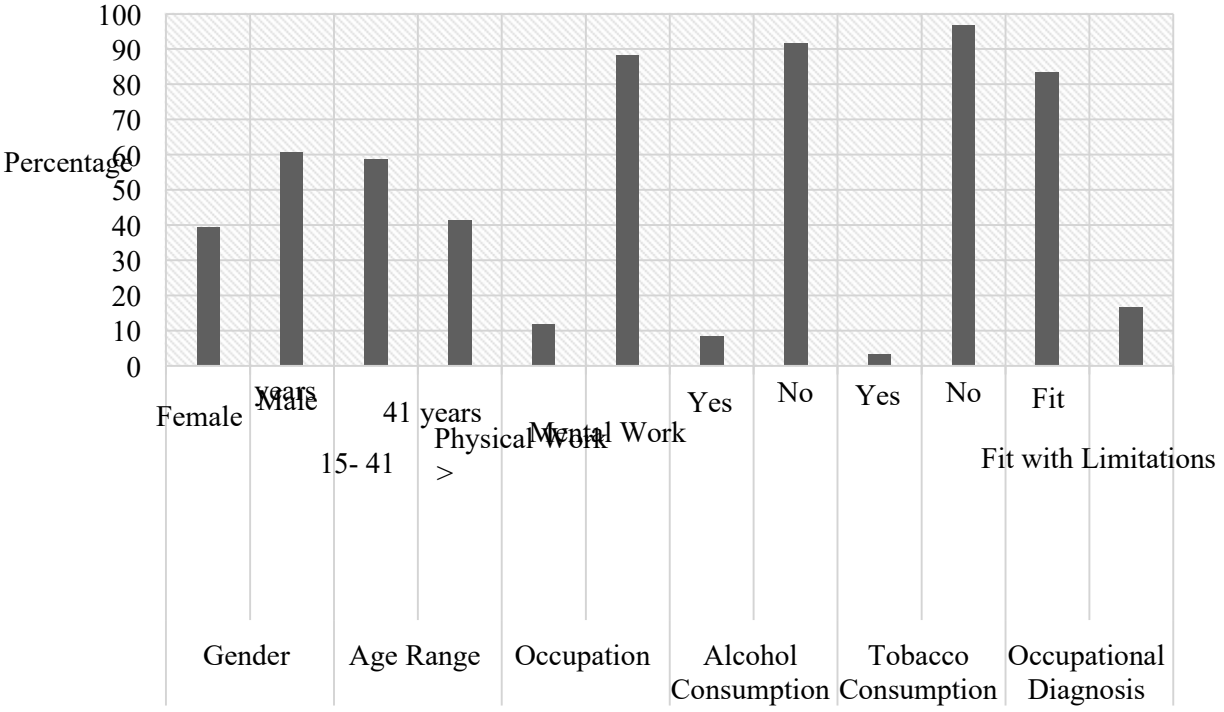


Table 2 indicates the number of problems in the laboratory exams and the physical examination. The laboratory exam results show that nearly 25 % of the workers are within the normal parameters regarding the blood and urine samples, while almost 50% of the participants present 1 or 2 abnormal results. In contrast to 12.7%, that has 4 to 8 problems of the 11 analyzed results. The physical exams indicated that 50% of the employees do not have physical issues, and just 3.4% have 4 to 8 problems over the 15 problems presented in the whole university community.

Table 2 enumerates the mean and standard deviation of height, weight, and BMI for females, males, and the total workforce. According to the World Health Organization (WHO), the mean height values for males and females are within the normal ranges for Ecuadorians, which are 1.67 and 1.54 meters, respectively(Metro Ecuador, 2016). However, there are extreme values to pay attention to obesity and diseases; this is reflected in the high standard deviation values, representing variation in the mean weight group.

Therefore, the mean BMI values for females (25.83) and males (27.11) fall into the range of pre-obesity, according to WHO (25-29.9), which shows that the workforce is slightly overweight. The standard deviation values are high, which is essential to pay attention to employees under extreme BMI classifications, such as obesity and morbid obesity(World Health Organization, n.d.). Besides, the overall BMI classification shows a considerable percentage (60%) of overweight or obese employees, and nearly 40% have a normal BMI, as shown in Figure 3.

Table 1:Results biometric screening height, weight, BMI (means and standard deviations)

	Height		Weight		BMI	
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
Female	1.58	0.07	64.86	11.85	25.83	4.15
Male	1.67	0.71	75.73	11.85	27.12	3.75
Total	1.63	0.08	71.45	12.98	26.61	3.95

Figure 3: Results BMI classification in percentage

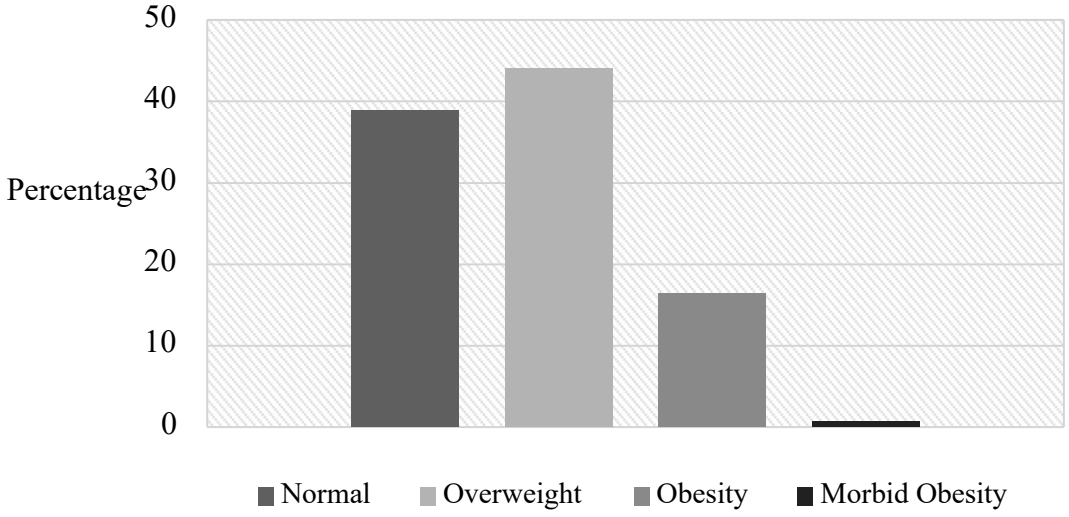


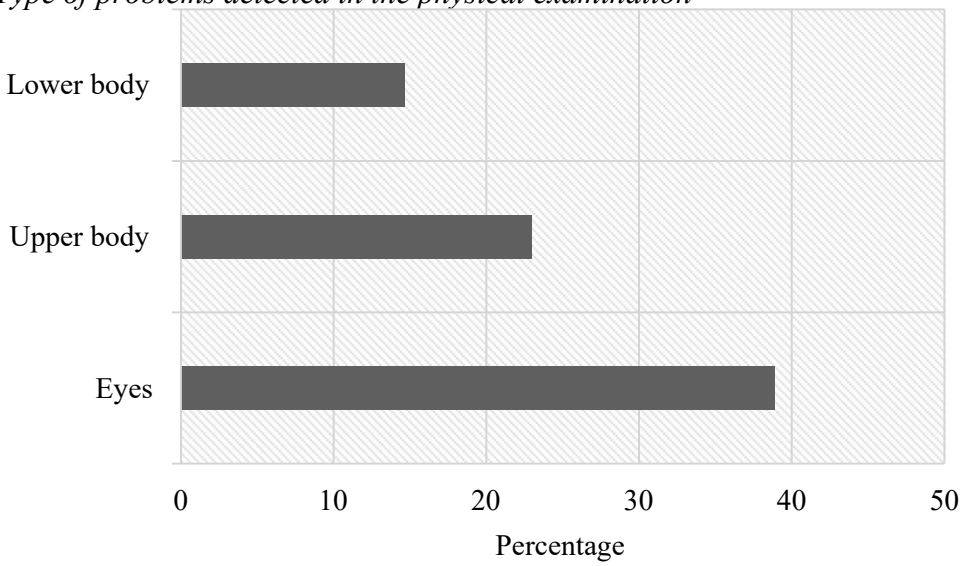
Table 2 indicates the number of problems in the laboratory exams and the physical examination. The laboratory exam results show that nearly 25 % of the workers are within the normal parameters regarding the blood and urine samples, while almost 50% of the participants present 1 or 2 abnormal results. In contrast to 12.7%, that has 4 to 8 problems of the 11 analyzed results. The physical exams indicated that 50% of the employees do not have physical issues, and just 3.4% have 4 to 8 problems over the 15 problems presented in the whole university community.

Table 2: Results blood and urine tests and physical examination in percentages

	No Problems	1 Problem	2 Problems	3 Problems	4 - 8 Problems
Abnormal Results Blood and Urine Laboratory Exams %	24.2	24	24.2	14.9	12.7
Problems detected during the Physical Exam %	50.4	34.2	8.1	3.9	3.4

Figure 4 presents the type of problems identified during the physical exam. The main physical problem encountered in the workers was in the eyes, followed by problems in the upper body (23%), such as in the neck, back, wrists and hands. Problems in the lower body such as knees, hips, legs, and feet were around 15%. The division of the body in the upper and lower eases the visualization of health affections related to musculoskeletal disorders.

Figure 4: Type of problems detected in the physical examination



In addition, the occupational doctor diagnosed clinical problems and pathologies through the analysis of the laboratory and physical exam results. Clinical problems such as polycythemia, sedentarism, hypercholesterolemia, and hypertriglyceridemia are prevalent among university workers, as shown in Figure 5. Also, for pathologies, metabolic, hematologic are predominant. Ophthalmologic and musculoskeletal pathologies comprise significant percentages (38.6 %, 19%), as presented in Figure 6

Figure 5:Clinical problems detected in percentages

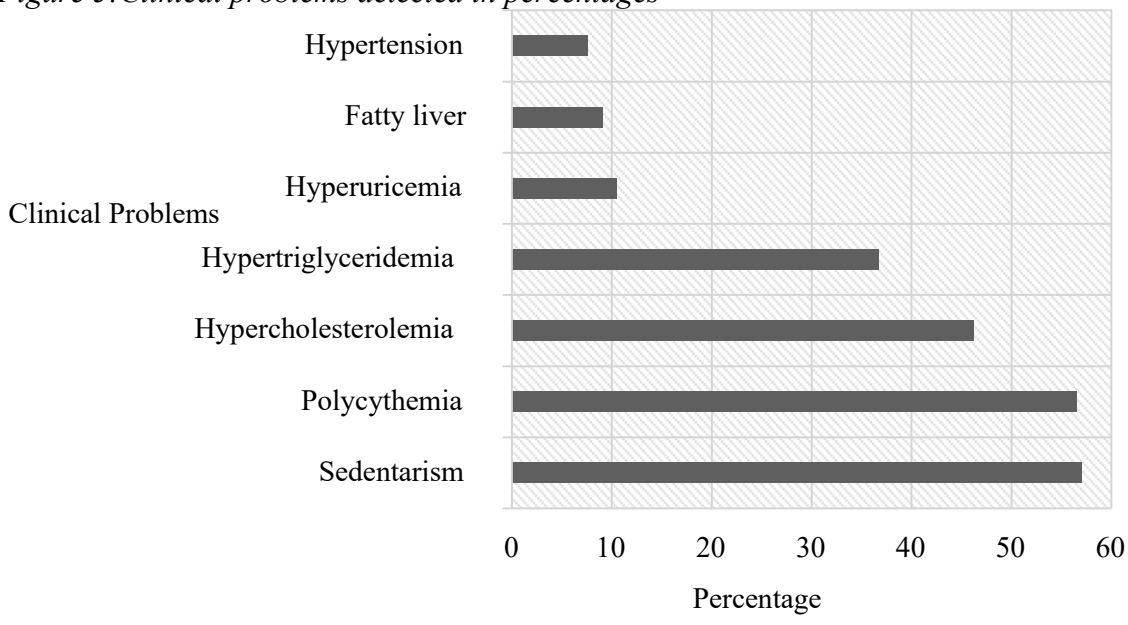
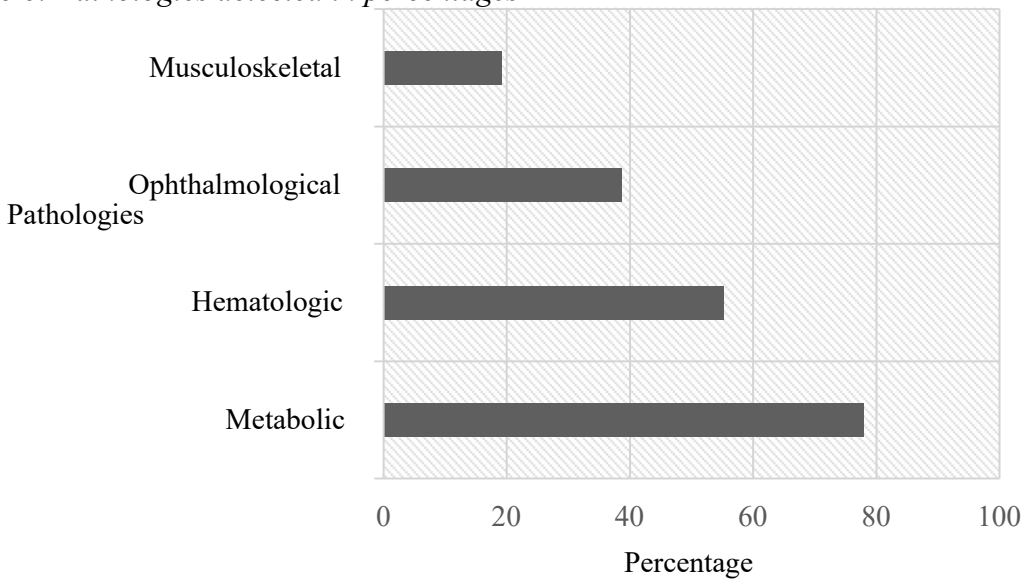


Figure 6: Pathologies detected in percentages



4. Discussion

The biometric data collection and analysis described in the case study was valuable for understanding the impact of biometric screenings on occupational safety and health.

Regarding blood and urine collection and the subsequent laboratory results, the dominant problems detected in the laboratory exams were:

- High hematocrit
- High cholesterol
- High triglycerides

The hematocrit test is a blood characterization analysis that evaluates the proportion of red blood cells. It can detect diseases such as anemia and certain cancers. Factors such as genetics, smoking, and respiratory problems can affect hematocrit levels. Another factor contributing to this abnormality is living in high altitudes(Shiel, 2021), where this research was executed. This aspect can explain the prevalence of high hematocrit counts in laboratory exams. Abnormal results in cholesterol and triglycerides coincided with the BMI values stating that the university employees are pre-obese.

The alcohol and tobacco screening results exhibit a low consumption percentage of 8.3% and 3.2%, respectively. These results do not coincide with the national percentage of alcohol consumption: 41%(Extra EC, 2019) and tobacco: 8.8%(El Comercio, 2017). This issue can be explained by the fact that self-assessment questionnaires were administered. Blood tests such as cotinine analysis will show more realistic results and determine if there is an alcohol or tobacco problem in the workforce.

Musculoskeletal, ophthalmological pathologies and sedentarism need to be paid special attention. These affections were considerably high among the employees and can be caused by mental work. During the physical exam, a decrease in visual acuity was detected as its most recurrent issue. Nearly 40% of the workers presented this problem, which indicates that the employee needs to wear glasses or change the eye prescription. Mental workers are prone to eye problems due to extended hours in front of a screen, which is typical for the job position

they perform. Additionally, upper body problems such as back, neck, wrists, and hands are predominant in employees performing mental work.

Almost all types of work demand hands and arms usage and analyze their safety and health implications. Therefore, 19% of the workers were diagnosed with musculoskeletal disorders. These disorders are generated by the overuse of muscles, tendons, and nerves. Back, neck, shoulder, and knee pain are some problems related to them. Activities that are repetitive or involve uncomfortable positions cause these disorders that are painful and affect work capability regardless of age(Business in the Community, 2017).

Furthermore, sedentarism was present in 57% of the workforce. Since the deployment of desktops in the workplace, sedentary jobs have increased along with chronic diseases and mortality. Mental work occupies a considerable proportion of the study type of occupation, with over 88%. Office jobs have been linked to a sedentary lifestyle(Biswas et al., 2015)(Gallagher, 2015). As a solution, wellness programs, including physical activity during office hours such as elliptical workstations, sit-stand, or treadmill desks, are introduced in workspace design nowadays(Carr et al., 2016).

5. Conclusions

The case study presented in this work aimed to use the data collected in a biometric screening event and characterize the leading health problems in mental work. Biometric screenings at one event can collect several health data that can establish common affections and can also demonstrate how a type of work affects employees' health. Thus, it provides a solid argument for making specific changes in the workspaces that can contribute to safety and health outcomes.

In the case study examined, biometric screening results analysis and data visualization indicate workforce primary health conditions. Descriptive statistics pointed out preponderant issues suffered by the university employees. More than half of the workers are overweight or obese. This fact explains the major clinical problems encountered, such as sedentarism, hypercholesterolemia, and hypertriglyceridemia. Eyes and upper body affections are the most frequent problems found in the physical examination, which can trigger two of the dominant pathologies: opthalmological and musculoskeletal disorders.

The case study findings coincide with a report entitled: "The work colleague of the future" (Higham, 2019). This report summarizes the hazards linked to an office workstation, which will be the most common workplace environment. In this report, generally, long-term health problems in the eyes, nose, hands, wrists, legs, veins, back, and also overweight are forecasted for the office employee in 2040. Thanks to the results found in a biometric screening event, the company can identify the specific hazards among their employees and invest in policies to eliminate or control them.

It is essential to annotate that biometric screenings provide a broad picture of the health status of the workforce but do not replace a consultation with a doctor. Nevertheless, the screening outcomes create health awareness among the employees and motivate them to change bad habits and live a healthier and safer life. Furthermore, the biometric screening results need to be considered to execute safer work processes and provide healthy workspaces.

6. Resources

Biswas, A., Oh, P. I., Faulkner, G. E., Bajaj, R. R., Silver, M. A., Mitchell, M. S., & Alter, D. A. (2015). Sedentary time and its association with risk for disease incidence, mortality, and hospitalization in adults a systematic review and meta-analysis. In *Annals of Internal Medicine* (Vol. 162, Issue 2, pp. 123–132). American College of Physicians. https://doi.org/10.7326/M14-1651

Breaux-shropshire, T. L., Whitt, L., Griffin, R. L., Shropshire, A. T., & Calhoun, D. A. (2012). Characterizing Workers Participating in a Worksite Wellness Health Screening Program Using Blood Pressure Control, Self-Monitoring, Medication Adherence, Depression, and Exercise. *American Association of Occupational Health Nurses*, 292–300.

Brown-Connolly, N. E., Concha, J. B., & English, J. (2014). Mobile health is worth it! Economic benefit and impact on health of a population-based mobile screening program in new Mexico. *Telemedicine and E-Health*, *20*(1), 18–23. https://doi.org/10.1089/tmj.2013.0080

Business in the Community. (2017). *Musculoskeletal health in the workplace: a toolkit for employers*. https://wellbeing.bitc.org.uk/sites/default/files/business_in_the_community_musculoskeletal_toolkit.pdf

Carr, L. J., Leonhard, C., Tucker, S., Fethke, N., Benzo, R., & Gerr, F. (2016). Total Worker Health Intervention Increases Activity of Sedentary Workers. *American Journal of Preventive Medicine*, *50*(1), 9–17. https://doi.org/10.1016/j.amepre.2015.06.022

CCOHS. (2016). *Fit to Work : OSH Answers*. Canadian Centre for Occupational Health and Safety. https://www.ccohs.ca/oshanswers/psychosocial/fit_to_work.html

Cowell, J. (1986). Occupational Medicine Guidelines for fitness-to-work examinations. In *Occupational Health in the Chemical Industry*.

El Comercio. (2017). El consumo diario de cigarrillo en el Ecuador bajó en 16 años. *EL Comercio Ecuador*. https://www.elcomercio.com/tendencias/consumo-diario-cigarrilloecuador-fumadores.html Extra EC. (2019). ¿Qué tanto 'chupamos' los ecuatorianos? *Extra Ecuador*. https://www.extra.ec/buena-vida/chupamos-ecuatorianos-alcoholismo-estadisticasconsumo-IY2742846

Fronstin, P., & Roebuck, M. C. (2015). Financial Incentives, Workplace Wellness Program Participation, and Utilization of Health Care Services and Spending. *EBRI Issue Brief*, *417*, 1–23. http://www.ncbi.nlm.nih.gov/pubmed/26477217

Gallagher, J. (2015). *Office workers "too sedentary."* BBC News. https://www.bbc.com/news/health-32069698

Goetsch, D. L. (2011). *Occupational Safety and Health for Technologist, Engineers, and Managers* (Pearson (Ed.); Seventh).

Hecht, M. (2020). Biometric Screening: What Is It and What's Tested? *Healthline*. https://www.healthline.com/health/what-to-know-about-a-biometric-screening#fast-facts

Higham, W. (2019). *The Work Colleague of the Future A report on the long-term health of office workers*. *June*.

KFF. (2019).*2019 Employer Health Benefits Survey*. https://www.kff.org/healthcosts/report/2019-employer-health-benefits-survey/

Loeppke, R. (2013). Biometric Health Screening for Employers: Consensus Statement of the Health Enhancement Research Organization, American College of Occupational and

Environmental Medicine, and Care Continuum Alliance. *Journal of Occupational and Environmental Medicine*, *55*(10), 1244–1251. https://doi.org/10.1097/JOM.0b013e3182a7e975

Maeng, D. D., Geng, Z., Marshall, W. M., Hess, A. L., & Tomcavage, J. F. (2017). An Analysis of a Biometric Screening and Premium Incentive-Based Employee Wellness Program:

Enrollment Patterns, Cost, and Outcome. *Population Health Management*, *21*(4), 303– 308. https://doi.org/10.1089/pop.2017.0110

Moreno, C., Bolaños, C., Monzón, C., Peña, R., & Acosta, E. Metro Ecuador. (2016). Retrieved April 29, 2021, from https://www.metroecuador.com.ec/

Rameswarapu, R., Valsangkar, S., Rizvi, A., & Kamineni, U. (2014). Trends shaping corporate health in the workplace. *Apollo Medicine*, *11*(3), 217–221. https://doi.org/10.1016/j.apme.2014.07.010

Ruiz Salvador, L. C., & Thinh, D. Van. (2016). Occupational Safety and Health: An overview. *2016 IEEE 11th International Symposium on Applied Computational Intelligence and Informatics (SACI)*, 355–360. https://doi.org/10.1109/SACI.2016.7507401

Shiel, W. C. (2021). *Hematocrit Ranges and Chart:Test, High, Low and Normal*. MedineNet. https://www.medicinenet.com/hematocrit/article.htm#how_is_the_hematocrit_measured

World Health Organization. (n.d.). *WHO/Europe | Nutrition - Body mass index - BMI*. Retrieved October 21, 2019, from http://www.euro.who.int/en/health-topics/diseaseprevention/nutrition/a-healthy-lifestyle/body-mass-index-bmi

IT BUSINESS TECHNOLOGIES IN EDUCATION

Andrey Rybalchenko[1], Gulmira Abildinova[2]

Abstract
Commerce is one of the most important engines for the development of modern technologies, including the IT-sphere. Nowadays, a lot of software has been created for businesses that help with organizing and analysing dat programs that simplify work, and help make a profit. Notable examples include data centralization, workflow simplification, and predictive analytics. In terms of education, this can help identify weaknesses in the learning process and help build a development trajectory.

In this article, we conduct a comparative analysis of existing developments that will help raise the automation of education to a new level.

Keywords
Education, commerce, innovative technologies, software.

1. Introduction

Automation processes at the present stage of society's development have penetrated all spheres of human activity. The introduction of information technologies and automation of management in educational institutions is one of the priorities of the modern education system especially considering that the current COVID-19 pandemic has shown us that it is necessary to respond immediately[1]. The main difficulty in implementing automated systems in education is the large gap between the development of computer technology, software and methodological support for control systems and their use. The transfer of the process of managing the activities of an educational institution to paperless technology allows you to save the management and teaching staff from routine work that takes a lot of time in office work, planning and maintaining documentation of the educational process. Management activity requires managers to constantly analyse the state of affairs, solve the tasks set. Information and communication technologies allow for optimal interaction of the managed and managing subsystems of the school. Automation of the management of the educational institution's activities should eventually become the norm in the daily work of managers and the entire teaching staff. The effectiveness of the automation of educational and managerial activities of the institution is determined by the training of managerial and pedagogical personnel in the use of information and communication technologies[2]. Understanding of the problems of the management process by the head of the educational institution, the availability of user skills and the use of an automated management system is a prerequisite for the successful implementation of ICT in the life of an educational institution. An educational institution, being a complex social management system, consists of separate units that are closely connected and interact with each other and implement: educational, methodological, educational, managerial, economic, personnel, and other types of activities. The flows of activities that form single processes can be depicted using traditional diagram techniques such as flowcharts, BPMN diagrams, Activity diagrams, etc. or newer techniques such as the Jurík-Schmidt's diagram technique for displaying complex business processes, which can be used for educational

[1] ENU im. Gumil'ova, Nur Sultan, Kazachstan, rybalchenkoas@gmail.com

[2] ENU im. Gumil'ova, Nur Sultan, Kazachstan gulmira2610@mail.ru

processes also (Jurík & Schmidt, 2020). The information flows that circulate in the automated system combine all the management links into a single information field. Making these flows purposeful ensures the effective functioning of the system based on information and telecommunications technologies.[3]

In contrast to the automation of educational institutions, you can take the automation of commercial activities. Business automation has long been a step forward, as quite impressive budgets are spent annually in this direction. Managers of large companies understand the value and profitability of this process.

When automating business processes, the level of work improves, employees ' workload decreases, and the data collected in this way includes fewer inaccuracies. The education system can and should learn from the experience of commercial organizations.

2. Research and analysis

Most universities and many other OUS have already implemented automation subsystems for the main management functions, but these subsystems are usually highly specialized. Their integration is hindered by the isolation on the functions of individual departments of the OU or on a small group of business processes.

Today, two alternative approaches to OU automation dominate. The first involves a strict unification of business processes and the creation of a similar product line (software modules). The second one appeared relatively recently and is aimed at creating the core of the system and a platform that allows you to integrate, adapt, geographically distribute and scale heterogeneous applications in order to optimize business processes and, if necessary, their gradual ("soft") unification.[4]

We analysed the already implemented automated system "Electronic Journal Kundelik", related to student performance.

The program "Electronic journal Kundelik" is used to reflect the results of the educational process in schools of the Republic of Kazakhstan[5].

"Electronic journal Kundelik" includes: a database for all students, an electronic journal of attendance and academic performance, a lesson schedule, calendar and thematic planning, tools for calculating statistics and preparing documents. Students are entered in the database and assigned to a specific class. The software product allows you to make final reports on academic performance and attendance.

The administration of an educational institution can monitor the educational process of classes, easily and in a short time receive a report on attendance and academic performance for a quarter.

Program features:

- input from the log of grades and student attendance, the formation of statements;
- automatic preparation of reports on academic performance and attendance;
- preparation of a report card;
- formation of a summary statement;
- The software product is designed for use in educational institutions. The system has two access modules;
- administrative (intended for the director, head teacher and teachers);
- viewing (for parents and students).

The administration of the educational institution could easily navigate through the class sites and exercise control over the information.

The grading system and the number of school days per week may be different for each class.

With this product, parents can monitor the progress and attendance of students, track the past and missed material, and solve important issues without waiting for meetings.

Features of the electronic journal.

Especially for teachers, the system "Electronic journal Kundelik" includes the following tasks::

- entry of grades in the electronic journal easily and in a short time (one click on the mouse to enter a single grade);
- simple entry of information about absentees, latecomers, sick people;
- generating reports on academic performance and attendance;
- preparation of the necessary surveys with the results in the form of tabular information;
- the ability to create a lesson plan that is not tied to dates, which speeds up the completion of the "homework" page for schoolchildren, as well as possible for use in subsequent years;
- placement of educational and methodological materials for preparing for lessons and performing homework by students (filled in using templates).

In the interests of managers, head teachers, the system "Electronic journal Kundelik" solves the following tasks:

- formation of administrative reports on the quality control of teachers ' filling in electronic journals;
- monitoring of academic performance, attendance of students and the formation of appropriate reports;
- formation of open reports of the educational institution with their subsequent publication on the school's website.

The analysis of the above system shows that there are software products on the market for automating student performance, but they have the following disadvantages:

Lack of inheritance.

The automation systems of educational institutions are scattered. In secondary schools, they can change every 3-5 years and the only thing that is inherited is the logic of building a business process, the software itself can be created every time from scratch, without using any platforms. At the same time, users will encounter the same errors as in the previous implementation.

Large software fragmentation.

In higher education institutions, software products may differ from university to university. This can lead to problems with data collection at the state level[6]. Dependency on remote servers or a single server.

There are cases in practice when failures on a single server led to a shutdown in all educational institutions.

Lack of personal improvements for schools with a single core.

These systems are not adapted to the activities of a particular school, thus causing inconvenience in use.

Integration problems.

Since systems are often written from scratch, there is a problem of integration with other subsystems. For example, educational institution portals, access control systems, or mobile applications.

For example, we have considered one of the most popular products from the side of commerce automation – 1C:Company.

Today, the level and prospects of informatization of society require automation of document management in an educational institution. This is especially relevant when collecting reports and calculating the main indicators of academic performance.

The most important purpose of an automated data processing system in our case is to increase the efficiency of performing the main functions of the class teacher, since the functioning of the grade statement block is associated with a very large information flow. In addition, the program should improve the efficiency of decision-making, increase labor productivity, reduce the number of computational errors by automating the data processing process, and promote efficient and secure storage and access to data.

The purpose of the software application is to create a unified information system that will effectively store and process information on the accounting of student performance and its detailed analysis.

School teachers spend a lot of working time performing numerous time-consuming accounting operations of data processing related to the performance of students. To perform these simple information processing procedures, you do not need special knowledge. The greater the amount of information, the greater the proportion of such works. This leads to a reduction in time for the execution of other important tasks.

Automating student performance requires careful consideration of the composition of variables and constant data. Information that characterizes, for example, students ' grades (permanent data) should ensure both the automation of grade processing and the operational accounting of the grades issued.

It was developed in 1C: Enterprise.

1C:Enterprise is a software product of the company 1C, and serves to automate the activities of the organization, here are already developed configurations for automation in education on this platform[7].

The simplicity of the language is one of the key characteristics of 1C:Enterprise: before you start programming, the developer does not have to study the programming language itself for a long time. Fundamental concepts of 1C:The rules are simple and clear, and developers can be productive from the very beginning.

The object - oriented paradigm is responsible for the tasks of distributed client - server systems: the use of the concepts of encapsulation, inheritance, and polymorphism. 1C: Enterprise provides a clear and efficient object-oriented development platform.

1C Platform - the company is designed to be able to create highly reliable application software. Special emphasis is placed on testing developed programs at the compilation stage, followed by the second level - dynamic verification (at the execution stage).

In 1C:Enterprise uses a huge number of different data access technologies. However, when designing database applications, the sequence of operations remains almost unchanged. In the course of writing programs, virtually the same components are used, which have been modified for use with a particular information access technology.

We also reviewed the IBM Planning Analytics software product. It is a cloud-based solution for planning and forecasting using high-performance analytics methods. The solution combines the advanced capabilities of several systems at once-IBM Cognos TM1 and IBM Watson Analytics. IBM Planning Analytics allows you to quickly build convenient systems of metrics and balanced metrics for business evaluation, which reduces labour costs for operational data processing and speeds up the receipt of business analytics results.

One of the most important things about IBM PA is Planning with AI

Automated model creation, natural language text processing, and cognitive assistance make it faster and easier to create accurate plans and forecasts using accuracy calculation schemes: the

Akaike Information Criterion, the Absolute error of the average value, the Average Root-mean-square Error, and so on. [8]

This functionality can be used to identify individual or global problems in education even before they appear. Help to build personal development trajectories or contribute to the development of strategies for the development of education in general.

3. Conclusions

Solutions for commerce eliminate many problems that exist in software products in the field of automation of the educational process. There are systems on the market that have been developing for more than ten years, have built-in integration schemes, fault tolerance policy, version inheritance and other advantages that the business has been improving in its work for years, as it has more serious opportunities for this. It is also necessary to use the experience of commerce not only to fix problems, but also to use the best practices. Now business automation has entered the era of Big Data and Deep Analytics, which allows them to build quite serious analytical reports for decision-making, predict possible scenarios. They use predictive models and Machine learning. All this experience can be used in Education. It is possible, based on the data of people already established as Professionals and Specialists in their fields, to build Individual development trajectories for students with personal recommendations. Or generate visual reports at the city\region\country level for making managerial decisions at the state level.

4. Resources

DRACHEVA V.I. (2011). *Avtomatizacija biznes-processov*. Retrieved April 20, 2021, from https://cyberleninka.ru/article/n/avtomatizatsiya-biznes-protsessov

GOLOSOV A.O., POLOTNJUK I.S., FILIPPOVICH A.Y. (2006). *Avtomatizacija obrazovatel'nyh uchrezhdenij na baze integracionnoj platformy*. Telekommunikacii i informatizacija obrazovanij 4, 78-83.

IBM Company (2021). *Modelirovanie v Planning Analytics Workspace*. Retrieved April 30, 2021, from https://www.ibm.com/docs/ru/planning-analytics/2.0.0?topic=workspacemodel-in-planning-analytics

JURÍK, P. & SCHMIDT, P. (2020). *Composition Diagram of a Complex Process: A Contribution to Business Process Modelling.* TIEES 2020: Trends and Innovations in E-Business, Education and Security: Eighth International Scientific Web-Conference of Scientists and PhD. Students or Candidates., 8, 81-92.

KUNDELIK (2021). *Description*. Retrieved April 21, 2021, from https://kundelik.kz/about/

OVCHINKIN O.V. & PYKHTIN A.I. & SHIROKOVA L.V.(2019) *Monitoring system of graduate employment as a factor of improving the efficiency of university educational activities.* International Journal of Emerging Trends in Engineering Research, 20, 849-853, doi:10.30534/ijeter/2019/207122019

ABELDINA Z. & MOLDUMAROVA Z. & ABELDINA R. & MAKYSH G. & MOLDUMAROVA Z.I. (2016) *Experience in education environment virtualization within*

the automated information system "Platonus" (Kazakhstan). International Journal of Environmental and Science Education, 11(18), 12512-12527.

KLYMENKO, E. Y., & ALPEISSOVA, S. E. (2021). *The experience of ukraine and kazakhstan of digitalization education under quarantine conditions.* Advances in Intelligent Systems and Computing, 161-172, doi:10.1007/978-3-030-71782-7_15

CYBERSECURITY ISSUES IN THE PENSION PAYMENT SYSTEMS

Zsolt Mihály Szabó[1]

Abstract
One of the most important values of economic and social life is information. Information is a resource for organisations, the basis of their efficient operation, an asset of the organisation and also often a product. Cybersecurity is the practice of defending computers, servers, mobile devices, electronic systems, networks, and data from malicious attacks. It is also known as information technology security or electronic information security. The term applies in a variety of contexts, from business to mobile computing, and can be divided into a few common categories. There is growing interest in the use of electronic payment (e-payment) systems in cash transfer programmes. When cash is transferred to beneficiaries through e-payment technologies such as mobile phone accounts or smartcards, there is potential to cut costs and reduce corruption compared with physical payment methods. The e-payment systems can also improve accessibility and security for programme recipients, which is important for reaching vulnerable groups including older people, people with disabilities and people in remote areas. We rarely think of information security as a problem, when in fact our daily lives are networked by a series of activities done to protect information. There are processes that can put an organisation in a critical situation in the event of a problem if we do not properly regulate them and are not effectively prepared to deal with a potential disaster. The study consists of three parts. The first part summarizes the theoretical design process of the Information Security Management System (ISMS). The second part focuses on the process of business continuity planning, with particular reference to risk management and business continuity management. The third part presents the data and information security sub-results of the research project "The Role of Self-Care in Our Lives - 2020" and a possible cybersecurity awareness program.

Keywords
Information Security Management System (ISMS), business continuity planning methodology, behavioural economics, cyber security awareness program, research project

1. Introduction

In Hungary, it is the task of the Pension Payment Directorate at the Hungarian State Treasury to pay old-age pensions, certain early retirement pensions, relatives' pensions (widow's pension, orphan's benefits, and parental pensions), benefits for people with disabilities and accident benefits. It also issues certificates, accounts and statements at the client's request or automatically, and helps clients with administration. Cases are managed primarily in writing, but people can visit the office in person, where they can book a time online. Naturally people with a customer portal can submit their request electronically, too. For the faster and simpler management of cases, the obligatory forms and information leaflets and filling guides can be accessed at the Form Library. The Pension Payment Directorate provides data for other organisations of public administration mostly in connection with legal assistance

[1] Ph.D student, Doctoral School on Safety and Security Sciences, Óbuda University, szabo.zsoltmihaly@uni-obuda.hu.

within Hungary and other issues regulated by law, and implements their decisions about pensioners and other beneficiaries concerning pension or benefit payment. Within the pension payment branch, the IT organisation, with its human, software and hardware resources, ensures the efficient and safe operation and development of the records, pension calculation and payment and other support systems. Changing laws necessitate the continuous modification and development of the systems during operation, according to the current laws.

2. Creating the Information Security Management System (ISMS)

Designing the Information Security management System (ISMS)

Nowadays, the state, all its organisations and all its citizens depend on very complicated electronic information systems within the cyberspace of Hungary. Without the proper operation of these organisations, the state cannot function and cannot provide essential services. Modern societies are not prepared to do without the blocked infrastructure, devices or services, therefore these need to be protected, with special attention to the fact that the information used and generated during their operation and the data handled within it constitute considerable national assets. By security it is meant that the activities important for the organisation can be performed uninterrupted. Systems ensuring the security of organisational activities must cover all activities that are connected to organisational strategy. Organisational and cybersecurity strategy and the Information Security Management System (ISMS) together guarantee overall and uniform security. Security is an essential element of the operation of the organisation; in the case of a state pension payment system, it has the same priority as organisational conditions (Szádeczky, 2014). Organisational security is a complex concept and the individual areas are closely related and depend on each other. The parts below have to be taken into account in planning:

- Physical security, defence of the premises;
- Information security;
- Business security;
- Safety of people;
- Incident handling and business continuity.

Taking into the account of the above, defence should be closed, complete, proportional to the risks and continuous (Muha & Krasznay, 2014). When the security management of an IT system is created, we should strive for centralization, uniform, transparent support with IT tools, automation to the highest possible degree, preferably eliminating the human factor.

The law requires that the IT systems of state organisations must be able to monitor and log critical security events of hardware and software tools essential for the operation of the organisation, and to handle these events automatically. In addition, the IT system and security management of a state organisation must also allow for the simple mapping and checking of the implementation of the security policy of the organisation. An organic part of the security management should be network, user, software and firewall management, the content filtering, and virus protection of the mailing system and other IT systems. Information security can be implemented if three basic requirements are met at the same time. These three requirements are connected to information (Michelberger & Lábodi, 2012, Dombora, 2019):

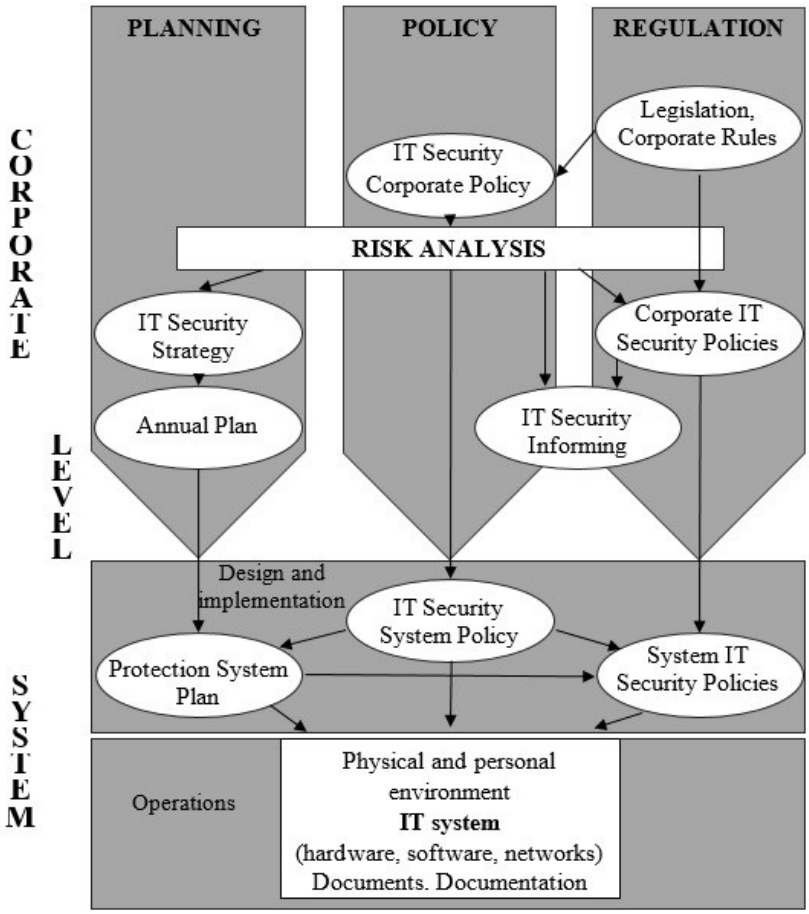
- confidentiality: only those authorized to do so can access the information,

- integrity: maintaining the accuracy and completeness of the information and the systems processing it,
- availability: the necessary information should always be available for authorized users.

If these requirements are not met, the IT system can be damaged and data can be damaged and lost (Muha - Bodlaki, 2001):

- confidentiality: other people than those authorized can access the data,
- integrity: information can change when it is transferred,
- authenticity: the information may not come from the required source,
- availability: the information is not accessible when it is required.

Fig. 3: Implementing IT defence



Source: my own editing, 2021 (Muha & Bodlaki, 2007)

Attack basically focus on data, which are surrounded by various system elements and handled by processes (Veszprémi, 2014). Threats threaten data and the processes handling data through a certain chain of system elements. Implementing defence does not only mean creating a system of tools. It means a process from design to implementation. This process covers the

full defence system of the organisation, including the protection of operation, logic (password generation, cryptography, incompatibility matrix etc.), administrative (regulation background) and human resources systems (Muha & Krasznay, 2014). In a larger organisation, where the IT infrastructure is extended and the number of applications is large, the administrative background cannot be created with a single policy and regulation, because if they go into details, then policy and regulation documents become unmanageable. Therefore, in large organisations, there should be a hierarchy of administrative protection with organisational and system levels, as shown by Figure 1.

The most important documents that govern the operation of the organisation and the most important regulations applying to the data protection of electronic information systems can be divided into two big groups:

- External documents and laws, such as Hungarian laws: Act L of 2013 on electronic information security of state and local government bodies, the information security law, Act CLXVI of 2012 on the Identification, Designation and Protection of Vital Systems and Facilities, Act CXII of 2011 on Informational Self-Determination and Freedom of Information and Act CLV of 2009 on the Protection of Classified Data. The Hungarian directives of monitoring organisations, such as: the National Cyber Defence Institute, the National Electronic Information Security Authority, the National Authority for Data Protection and Freedom of Information and the State Audit Office, and finally, international and domestic IT security standards and recommendations (European Parliament and Council (EU), MeH ITB 8, KIB 25).

- Internal documents: Security policy, IT security policy, other security policies, Security Strategy, IT security strategy, other security strategies, IT Security Regulation, Business Continuity Plan and Disaster Recovery Plan, Privacy Policy, IT Operating Order, IT Development Regulation.

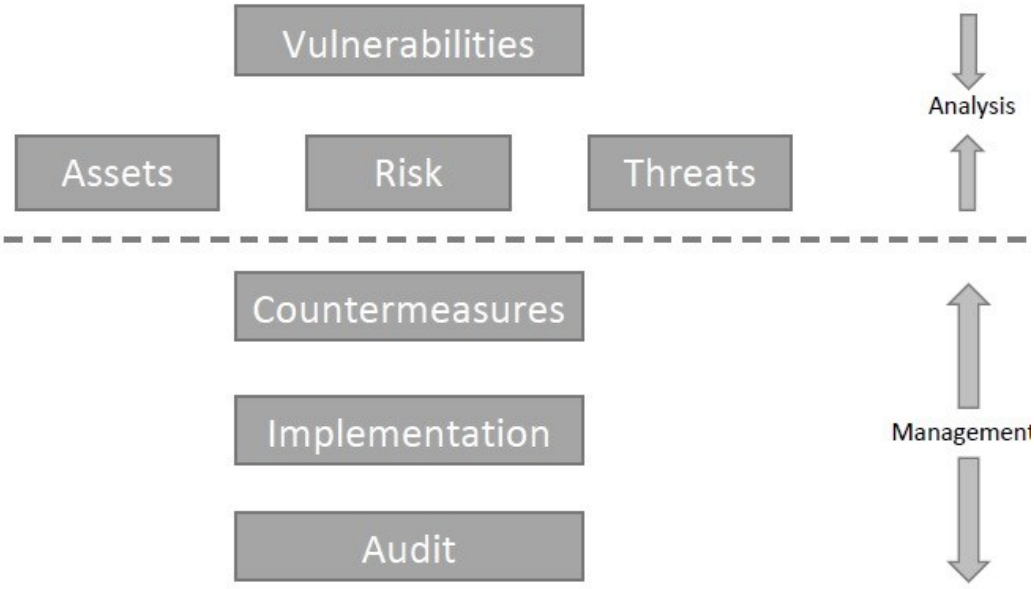
Implementing the Information Security Management System (ISMS)

Each organisation has goals, written down in their organisational strategy. The cybersecurity strategy is a review of the IT applications necessary to achieve organisational goals and the threats threatening them. The cyberstrategy of the organisation is part of the organisational strategy. The steps of outlining, strategy development, implementation planning and the fine-tuning of monitoring of IT security should be carried out in complete harmony with the cybersecurity strategy planning cycle. The two projects can be implemented simultaneously, their common elements can be handled together, their partial results can be used and they can reach the goals of the organisation together (Szádeczky, 2014).

The ISMS, i.e. planning the physical, logical and human resources defence system and setting up its administrative defence concept must be part of every IT project. The “life” of every IT system is divided into four sections starting with the decision to create it: preparation, planning, implementing, and operation. This lasts until the system is stopped or reconstructed. This life cycle approach should be applied to IT security systems, too: implementation must be prepared, designed, and after implementation, the system must be operated, maintained and further developed. Security measures afterwards can only be implemented at extra cost in most cases. Planning IT security systems consists of many steps and require much and diverse expertise. The “life” of an IT security system starts with the decision to create it. The document on the decision must include the necessity of creating the security system and the conditions of financing. Also, the minimal IT requirements must be determined. Designing an IT security system should be handled as an IT project (Muha & Krasznay, 2014).

The steps of planning should be synchronized with the usual steps of project management. In order to keep the harmony with the functional development of the IT system, security planning should be matched with the organisation's own development methodology. The full IT security project is followed by IT security auditing. The checking points of auditing and the monitoring activity must be contained in the quality assurance plan of the project. During implementation, the defence functions and measures defined in the detailed system plans are implemented in the physical environment of the hardware and software tools. During implementation, the change management applied for the whole project must be applied for security planning and implementation as well so that the final documentation and the implemented, tested system are in harmony. The defence system is tested according to the plans produced in the planning phase, and should be tested with special attention to the earlier defined weak points and potential threats: the security functions and the strength of their protection mechanism, the connections between the security functions and their potential reinforcing or weakening interactions. The testing phase is closed by a system-level test, where several security functions are tested together. Then, the test results are documented. Before the IT system starts operating in full, the system-level IT Security Regulation should be created. It is commented on by the operation manager and approved by the top management.

Fig. 4: Risk factors according to CRAMM



Source: my own editing, 2021

In planning IT security, risk management methods and tools must be used. The human factor is very important, just like in other areas. There is no risk analysis without modern methods, tools and techniques (Jenei, 2016). However, risk management is not just a collection of methods and tool (software, database and model), but also an approach, process and extended teamwork. Exploring, analysing, handling and following risks is a complex process that has to be supervised, and controlled in the organisation. The top management usually don't like new regulations because they always visualize a considerable increase in administration costs. They do not think of the effect and costs of unknown (and therefore unhandled) risks, even though they are usually far higher than the assumed administration costs. Risk management is a conscious activity which is built on risk awareness, and risk analysis and handling methods, tools and processes. The CRAMM (CCTA Risk Analysis and Management Method)

methodology, which is the MeH ITB 8 recommendation in the Information security manual, is the most often used methodology nowadays. This method describes the vulnerabilities of IT systems and makes recommendations to counter threats. Figure 2 shows the risk factors and risk management according to CRAMM (Szabó, 2017).

The three pillars of IT security are physical protection, business continuity protection (regulation) and algorithmical protection (IT tools). Sometimes, protection measures, which are basically aimed at reducing risks, cannot provide uniform security together because certain individual protection measures may be exaggerated and others may be too weak. This typical disparity makes it necessary to define protection measures based on previously determined risks and residual risks (that remain after the measures). Therefore, basis of protection measures (business continuity, physical and algorithmical) is that unacceptable risks must be reduced. The CRAMM methodology (which is also found in the ITB 8 recommendation) makes it possible to assess and analyse risks, and define the necessary measures to handle them through predefined steps. The document prepared after the risks are assessed, analysed and handled, and which defines the measures to handle the risks, is called information security concept. Planning the ISMS can be divided into the following 4 sections:

Section I: exploring what needs to be protected. In this section, the objects that need to be protected (IT systems, applications, subsystems and data) are found and selected.

Section II: threat analysis. The threats that may threaten the elements listed in Section I have to be analysed—what can harm the organisation and its clients.

Section III: risk analysis. In this section, the threat (II) and its frequency to the selected elements (I) has to be quantified. Based on this, the security risk can be assigned to the given threat–object pairs. The most common quantification method regards the product of the damage value and its frequency as the risk.

Section IV: risk management. In this final section, the given risks are either accepted or reduced. In the case of risks to be reduced, the risk management methodology prescribes the measures that reduce the given risk to an acceptable level. Table 1 shows a possible example for a CRAMM 5x5 matrix. Those fields that have acceptable risks do not include risk reduction measures. In fields that require a decision there must be a decision according to the operation of the organisation about the further management of the risk. In the case of fields with unacceptable risks, a plan of action needs to be created (Répás & Dalicsek, 2015)

Table 1: An example of the CRAMM matrix

		Effect on the organisation (damage value)				
		1	2	3	4	5
Frequency of risk	1	Acceptable	Acceptable	Acceptable	Acceptable	Decision
	2	Acceptable	Acceptable	Acceptable	Decision	Unacceptable
	3	Acceptable	Acceptable	Decision	Unacceptable	Unacceptable
	4	Acceptable	Decision	Unacceptable	Unacceptable	Unacceptable
	5	Decision	Unacceptable	Unacceptable	Unacceptable	Unacceptable

Source: my own editing, 2021

After the risks are assessed and managed, the security plan can be made. After IT security planning, the work should be summarised in a document and approved by the management of the organisation. However, making the plan and having it approved does not mean that IT security is created. The plan is only a basic requirement of IT security; later, it has to be implemented, continuously enforced and made up-to-date. In larger organisations, it is

recommended that IT security is governed at two levels, as can be seen in Figure 1. The goal of creating the IT Security Regulation (ITSR) is to define a framework for the rules that all employees must know to the degree it affects them and use in order to fulfil security requirements. Another goal of the ITSR is to regulate the security system and make it easy to check, and provide help with its planning, which help implement the technological and non-technological elements of the system, and make the required security system measureable. The Implementation Instructions are created based on the regulations contained in the IT Security Policy, and the job descriptions in the organisation. The Implementation Instructions contain the measures in the IT Security Policy which have to be used in the daily work, broken down in detail to individual operating areas. Another important role of the IT Security Policy is to ensure that the principles of data protection and the requirements of data security are observed.

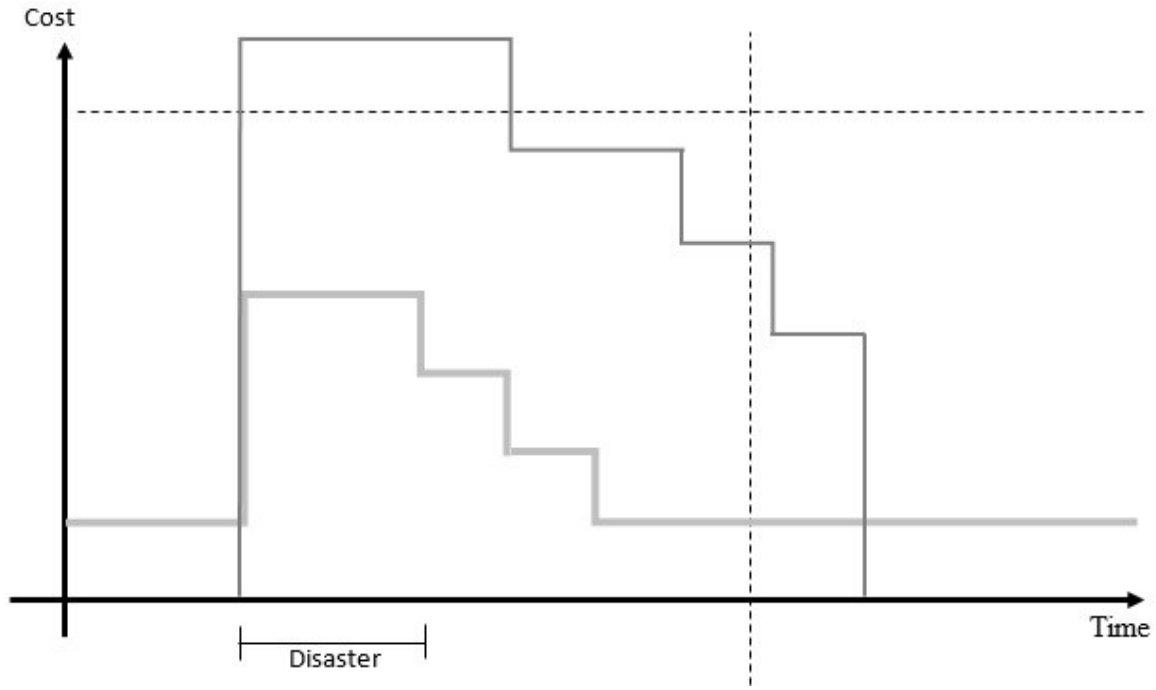
3. Creating Business Continuity

Planning business continuity

When the security of an IT system supporting administrative processes is mentioned, the most common terms used are disaster prevention/recovery and business continuity. Sometimes these are used interchangibly but that is not correct. Business continuity covers a much larger area. While disaster prevention originally meant maintaining the operation of IT systems, business continuity means ensuring the continuous operation of the basic activity of the organisation (e.g. serving clients) and minimizing financial loss. A part of business continuity is identifying the critical organisation processes necessary for organisational activity, and mapping and protection of the IT applications supporting them. Business continuity includes the processes, procedures and technologies necessary for the achievement of organisational goals, and their planning and implementation. The basic goal of business continuity is therefore providing a cost-effective solution which allows the continuation of organisational activity even in the case of unforeseen events, and thus reduces the risk of the disruption of business activity to alevel acceptable by the leadership of the organisation (Nyikes & Németh & Kerti, 2016).

Figure 3 shows two extremes: in one case, we have a business continuity plan (thick line), and in the other case (thin line), we do not. The intersection of dashed lines is when the Business Continuity Plan is not updated or the organisation does not spend enough on ensuring resources. In this case, it may happen that the system cannot return to its normal operating state within the critical recovery time. As can be seen, the organisation must find the optimal situation between the two extremes.

Fig. 5: Costs and time requirement with a plan (thick line) and without a plan (thin line)



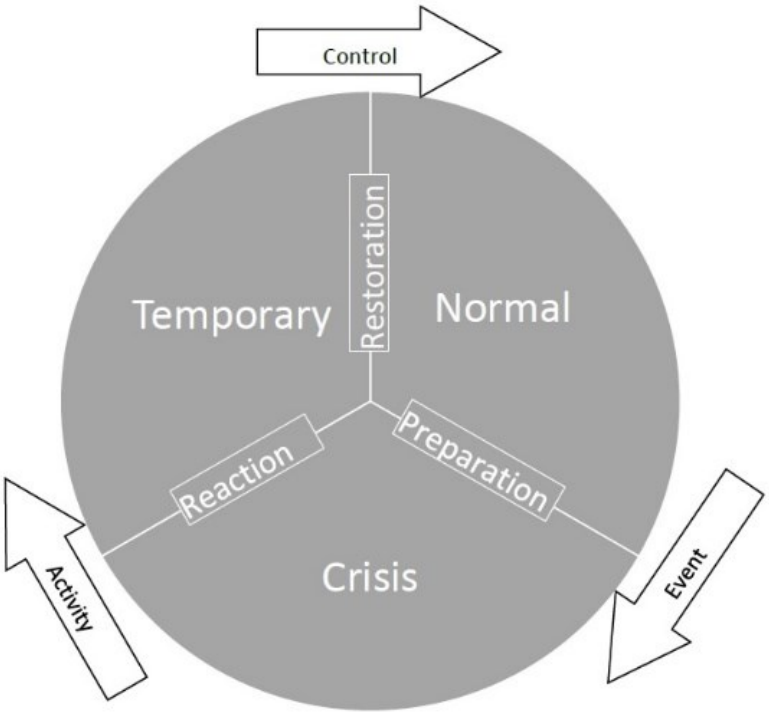
Source: my own editing, 2021

The Business Continuity Plan is the sum of tasks necessary to continuously maintain the operating processes of the organisation. It analyses

- the possible threats against the individual processes,
- the probability of the threats,
- the damage the disruption of the process may cause,
- and defines the procedures necessary to maintain the functionality of the organisation based on risk analysis.

The IT Disaster Recovery Plan is based on the Business Continuity Plan. It contains action plans and tasks that need to be implemented if processes critical for the operation of the organisation and the resources supporting it are damaged to such an extent that the critical processes cannot be maintained, or are unacceptably limited concerning the functionality of the organisation. The Business Continuity Plan and the IT Disaster Recovery Plan are strongly interrelated and are worked out together. They provide effective responses and solutions for business continuity and handling crises (Dombora, 2018).

Fig. 6: The process of business continuity



Source: my own editing, 2021

Planning business continuity is a complex and costly process, therefore it is recommended that it is implemented like a project, in cooperation with a project management. As in the case of all security measures, the cost of business continuity planning and disaster recovery planning cannot be neglected. The solution most appropriate for the goals of the organisation should be selected, taking into account of the costs. Loss of income and other damage should also be calculated. For planning business continuity, the three phases below should be separated, in addition to preliminary planning, as shown by Fig. 4:

- Immediate reaction phase: the period immediately after the unforeseen event has happened;
- Transient phase: providing the minimal services concerning the processes critical for the organisation;
- Recovery phase: returning to normal operation.

Experience shows that in larger organisations, decision-makers cannot fully understand the IT complexity and pitfalls of implementation, and assess the risks resulting from them. Managers responsible for applications and the operation of the infrastructure do not fully know the priorities of business. Therefore, there may be a large gap between what the two groups see and bridging this gap is an element of major importance for business continuity planning, and perhaps its greatest challenge. All factors endangering the safety of operation must be taken into account and providing complete protection against all is not only hopeless but it is not worth either, since protection about most unforeseen events costs far more than the damage they can cause. Therefore, before business continuity planning, possible threats should be

prioritized based on the preferences obtained from professional organisations, their risk must be analsed and the optimal soultion has to be found based on this (Rajnai & Puskás, 2015).

Table 2: Five-level business continuity maturity model

Level 5	Dedication of the top management – integrated cross-functional planning, maintenance of business continuity plan			Optimized
Level 4	Dedication of the top management – Regular business continuity planning	Managed – planned		
Level 3	Dedication of the top management – tested business continuity plan	Defined – Conscious	Supported, extends to all organisational units	
Level 2	Based on organisational units – repeatable	Business continuity plan documented in some units, some dedication and support form management		
Level 1	Starting level	No business continuity plan, the management is not dedicated; self-regulating		

Source: my own editing, 2021

When planning business continuity solutions, we have to start from the existing business continuity state, and have to exploit the existing possibilities. We primarily examine the preparedness of an organisation to avert an unforeseen disaster event from a personnel-organisational and IT aspect. In addition to the IT background of business continuity, the maturity of the organisation is decisive - there are several methods to assess it. These models categorize the preparedness of the organisation based on the dedication of the management, the level of detail of business continuity processes (levels of business continuity) and scope (5-level business continuity maturity model), as Table 2 shows.

The theoretical and methodological knowledge mentioned earlier can show the current IT security state of the organisation, the areas to be developed can be determined exactly based on the requirements of the organisation, as can the degree of lag and the most important tasks. At the end of this section, therefore, we can outline the exact task to be done.

Implementing business continuity

The business continuity planning below was done at a state pension payment institute. The important methodologies are shown in practice in a possible case. The pension payment institute started its business continuity project to achieve its numerous goals defined in its organisational and IT strategy. The project was aimed at improving the reliability of the services of the organisation, and ensuring its uninterrupted operation by preparing the organisation for unforeseen events considerably influencing business continuity (disasters). The project consisted of three parts:

1. General IT security audit;
2. Risk and business impact analysis;
3. Creating the Business continuity Plan and the IT Disaster Recovery Plan.

Since IT support systems are indispensable for the organisation, the project started with a full-scale security audit. The results of this phase were summarized in an independent report. For the Business Continuity project, we mostly used the results connected to the continuity of

the systems, which had to be included in the action plan. The basis of business continuity planning was a risk-damage analysis, in which the business processes of the organisation and the IT systems supporting them were reviewed. Organisational risk was the damage, which included image loss, and potential loss of of clients, in addition to the financial loss. The risk classification of IT systems and the data they handled was based on the damage resulting from accidental or intentional loss of integrity, availability, confidentiality of the data. All information systems had to be provided a minimal level of protection, which included physical and logical protection procedures, and the use of a virus protection system and an authorization system.

Table 3: High priority processes (part of the IT Disaster Recovery Plan)

Name of the Process (System)	System priority level	Time to avert
Pension payment IT system	5 (Critical)	4 hours
Pension payment data query system	5 (Critical)	4 hours
Pension assessment	5(Critical)	4 hours
Filing system	5(Critical)	4 hours

Source: my own editing, 2021

A section from the risk classification of the information systems of the organisation can be found in Table 3. The possible protection solutions, the necessary aversion time and availability had to be assigned to the priority level of individual information systems. The systems have to be ranked by the person responsible for the system, system organizer and systems administrator and the organisation commissioned with the handling of IT risk, with the leadership of the manager responsible for IT security. The ranking of systems has to be supervised at least once a year. The processes that have action plans from the business continuity planning phase are selected together. The form contains the summary of data previously processed in parts. The form contains all the dependencies of the process, its risk classification and other important characteristics in detail, as cimpliance with the law (Phuoc & Kerti & Rajnai, 2016).

The third phase of the project is creating the action plan, the IT Disaster Recovery Plan, whose main aim is to ensure the operation of the organisation even in the case of unwanted events. Its other goal is to ensure continuous operation in the case of such a event with minimal cost and the shortes possible time of imoperation. The action plan contains the high-priority processes, such as pension payment, which depend on the pension payment system. The Business Continuity Plan has to be taught at the managerial, operational and end-user level as well. The importance of business continuity has to be stressed, the basic information has to be taught, as well as the contents of the prevention and recovery plan. The Business Continuity Plan can be tested with a simulated event and recovery according to the plan. This way, the operator and the users can practice the contents of the action plans in a real-life situation. Then, the results are evaluated, and the plan may be modified based on this. After successful testing, business continuity planning is finished. According to the decision of the management of the organisation, the Business Continuity Plan is modified and updated at regular intervals and after bigger changes in the organisation. Updating includes a review of the various organisational

processes, deciding which processes can be considered critical at times of maintenance, a listing of IT applications in the organisation (e.g. maintenance of forms), and whether there is new hardware or there are new network elements etc. (maintenance of the IT Disaster Recovery Plan and the Business Continuity Plan). After the plan is updated, it has to be presented to the employees.

4. The Role of Self-Care in Our Lives - 2020

Security in our decisions

The primary research is directed at both the present and the future state of the individuals. We would like to know what influences the respondents in deciding about self-care, what habits, processes influenced the decision, in other words, what motivated them to choose the given form of saving. In the research project, the respondents filled in a questionnaire, then we examined them with the methodology of behavioural economics and factor analysis to see what choices we have about our expected future level of pension. The research examines the role of self-care as a supplementary pension pillar in public thinking and our decisions, what possible pension system young people consider desirable in the future, what will comprise the pension of the future generation, how many people will work and how in the future. The basic assumption of the research is that people think about pension with fear and uncertainty. Since the PAYG system is in crisis, the second pillar, self-care, receives more and more attention.

We processed the data of the survey with statistical methods. The online survey was conducted in 2020. The total number of respondents was 1410 (n=1410). The basic questions we sought an answer to were connected to pension systems, pension savings and self-care, and pension security because these are the elements that determine our financial background in the future, that is, the degree of self-care. The responses were divided into three groups: 1) Knowledge of pension systems (mandatory, voluntary); 2) Financial planning (different kinds of savings); 3) The role of self-care (mapping the personality).

In the present study, we only examine data and information security issues connected to the pension system. Based on the answers to the questionnaire, as shown by Table 4, the respondents consider online services important and data and information security are of outstanding importance.

Citizens can use countless services through e-administration, among others services connected to pension payment. Requests for pensions can be submitted electrinically too. The e-administration system allows people with a customer portal to fill in and submit electronic forms.

Cybersecurity awareness programmes

Today, we live in an information society. There are more and more IT devices, such as laptops, tablets, smart phones, smart TVs, smart fridges etc. Also, the direction of technological development is to produce more smart devices, which results in an enormous amount of data and data transfer when these devices are connected to the Internet. One of the most important goals of information security awareness and the related IT security programmes is to properly create and use the information security and IT safeguards necessary for the healthy operation of organisations. Each organisation must creat its information security awareness improvement and IT security training programmes according to the needs and the existing and planned safeguardsThe training programmes must include not only the internal regulations of the organisation but also the applicable laws. The IT security manager of the organisation has a fundamental role in creating these training courses: they not only contribute to the content of the course but also typically take part in teaching it, too (Mádi-Nátor & Kardos, 2014).

Table 4: Answers about security

How important is security in savings?		
Not at all	4	0.3%
A little bit	13	0.9%
Fairly	220	15.6%
Very	1173	83.2%
How important are online services?		
Not at all	41	2.9%
A little bit	235	16.7%
Fairly	774	54.9%
Very	360	25.5%
How important is data security?		
Not at all	10	0.7%
A little bit	37	2.6%
Fairly	227	16.1%
Very	1136	80.6%
How important is information security?		
Not at all	8	0.6%
A little bit	28	2%
Fairly	251	17.8%
Very	1123	79.6%

Source: my own editing, 2021

As can be seen above, respondents consider online services important and also think data and information security are of paramount importance. But what should they do to use online services securely when they may not have received information security awareness training? The state and organisations have an important role in this because they have to provide secure services, and training courses and awareness programmes, as much as possible. Fortunately, states and organisations have recognized this problem, and more and more cybersecurity training courses and awareness materials or programmes are available. I would like to stress the role of the Hungarian National Cybersecurity Institute and the National Data Protection and Freedom of Information Authority, and the European Union also regards cybersecurity as especially important, where data and information security awareness materials and programmes are made for the citizens.

IT security awareness education

The organization of information security training and further training is primarily the responsibility of the National University of Public Service. The training of organizational leaders, experts, contributors and employees began on the basis of the Information Security Act, which was enacted in 2013, and about 150 people have successfully completed the courses

at NKE so far. The task of the Cyber Security Academy, which has been operating since March 2017, is to coordinate the existing resources of certain faculties, research institutes and workshops of the National University of Public Service, and to support cyber security research and specialists. They initiate and organize training programs and plan cyber security exercises. The aim of the academy is to act as a catalyst in cyber security training in the fields of public service, law enforcement and national defense.

Table 5: Stages of preparing the IT security awareness material

Project phases	Milestones	Finished tasks	Notes
Conclusion of a contract	28/12/2017	The User asks the Author	Defining project requirements
Planning and preparation	29/12/2017 - 01/11/2018	Elaboration of interview questions	Structure of the specialization, determination of chapter titles
Transaction	01/12/2018	Interviews with the four professional leaders	Taking into account information security trends
	13/01/2018		
	16/01/2018		
	17/02/2018		
	18/02/2018 - 15/03/2018	Processing the literature related to the topic	
Closing	15/03/2018 - 30/03/2018	Finalization, proofreading and submission of the material	Under the terms of the contract

Source: my own editing, 2021

In its Decision C (2015) 5262, the European Commission approved the Operational Program for the Development of Public Administration and Public Services (hereinafter: KÖFOP) and the Government Decree 1004/2016 establishing the annual development framework for the Operational Program for the Development of Public Services and Public Services (I. 18.) in its Government Resolution established the annual development framework of KÖFOP, which in its Annex No. 2 named KOFOP-2.1.1-VEKOP-15-2016-00001 “Complex competence of public service, career development program and technological development of education” as a priority project. project. In connection with the project, I participated as an author in the preparation of the professional material belonging to the training program "Annual training for the person responsible for the security of the electronic information system - Targeted cyber attacks" (50 hours). The 20-page professional material entitled "Targeted Attack on the Public Administration Sector" was completed by March 30, 2018 (Bodó & Cser & Gyaraki & Kaczur & Lattmann & Solymos & Szabó & Tikos & Váczi & Zámbó, 2018). Table 5 shows the table of contents of the document, which can later be accessed and downloaded as e-learning curriculum from the website of the Specialized Continuing Education Department of the Faculty of Political Science and Public Administration, Center for Management and Continuing Education of the National University of Public Administration. The professional

material has been completed, so the information security training project can be said to be successful and successful. Success was determined by meeting and achieving project milestones and milestones by meeting the conditions set out in the contract. In each case, the chapters start with the processing of the literature and the theoretical background, and then the interviews with the managers turn the theoretical knowledge into knowledge that can be applied in everyday practice. The economic background of the project was provided by the European Social Fund and Széchenyi 2020. It is financed on a pro rata basis from the Competitive Central Hungary Operational Program (VEKOP) and the KÖFOP: 80.9% of the total eligible costs of the project are paid from the Convergence (KÖFOP) source, 19.1% from the KMR (VFKOP) source. , taking into account that the defined tasks are related to the project with the identification number KÖFOP-2.1.1-VEKOP-15-2016-00001.

5. Conclusion

The thorough planning of information security defines the IT resource and investment needs, and defines the framework in which the priorities are ranked, responsibilities for implementation are assigned and the resources are concentrated to the key areas. Detailed preliminary planning is necessary to exploit all the possibilities of IT to support the aspirations and goals of the organisation. Planning ensures that the solutions to be used are within the budget, technologically feasible, and can be controlled and easily interpreted by all the people involved. The use of up-to-date planning and design methods is indispensable for a the planning of a complex IT security system. As can be seen from the study, very many aspects need to be taken into account. The security of the information systems of the state are very similar to the security problems of other IT systems but the above-mentioned factors require special procedures, tools and solutions. Planning is always orgnaization-specific; the structure and characteristics of the organisation have to be taken into account. Without this, a secure IT system is not possible to create. All employees and external workers of the organisation must know and use the guidelines for information security and rules applying to the use of IT systems. Finally, it must not be forgotten that it is necessary to alwas improve and update our knowledge of IT security due to the complexity and continuous development of the topic. In the future, it is expected that security experts will often hear the term multi-layer protection. It is no coincidence that we need to be prepared for complex threats that are no longer sufficient for traditional security technologies or protection systems that operate in isolation. An adequate security level can only be achieved if different levels of protection work in concert with one another. This aspect should be considered when purchasing and deploying new tools. There is no single magic weapon, an all-in-one security solution. The key to multilevel protection and collaboration lies in advising professionals and researchers as well.

6. Resources

BODÓ, A. P., CSER, O., GYARAKI, R. E., KACZUR, G., LATTMANN, T, SOLYMOS, Á., SZABÓ, Zs. M., TIKOS, A., VÁCZI, D., & ZÁMBÓ, N. (2018). Targeted cyber attacks. Annual training for the person responsible for the security of the electronic information system. National University of Public Service. Hungary. 156-167.

DOMBORA, S. (2019). Parameters and Guidelines of Enforceable Information Security Management Systems. Interdisciplinary Description Of Complex Systems 17: 3A, 485-491.

DOMBORA, S. (2018). Integrated Incident Management Model For Data Privacy And Information Security. Book of Proceedings Bor. Szerbia. 430-439.

JENEI, T. (2016). Comparison of the most frequently used models of risk management, International Journal of Engineering and Management Sciences (IJEMS) Vol. 1. No. 1, 1-11.

MÁDI-NÁTOR, A., & KARDOS, Z. (2014). Information security awareness practice. National University of Public Service, Hungary. 1-50.

MICHELBERGER, P., & LÁBODI, Cs. (2012). After Information Security - Before a Paradigm Change: A complex Enterprise Security Model, Acta Polytechnica Hungarica 9:(4), 101-116.

MUHA, L., & BODLAKI, Á. (2007). IT security. Hungary. Pro–Sec Kft. 176.

MUHA, L., & KRASZNAY, Cs. (2014). On the security of electronic information systems for managers. National University of Public Service. Hungary. 8.

MUHA, L., & KRASZNAY, Cs. (2014). Security management of electronic information systems. National University of Public Service. Hungary. 10.

NYIKES, Z., NÉMETH, Z., & KERTI, A. (2016). "The electronic information security aspects of the administration system," 2016 IEEE 11th International Symposium on Applied Computational Intelligence and Informatics (SACI), Timisoara. 327-332.

PHUOC, D. KERTI, A. & RAJNAI, Z. (2016). E-Learning Security Risks and Countermeasures, International Journal of Emerging Research and Solutions in Ict 1:(1). 17-25.

RAJNAI, Z., & PUSKÁS, B. (2015). Requirements of the Installation of the Critical Informational Infrastructure and its Management, Interdisciplinary Description of Complex Systems 13: (1). 48-56.

RÉPÁS, S., & DALICSEK, I. (2015). Methodological issues of information security risk analysis for organizations operating critical infrastructure elements. Pro Publico Bono - Public Administration Vol. 3 No. 4, 22-33.

SZABÓ, Zs. (2017). The Information Security and IT Security Questions of Pension Payment. Key Engineering Materials Vol. 755, 322-327.

SZÁDECZKY, T. (2014). Information security standards. National University of Public Service. Hungary. 43-44.

VESZPRÉMI, B. (2014). Data protection, freedom of information. National University of Public Service. Hungary. 3-6.

COMPOSITION DIAGRAM OF A COMPLEX PROCESS A NEW METHOD FOR DISPLAYING COMPLEX BUSINESS PROCESSES

Peter Schmidt[1], Pavol Jurík[2]

Abstract
In practice, we often encounter business processes with a complicated hierarchic (i.e., vertical) structure of sub-processes. The flow of the individual sub-processes can be depicted using traditional diagram techniques, such as flowcharts, Business Process Model and Notation (BPMN) diagrams, activity diagrams, etc. These techniques seem to be the best tool to display the sequence of activities creating a single process (i.e., the horizontal structure of a process). However, they are not well suited for displaying the vertical structure of a process, which consists of many levels of interconnected sub-processes. With a large number of sub-processes, the overview of how exactly the individual sub-processes relate to each other and what is their exact position or role in the parent process can be lost, if these techniques are used. Therefore, we have developed a new diagram technique, which we have called Composition Diagram of a Complex Process or a Jurík-Schmidt's diagram. In this paper we would like to present the rules how to use this technique. The consequent effect of these rules can be referred to as a Jurík-Schmidt's method for displaying complex business processes.

Keywords
Business process modelling, diagram technique, complex process, hierarchic structure, hierarchic decomposition

1. Introduction

A business process (BP) can be defined as "a workflow flowing from one person to another and, in the case of a large process, probably from one department to another" (Robson & Ullah, 1998). According to Davenport and Short, "a process is a series of logically linked tasks performed to deliver a defined business output" (Davenport & Short, 1990). A business process is characterized by its repeatability. It represents a steady and repeatedly performed activity taking place in a particular company, which consists of a number of steps designed to transform a certain set of inputs into a set of outputs and to achieve a predetermined objective. A business process is; therefore, a generalization of a complicated business activity and it represents a workflow. One particular execution of a process is then referred to as a process instance. As we mentioned earlier, every process can be executed repeatedly and there can be many instances of the same process in the state of a simultaneous execution (Jurík, 2018).

A process may have its horizontal and its vertical structure. The process horizontal structure is represented by a sequence of activities in a specific order leading to the pre-defined goal of the process. Some of these activities may be performed in a parallel manner in order to shorten the total process execution time if it is possible. All activities forming the horizontal structure of a process are on the same hierarchical level. On the other hand, the vertical structure

[1] University of Economics in Bratislava, Faculty of Economic Informatics, Department of Applied Informatics, Dolnozemská cesta 1, 852 35 Bratislava, Slovakia, peter.schmidt@euba.sk

[2] University of Economics in Bratislava, Faculty of Economic Informatics, Department of Applied Informatics, Dolnozemská cesta 1, 852 35 Bratislava, Slovakia, pavol.jurik@euba.sk

of a process represents a hierarchical breakdown of the process activities into several levels of sub-processes until we gradually get to elementary steps that cannot be further broken down (Jurík, 2020). A process that consists only of elementary steps that cannot be further broken down can be referred to as an elementary process. In other words, an elementary process is a process that can be described by its horizontal structure only (i. e. it has no vertical structure). On the contrary, a process with a vertical structure can be referred to as a complex process. Thus, a complex process is a process that can be hierarchically broken down to several levels of sub-processes until we gradually get to elementary steps that cannot be further broken down.

The main goal of this article was to create a simple (easy to write and easy to read) new diagram technique able to provide a managerial (i.e., global) overview of the structure of a complex process at the highest level of abstraction. The partial goal was to demonstrate how to use this modelling method on an example of a complex business process.

2. Business process visualisation - state of the art

The BP horizontal structure might be visualized with the use of standard diagram techniques, such as flowcharts, BPMN (Business Process Model and Notation) diagrams, Activity diagrams, etc. However, they are not able to capture a hierarchical breakdown of a complex process into several levels of sub-processes in a single diagram. Of course, these techniques are able to depict the vertical structure of a complex process in the form of a few interconnected diagrams but with a large number of sub-processes the global overview of the process structure can be lost using these techniques. Moreover, each of the standard diagram techniques has its individual shortcomings related to complex processes with a complicated vertical structure, which we have described in detail in another article already (Jurík & Schmidt, 2020). The process thread diagram, which is part of the tools of the Select Perspective methodology (Kanisová & Müller, 2006), is indeed intended to provide a global overview of the process, but it has some disadvantages mentioned above and cannot capture some situations that may occur in real life (Schedlbauer, 2010). The notation of BPMN diagrams is explained by Révészová and Paľová very well (Révészová & Paľová, 2009).

The main goal was to create a simple (easy to write and easy to read) new diagram technique able to provide a managerial (i.e., global) overview of the structure of a complex process at the highest level of abstraction. This managerial overview of a complex process is important not only for the BP *owner*[3] within its flow and execution managing, but also for designers of an information system which plays a role of the BP information support tool.

The new diagram technique can be referred to as a composition diagram of a complex process or as the Jurík-Schmidt's diagram. This technique was introduced in another article shortly, where we have described its symbols and their basic meaning (Jurík & Schmidt, 2020). In this article we would like to present exact rules how to use this technique in order to display complex business processes. These rules can be referred to as the Jurík-Schmidt's method for displaying complex business processes.

[3] *The owner of a process* is a person who is responsible for achieving the goal of the process and for its long-term functioning. This person is responsible for monitoring the performance of the individual process instances and for systematic improving the flow of the process and its rules in order to achieve best results possible.

3. Research methodology

Based on many years of teaching experience, we found that it is very difficult for most students to see the dependencies across individual diagrams. Without understanding these contexts, it is not possible to perceive the vertical and horizontal side of a process at the same time, which results in a misinterpretation of the described process. Various techniques exist to facilitate communication between the designer and the customer, such as UML, flowcharts, etc. However, even while using these techniques, it is often necessary to use several diagrams in order to capture the essence and know how to properly explain the solution to the customer. With complex processes of a slightly more complicated nature, this no longer works and customers do not understand the interconnections among the diagrams and the solutions had to be explained verbally, which often leads to great inaccuracies and misinformation. These two reasons led us to the idea of creating a new method and diagram technique that would be easy to understand for customers and pleasing even in teaching, and would include both vertical and horizontal planes.

4. Results and discussion

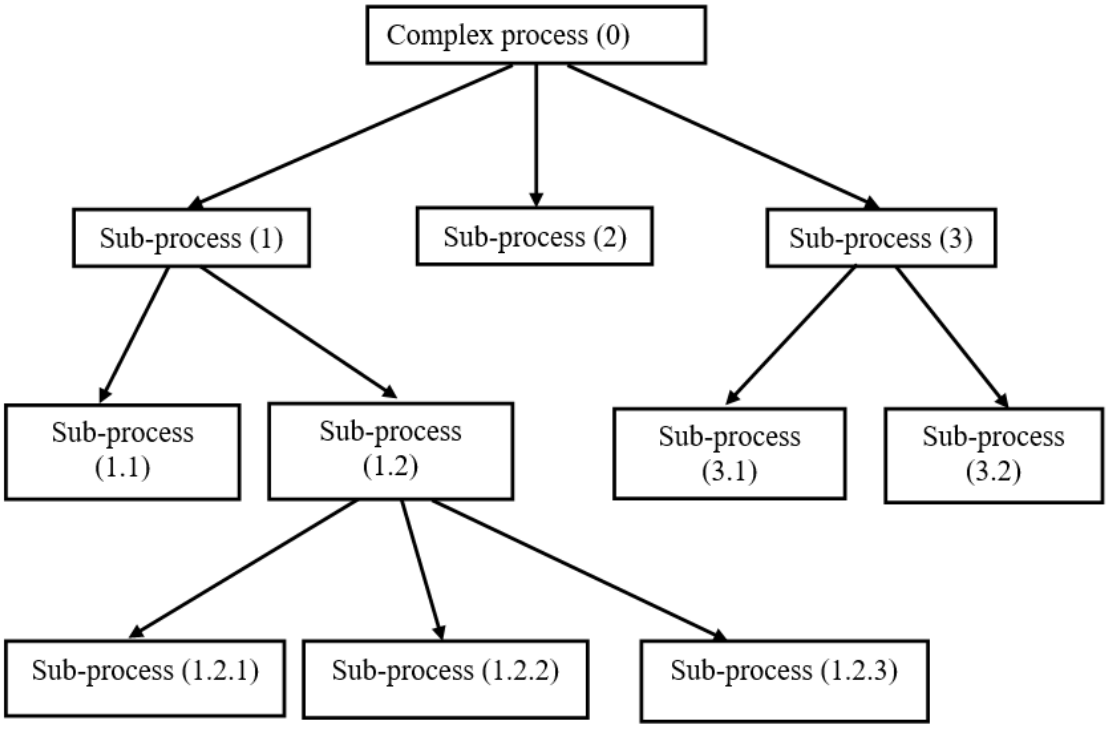
4.1 The Jurík-Schmidt's method for displaying complex processes

In this chapter, the Jurík-Schmidt's method for displaying complex business processes using the composition diagram of a complex process, also called the Jurík-Schmidt's diagram is described. However, before doing that, the more precise explanation of the term process should be made as well, while a complex business process is created by set of subordinated processes and the sub-processes might contain further subordinated processes within different hierarchic levels.

The highest hierarchical level in this structure is being created by the complex business process denoted as a core process. The highest hierarchical level can be marked as the zero level. Processes at the first hierarchical level are core process sub-processes at the zero hierarchical level. Furthermore, there are core process sub-processes denoted as the main processes, while they might content further subordinated processes at different hierarchic level

In Fig.1, there is shown an illustration of a complex process vertical structure, however, this is not the Jurík-Schmidt's diagram technique. The simple schematic representation of a complex process visualized in Fig.1, which resembles a process map, does not contain a lot of important information, about the expected duration of the individual processes, about the owners of individual processes, about their trigger events, inputs and outputs, possible cycles or branches. etc. Therefore, we have created a special new diagram technique that allows all this information to be included in a comprehensible way, which makes it a suitable tool for representing a global view of a complex process vertical structure.

Fig. 1: An example of a complex process structure



As we have already outlined, we understand a complex business process as a continuous chain of sub-processes on a lower hierarchical level. Each sub-process can again be formed by a continuous chain of sub-processes on an even lower hierarchical level. The hierarchical structure of a compound process is thus actually a hierarchical arrangement of such chains of sub-processes. *According to the Jurík-Schmidt's method for displaying complex processes* ***each chain of sub-processes should follow these rules****:*

- *A chain can start with* ***a trigger event*** *(a complex process as a whole should have a trigger event defined, but chains at lower hierarchical levels don't have to have a trigger event, only if it is necessary based on the specifics of a particular chain)*. A trigger event is an event that starts the execution of the first sub-process in the chain. If the trigger event does not occur, execution of the chain can not begin and this event should be waited for. The trigger event is drawn as an empty arrow. If it is necessary to supply any inputs to the chain, then these inputs are written inside this arrow. In general, the inputs can be of a material, financial or informational nature. An example of a trigger event may be e.g. receiving an order from a customer. This order itself is also understood as an input and the act of receiving it is the trigger event. In this case, the event of receiving the order starts the chain of interrelated sub-processes ensuring processing of the order.
- *Each chain is made up of* ***a sequence*** *of sub-processes, but it can also contain* ***branches*** *or* ***loops*** *if necessary.*
- *The execution of individual sub-processes in a chain may be conditioned by the fact that* ***a transition event*** *should occur.* For example, it may be necessary to receive a specific

document, signal, material, or other specific stimulus at first because without this stimulus (i. e. without the fulfillment of the transition event) the chain of sub-processes cannot be started.

- *Each sub-process in a chain* ***can be itself an individual chain of sub-processes****, which are present at a lower lower hierarchical level.*
- *A chain can be terminated by* ***an output symbol*** *if it is supposed to give a specific material, financial or information output as a result (the complex process as a whole should be always terminated by an output symbol).*
- *The individual chains of sub-processes are connected to each other* ***by arrows****, depending on the order in which they are to be executed*. If two chains are not connected by an arrow, then their execution shall start at the same moment (i. e. these chains shall be executed **in a parallel manner**).
- *We also distinguish a specific type of parallelism, where a secondary branch is separated from a primary branch, and these two branches no longer need to be merged*. The secondary branch is just a digression from the primary branch and it is usually associated with carrying out some trivial activity, which is not important for the further course of the process, such as sending a message for informative purposes only.
- *If it is necessary to redirect the flow of the process from a certain place in the diagram to another remote place and we do not want to connect these two places with an arrow because it could affect the comprehensibility of the diagram in a negative way* (because many arrows and the arrow crossing are not suitable) *a pair of symbols denoted as the* ***redirection symbols*** might be applied. The pair of symbols represents a kind of teleport, which connects two distant places and progresses the course of the process from one place to a different place in the diagram.

The following rules should be applied related to the complex process and sub-process at any hierarchic level:

- *Each process or sub-process is drawn using* ***a frame that contains three basic pieces of information:*** the name of the process or sub-process, its expected duration, and the specification of the owner of the process or sub-process.
- *If the particular process or sub-process consists of one or more other sub-processes, then they might be drawn via separated frame inside the frame for their parental process* ***on the principle of a matryoshka doll***. Each frame should again contain three basic pieces of information, namely: the name of the sub-process, its expected duration and the specification of the owner of the sub-process.
- *For each process or sub-process, it is necessary to specify its* ***expected execution time*** *based on the expected execution times of its individual sub-processes.* For example, if the process consists of a sequence of three interrelated sub-processes, each of which has its own expected execution time, then the total execution time for the parental process is *the sum of these sub-times*. The proper determination of the aggregate expected time for the parental process is complicated by the fact that some of its sub-processes may be planned to be executed in a parallel manner. For example, if the parental process consists of two sub-processes that shall be executed in a parallel manner and each of

these sub-processes has a different expected execution time, then the total execution time for the parental process is *equal to the higher of these two partial times*.

- *For each process or sub-process, it is necessary to* ***specify its owner****, that is a person who is responsible for the successful execution and operation of that particular process or sub-process.*

The following rules should be applied especially related to the entire complex process:

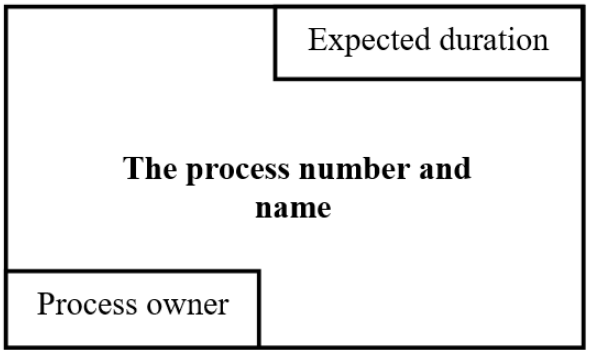
For the complex process as a whole **a trigger event should be specified**, which is also a trigger event for the chain of sub-process at the first hierarchical level. Each time when the trigger event occurs, a new instance of the complex process is getting started. Several instances of the same process can be running and be in an unfinished state at the same time. If the trigger event does not occur, the start of a new process instance might not be done.

- *For each complex process an appropriate set of* ***inputs and outputs*** *should be assigned, while* the inputs as well as outputs might be of a material, financial or an information nature. A complex process can produce some outputs not only at the end of its execution, but also during its course (i. e. while it is running it can produce some intermediate outputs besides the final outputs).
- *Each complex process should be terminated by the* ***symbol of an end of its execution****.* However, there may be several qualitatively different ways in which the same process can end. In other words, there may be multiple qualitatively different end states that can be reached at the end of the process execution (of course, every instance of the process may end only in one of these end states, but there may be multiple possibilities how the individual instances may end). If this is the case, the process contains a plurality of branches terminated by the symbol of an end where each end represents a different end state.
- *Each end symbol indicates* ***an immediate termination*** *of the execution of the current process instance regardless of whether its course was successful or not.* Thus, if a complex process contains multiple branches ending with an end symbol, then the execution of the current instance of the process should be terminated immediately when one of these end symbols is reached.

4.2 Description of symbols used in the Jurík-Schmidt's diagram technique

Every process or sub-process should be displayed using the following frame, which is portrayed on figure 2.

Fig. 2: Frame representing a process



The sub-processes that form a parent process together are drawn inside this parent process under the principle of a matryoshka doll as displayed on figure 3. If the execution order of the individual sub-processes matters, then we should connect them with arrows. If their execution order is not important, then we omit the arrows, thus, the sub-processes remain unconnected. We distinguish tree types of flow from one sub-process to another. These flow types are depicted in figure 4.

Fig. 3: A process and its sub-processes

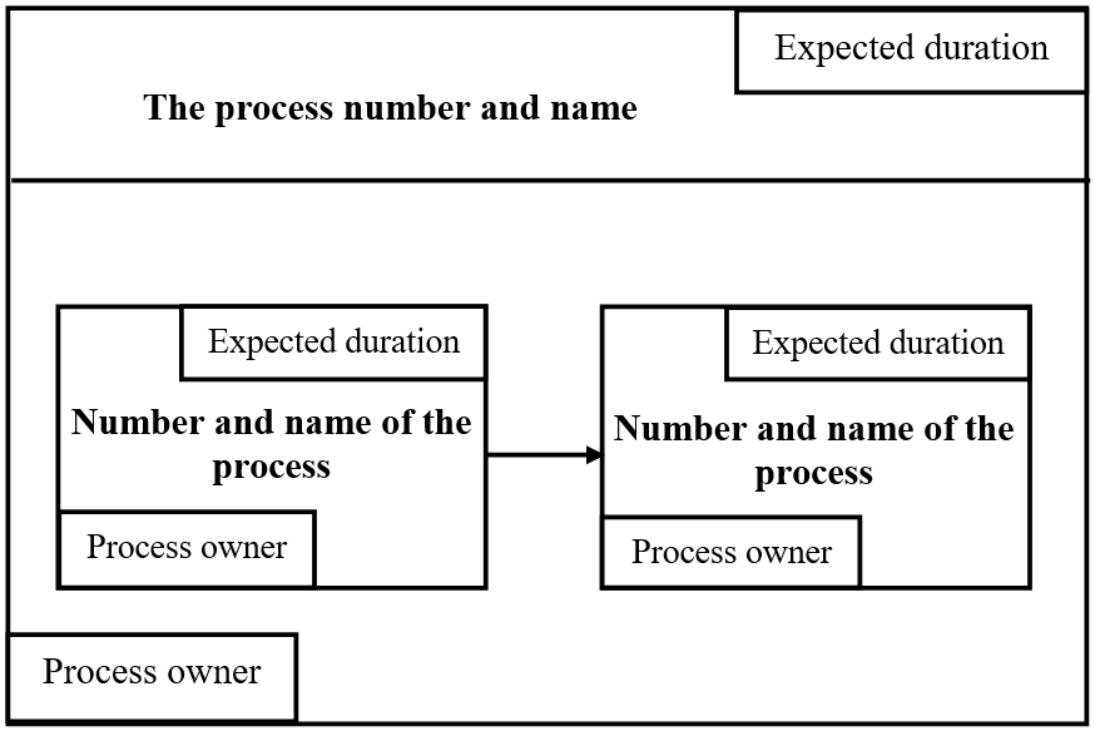


Fig. 4: Different types of flows between sub-processes

Flow type	**Symbol**	**Meaning**
Physical flow	Name of the physical object →	The output of a sub-process is a physical object which is transferred as an input to the following sub-process.
Information flow	Name of the information or document →	The output of a sub-process is information which is transferred as an input to the following sub-process. Typical examples are requirements, orders, signals, documents, reports, statistics, etc.
Sequence flow	→	It only expresses the sequentiality of sub-processes. There is no special output that should be transferred from one sub-process to another.

A **trigger event** is an event that has to occur to start a new instance of the complex process. However, it can also be used to start a chain of sub-processes. A trigger event is drawn

by a symbol of an empty arrow as pictured on figure 5. It can be an internal or an external event. An example of an internal trigger event is the drop in the amount of material in stock below a critical threshold value, which is a signal to trigger a new instance of a material purchasing process. An example of an external trigger event is the arrival of an order sent by a customer, which triggers a new instance of a customer order processing process. The symbol of an empty arrow may also symbolize a process input.

The **output of a whole complex process** or **the output of its individual sub-processes** is drawn by a symbol of a full arrow as displayed on figure 5. As we said earlier, there may be situations in which the complex process has to provide some output (for example a document), but this output does not mean the termination of the execution of the process instance. Thus, in these situations the execution of the process instance has to continue after providing the output. Because of this we use a separate symbol for a process output, which is different from the symbol for the end of a process instance execution.

Sometimes, it is necessary to wait for a special event to occur in order to continue with the execution of a process. Such an event can be marked as a **transition event.** For example, it is necessary to receive a special document, signal, material, or an instruction. A transition event is drawn by an oval symbol, which is portrayed on figure 5.

The **end of a process instance** execution (i.e. the place where the execution of the process instance has to be terminated) is drawn by a symbol of a circle with the letter "E" inside as pictured on figure 5. A symbol for the start of a process instance is not needed because it is represented by the symbol of the trigger event, which was introduced already. When the trigger event occurs, the execution of a new process instance starts.

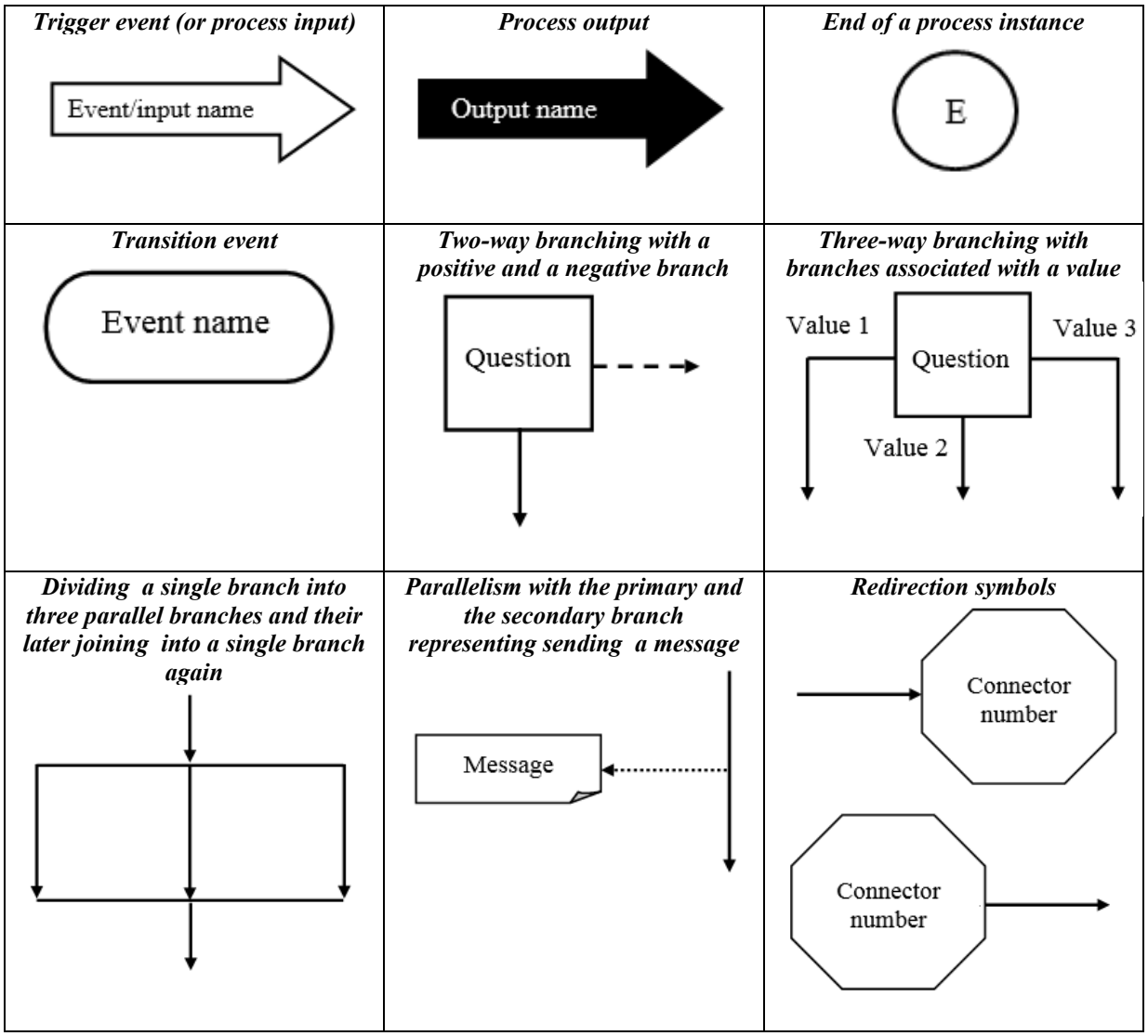
As mentioned above there may be situations, in which the flow of the process instance is conditioned by the fulfilment or non-fulfilment of a certain condition. The branching of the process flow can be depicted by a square symbol, in which the wording of the condition is written. The condition should be formulated as a question. If the question is formulated so that it can be answered only by a "yes" or "no", then there are two branches leading from this square. A solid arrow (i.e., an arrow drawn with a solid line) represents a positive branch (the answer is "yes") and a dashed arrow represents a negative branch (the answer is "no"). The notation for this situation is depicted on figure 5.

If the question is formulated in such a way that it cannot be answered simply by a "yes" or "no", then both branches may be drawn as a solid arrow and every one of them should be marked with a specific value representing one of the possible answers to the question. There may be also situations, in which we need more than two branches. This is not a problem since we may add as many branches as we need. The notation for this situation is depicted in Fig.5.

When parallel execution of the sub-processes is needed, it can be depicted using a simple **fork-and-join mechanism** as shown in figure 5.

We also distinguish a specific type of parallelism, where a secondary branch is separated from a primary branch, and these two branches no longer need to be merged. The secondary branch is just a digression from the primary branch, and it is usually associated with carrying out some trivial activity, which is not important for the further course of the process, such as sending a message for informative purposes only. This type of parallelism can be depicted using the notation on figure 5.

Fig. 5: Other symbols of the Juríк-Schmidt's diagram technique



A very important rule for the Juríк-Schmidt's diagram technique is that every line between two symbols should be directed. However, in large diagrams there may be a need to move from one point to another without drawing a direct line between these two points because the line would be too long and could cross with other lines in the diagram. Crossing lines is undesirable because it makes the diagram less transparent. In such situations we can use a pair of special **redirection symbols**. This pair of symbols represents a kind of teleport, which progresses the course of the process further to a different place in the diagram. Since there may be multiple pairs of redirection symbols in a single diagram, it is necessary to distinguish these symbols by numbers so that those symbols which form a pair have the same number. Two redirection symbols forming a pair can be drawn as shown in Fig. 5.

5. An example of a complex process drawn using the Juríк-Schmidt's method

In this chapter, an application of the Juríк-Schmidt's method should be demonstrated, while *tomato puree production process* might be applied for those purposes.

It is a complex process that consists of two hierarchical levels of sub-processes. With respect to the above-mentioned method and *tomato puree production process* as an example, Fig.6 might be postulated and visualized.

If there are not enough tomatoes the tomato puree production should be postponed and the current instance of the tomato puree production process is being terminated shortly after the beginning. If there are enough tomatoes the process can continue. The owner of the whole process is the Puree production department, and the expected execution time is 470 minutes.
At the first hierarchical level there are three sub-processes:

- 1 Puree cooking
- 2 Puree preserving
- 3 Packaging in paper boxes

The puree cooking process consists of a chain of the following sub-processes on the second hierarchical level:

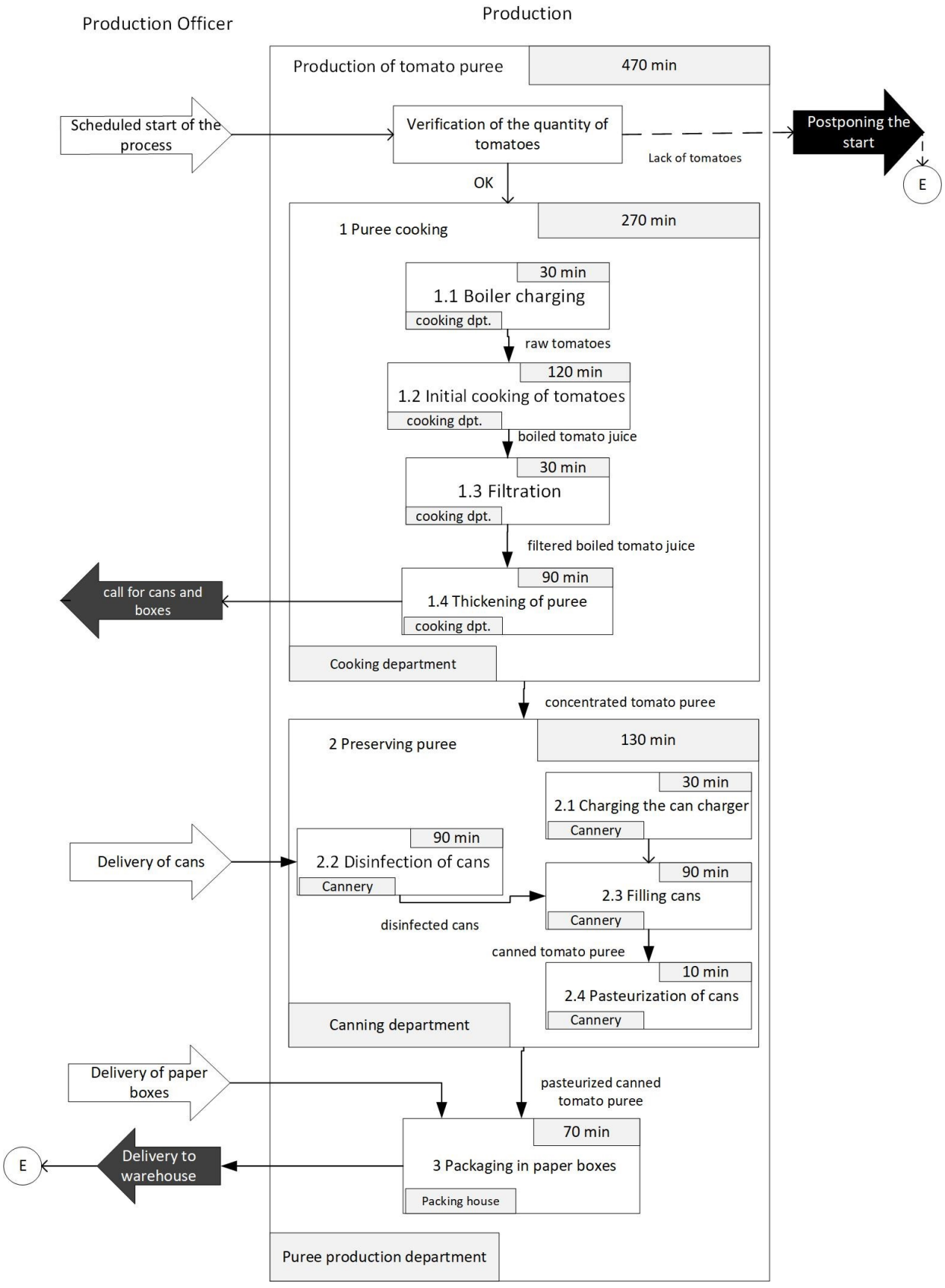
- 1.1 Boiler charging
- 1.2 Initial cooking of tomatoes
- 1.3 Filtration
- 1.4 Thickening of puree

The Puree preserving process consists of a chain of the following sub-processes on the second hierarchical level:

- 2.1 Charging the can charger
- 2.2 Disinfection of cans
- 2.3 Filling cans
- 2.4 Pasteurization of cans

Each sub-process has its owner and its expected execution time, might be shown in Fig 6. However, the matryoshka doll principle is shown at the same figure as well.

Fig. 6: An example of a tomato puree production process



6. Conclusion

In that contribution, there is introduced the Jurík-Schmidt's method for visualization the horizontal flow and also the vertical structure of complex business processes using the Jurík-Schmidt's diagram technique, which is also called the Composition diagram of a complex process. Traditional techniques such as flowcharts, BPMN diagrams, activity diagrams, etc. have their limitations regarding complex business processes with a complicated vertical structure and are not appropriate for their visualization, while

our method seems to be very simple and practical and might be applied to design easy-to-write and easy-to-read business process diagrams, which contain the most important information in a comprehensible form.

7. Resources

Davenport, T. H., & Short, J. (1990). The New Industrial Engineering. Information Technology and Business Process Redesign. Sloan Management Review, 31(4), 11 – 27. Retrieved from http://dspace.mit.edu/bitstream/handle/1721.1/48613/newindustrialeng00dave.pdf.

Jurík, P. (2018). Informačné systémy v podnikovej praxi (2nd ed.). Nové Zámky, Slovak Republic: Tlačiareň MERKUR, s. r. o.

Jurík, P. (2020). Servisne orientovaná architektúra v procesne riadenom podniku (1st ed.). Nové Zámky, Slovak Republic: Tlačiareň MERKUR, s. r. o.

Jurík, P. & Schmidt, P. (2020). Composition Diagram of a Complex Process: A Contribution to Business Process Modelling. TIEES 2020: Trends and Innovations in E-Business, Education and Security: Eighth International Scientific Web-Conference of Scientists and PhD. Students or Candidates., 8, 81-92.

Kanisová Hana, & Müller Miroslav. (2006). *Uml SROZUMITELNĚ*. Computer Press.

Révészová, L., & Paľová, D. (2009). *Základy modelovania podnikových procesov* (1st ed.). Košice, Slovak Republic: Technical University in Košice.

Robson, M., & Ullah, P. (1998). Praktická príručka podnikového reengineeringu (1st ed.). Praha, Czech Republic: Management Press.

Schedlbauer, M. (2010). *The Art of Business Process Modeling: The business analyst's guide to Process Modeling with Uml & Bpmn*. The Cathris Group.

PRODUCTIVE HOME OFFICE

Mária Szivosová[1]

Abstract
Article is devoted to productive home office. The aim of the article was to select tools for home office, categorize selected tools, describe their possibilities and their use, which are conditioned by new circumstances in society, which was affected by the pandemic of COVID-19. The work from the office has been transferred to the household, so many people currently have to work from home.

Key words
Zoom, Microsoft Teams, Skype for business, Covid-19, products and applications to ensure communication

1. Introduction

The real term home office in the sense of working from home originated in the middle of the last century. During this period, the present appearance of offices was slowly forming in developed countries, which was significantly helped by inventions such as the telephone, the Internet, the fax and the like. Working from home occurred especially in the administrative, accounting and development sectors.

The turning point came at the beginning of 2020, when the vast majority of countries were affected by the new COVID-19 virus. It has been shown to have a very high reproductive number and to behave more aggressively than other respiratory diseases, so it has been labeled as dangerous by experts around the world. Many employers therefore decided as much as possible to enable their employees to work from home in order to protect their health and employability

2. Comparison of individual products primarily intended to ensure communication

In this section, we will focus on applications providing business communication, summarize their functions, advantages and disadvantages, and estimate the price. In the price analysis, we will consider small to SME with up to 100 employees (Reynolds, 2020).

The first application is **Microsoft Teams**. Its biggest competitive advantage can be considered that it is part of the Microsoft Office 365 package. It provides full integration with other applications and services of the MS Outlook, MS Planner, Word, PowerPoint and other packages. It also matters whether the company already uses MS Office 365 (Serdar, 2020). If so, Microsoft does not charge any additional fees for its use. Another advantage is that each channel in the Teams platform has its own backup, which in practice means that even after deleting the channel, it is possible to access the platoon history and associated files. The application also allows searching in channels using the search window. MS Teams also has

[1] University of Economics in Bratislava, Faculty of Economic Informatics, Department of Applied Informatics, Dolnozemská cesta 1, 852 35 Bratislava, Slovakia, e-mail: maria.szivosova@euba.sk

many positives in the field of security, such as the creation of private and public teams and channels, the authorization management center, control of administrators over guest access and the like. The advantages can be summarized in the following points:

- is part of the MS Office 365 package;
- guarantees compatibility and integration with other applications in the package;
- does not represent any acquisition costs for companies using MS Office 365;
- backup and search in channels;
- possibility to create private and public teams, administration through a clear admin center;
- up to 20,000 participants in one call.

However, Microsoft Teams also has several disadvantages that are worth considering. One of the biggest problems is its size. By default, any user can create teams on this platform, which can lead to the creation of unnecessary teams and congestion of corporate storage. This setting can also be changed, but it is not easy and requires knowledge of power shell scripts. Another disadvantage is the limited number of channels in each team. There are 200 public and 30 private channels per team. This may be sufficient for SME enterprises, but for larger ones there may be significant restrictions. The integration with other Microsoft tools can be a surprising disadvantage. There are too many of them and many of them have redundant functionality, which can confuse the user to use. Teams is developed by Microsoft, which guarantees full compatibility with the Windows operating system, but others, such as Linux and Mac OS, may have problems. The disadvantages can be summarized in the following points:

- default setting, each user can create by, which leads to frequent disk congestion;
- limited number of channels per team;
- integration with overly redundant tools;
- occasional problematic compatibility with operating systems other than Windows.

Another application is the **Zoom platform**. The biggest advantages include very good handling and a high level of security of online calls with a large number of participants, which will find use in the organization of various webinars and online conferences. The advantages include the multiscreen share function. With it, the user can share more than one of their screens, which can be an advantage especially when presenting a product online to potential customers. Zoom also offers the possibility of broadcasting an online conference or webinar via Facebook, which is a great advantage when organizing advertising campaigns. The advantage is also its free version. Although it supports a maximum talk time of 40 minutes, it can be sufficient for smaller companies that do not have high requirements. The application is also very user-friendly, so its use can be implemented by companies without IT-focused employees. The benefits can be summarized in the following points:

- handling and securing a large number of participants in one call up to 10,000,
- multiscreen share function, sharing multiple screens,
- posting online conferences on Facebook,
- free version
- ease of use,

• many interactive functionalities.

However, Zoom also has several disadvantages. One of the biggest is its safety itself. Although calls are encrypted from the outside and records are stored on encrypted servers, there are a few known cases in the past where hackers hacked directly into an online conference and stole participants' data and personal information. The so-called Zoombombing is also a specific case. This is a relatively new concept associated with the pandemic period, when unwanted guests who disrupt the course of the conference by playing loud music or sharing inappropriate content try to join the online call. Another disadvantage may be the deteriorated video quality (720p by default), which is mainly caused by the excessive utilization of this platform due to the current situation. For some companies, the higher price can also be a decisive factor, which can climb up to several thousand euros when organizing large webinars. The disadvantages can be summarized in the following points:

• lower security,

• Zoombooming, disturbing the call by unwanted guests,

• poorer image quality,

• higher price for premium packages.

Meet Google, like Microsoft Teams, belongs to a larger suite of applications G Suite and within it offers full compatibility. This makes it very easy to use and organize meetings for companies that use this platform. The administrator may allow or deny participants to upload the meeting, which will be saved to Google Drive. Another big advantage is the better free version compared to others - the call can last up to an hour and have up to 100 participants. The price of more premium packages, which is lower than with Zoom or Microsoft Teams, can also weigh in the decision-making process. The benefits can be summarized in the following points:

• strong free version;

• is part of the G Suite (Google Workspace);

• guarantees compatibility and integration with other applications in the package;

• lower price for premium packages.

The biggest disadvantage of Google Meet is its limited functionality compared to others. The Enterprise package offers a maximum of 250 participants in a call, which can be a significant problem for larger companies. Google Meet also doesn't support sharing multiple screens at once. Higher hardware requirements, especially for RAM and processor, can also be a disadvantage. The disadvantages can be summarized in the following points:

• limited functions compared to other mentioned applications,

• limitation of a maximum of 250 participants even with the most expensive version,

• higher hardware requirements.

Slack is another of the communication applications worth mentioning. Its biggest benefit is a very simple implementation and a user-friendly environment. We can also consider its ability to integrate with a wide portfolio of applications, such as Google Drive, MailChimp, GitHub and many others, as a great advantage. Slack can also be a very flexible tool, which makes it suitable for larger companies. It is mainly a chat platform, but it also allows you to make calls. The benefits can be summarized in the following points:

- very simple implementation and friendly user environment,
- integration with a large number of third-party applications,
- flexible enumeration tool for every size of firms.

This application also has a few disadvantages. One of them is a small storage space, so sending files via Slack is not highly recommended. Several users described Slack as less transparent, especially for larger companies. Some experts also consider Slack to be addictive, as it is a chat platform similar to Messenger or WhatsApp, with the risk of employees losing too much time communicating, which can lead to procrastination. The disadvantages can be summarized in the following points:

- small storage space, unsuitable for forwarding files;
- with a large number of channels, we can easily lose track;
- can be a big "eater" of time.

Based on a search of several sources and an examination of the product documentation of individual providers, it is not possible to unambiguously determine the best communication tool. It depends mainly on the preferences and requirements for its functionality on the part of users.

- Google Meet is a reliable and affordable option for less demanding users. It's simple and just needs a Google Account to use it. It is also suitable if the company works with the G Suite. For larger companies or organizations where users need to present projects or organize training, it may be insufficient and it is therefore appropriate to consider other alternatives.
- Microsoft Teams is a powerful communication tool for businesses of all sizes and has many features to help you better organize, such as creating teams, channels, and more. It's part of Office 365, which is used by many organizations today. They can use the full potential of Teams without the need for investment. It works great especially within the domain of the organization. For smaller businesses that don't have many organizational units and don't use Office 365, it can be unnecessary (Strutt, 2020).
- Zoom has the most tools that can be used to present projects or learning in schools. It is described as a great tool for organizing webinars. It can also be used in primary and secondary schools, where students do not have Office 365 licenses available. Its security can be problematic, so companies working with sensitive data should consider using it.
- Slack is primarily a chat platform, but thanks to integrations with various programs, it can be a powerful support tool in the right hands. It is used mainly in companies focused on application development. It offers very little for telephone and online presentations and is therefore more suitable as a supplement to the above-mentioned services.

• Zoiper5 is not a platform for organizing meetings. It is software that replaces physical VoIP phones and thus allows users to communicate via these protocols even in places where such a phone is not available.

Tab.1 :Comparison of some aspects of the above mentioned programs

Application	Microsoft Teams	Zoom	Google Meet	Slack
Name of plan	Microsoft 365 Business Standard	Business	Google Workspace Essentials	Standard
Maximum number of users in one call	300	300	150	15
Call recording	yes	yes	yes	no
Maximum call duration	24 hours	30 hours	24 hours	not specified
Desktop application	yes	yes	no	no
Live closed captions	yes	yes	yes	no
Surveys, polls	yes	yes	yes	nie
Web application	yes	no	yes	yes
Sharing multiple screens	no	yes	no	no
API integration	yes	yes	no	yes
Price	10,39 € user / month	18,99 € month/licencia	6,65 € user / month	5,55 € user / month

Source: author

3. Comparison of individual products primarily designed to provide shared storage

First, we focus on the benefits of the **Google Drive**, which was identified by respondents as the most popular. Its biggest advantage is a great free version. It offers users up to 15 GB of storage space, which may be sufficient for less demanding users. The service is part of the G Suite, which guarantees compatibility with other services, such as Google Meet, Docs, Sheets and the like. Google Drive also supports integration with many third-party applications. He also enjoys popularity thanks to its user-friendly interface. We can also consider the more affordable price for premium packages as an advantage. The benefits can be summarized in the following points:

• up to 15 GB of space with the free version;

• is part of the G Suite, compatibility with other services from this package;

• simple and user-friendly interface;

• lower price for paid plans than for competitors;

- support for third party applications;
- compatibility with Microsoft Office 365 applications.

There are a few disadvantages associated with Google Drive. The biggest is insufficient security, especially when sharing files. If files are shared using a link, anyone with that link can access those files. Another significant disadvantage is the daily limit when uploading files. After exceeding the 750 GB limit, the application will no longer allow you to upload more files that day. There is also a size limitation when converting some documents to Google formats, eg: Google Docs document up to 50 MB, Google Slides presentation up to 100 MB. Disadvantages include the fact that some third-party applications may pose a potential risk. The disadvantages can be summarized in the following points:

- password protection is missing for shared files;
- daily upload limit 750 GB;
- limit the size of some file formats;
- Some third-party applications may pose a threat.

The second most significant alternative in the survey was the **OneDrive** cloud storage service. Its popularity stems from several factors. Above all, it can be pre-installed with Windows 10 pre-installed. It has a simple and user-friendly web interface in which everything is neatly organized. It is ideal for backing up folders and files in Windows 10. Files added to folders marked for synchronization are automatically mirrored to OneDrive without user intervention. Another positive function is the so-called Personal Currency. Files stored in it require two-step authentication to open. Also, they can't be shared with a few clicks, which can avoid the inconvenience associated with accidentally clicking on a share. Personal Currency will automatically lock after a few minutes without recording user activity. As a positive, we can also mark a two-level recycle bin, which prevents deleting files by mistake. Even after emptying the first level recycle bin, the files are still left in the second level. The benefits can be summarized in the following points:

- is preinstalled with Windows 10;
- clear web interface;
- the ideal tool for synchronizing and backing up files in Windows 10;
- Personal Currency service;
- two-level basket.

OneDrive also has several disadvantages, the biggest of which are the company's policy itself. It reserves the right to scan files stored on OneDrive, so we can talk about invasion of privacy. The service lacks several useful tools for managing files within an organization, such as Microsoft SharePoint. Another disadvantage is the unavailability of a desktop application for Linux operating systems. For more demanding users, limiting the maximum size of the uploaded file can also be a problem. During the first pandemic, Microsoft increased this limit from 15 GB to 100 GB, which is a significant shift, but it still represents a certain limitation. The disadvantages can be summarized in the following points:

- Microsoft's privacy concerns
- missing file management tools,

- absence of client application for OS Linux,
- limited maximum size of the uploaded file.

The third most popular cloud storage service among respondents was **Dropbox**. We can consider the advantage of a sophisticated file sharing system. It provides features such as expiration dates, password protection, and removal as soon as a link is clicked. Dropbox is the only one mentioned to offer a desktop application for the Linux OS. It also offers integration with many applications as well as a digital signature function. The advantage over the competition is also the non-existent limitation of the size of uploaded files through an application available for all known operating systems. In surveys, users also often highlighted the sync speed, which was faster than competing services. The benefits can be summarized in the following points:

- sophisticated file sharing tools;
- client application also available for Linux OS;
- integration with many applications and digital signature function;
- the size of uploaded files is not limited;
- faster synchronization than the competition.

Dropbox also has several disadvantages, the biggest of which is the price, which is significantly higher than the competition, which is often a decisive factor. Another disadvantage is the bad reputation in the field of security, as several cases of large amounts of data leakage have been known in the past. The free version provides only 2 GB of storage space, which may discourage some users. The disadvantages can be summarized in the following points:

- higher price than competing programs,
- bad reputation in the field of security,
- weak free version.

The last mentioned application is **Microsoft SharePoint**. Unlike OneDrive, it is not a preinstalled component of Windows 10, but a service from Microsoft Office 365. This factor ensures that it works best with other services in the suite. Compared to its sibling, it has several functionalities that can be used especially in the corporate sphere. It provides its users with several customization options, so that each user can customize it to their needs. It supports synchronization with Azure Active Directory, which helps administrators manage it. It also provides the highest level of security of the mentioned programs. The benefits can be summarized in the following points:

- part of Office 365, which guarantees full compatibility with other services,
- Multiple business-specific features,
- the possibility of a significant degree of adaptation,
- Sync with Azure AD,
- the highest level of security among the other instruments mentioned.

SharePoint may seem like the best solution for businesses, but it also has several disadvantages that we must take into account when making decisions. Its greatest complexity can be described not only for ordinary users, but also for the more technically proficient.

Therefore, the effective use of SharePoint will most likely require training for employees and system administrators to avoid common problems in its use. This is associated with the slow adaptation of the service to new realities and needs of the company. Searching for files in SharePoint can also be problematic. The disadvantages can be summarized in the following points:

- complex use and initial setup;
- often requires support and training from external companies;
- slow adaptation to new business needs;
- problematic search.

The following table summarizes and compares the individual plans of the above services. As with communication applications, we will consider the needs of SME businesses with up to 100 employees.

Tab.2 : Price and comparison of functionalities of individual services for packages intended for SME

Name of services	**Disk Google**	**OneDrive**	**Dropbox**	**SharePoint**
Name of plan	**Business Standard**	**OneDrive for Business**	**Business Standard**	**Microsoft 365 Business Standard**
The size of the storage space	2 TB per user	1 TB per user	5 TB per user	1 TB per user
File sharing via link	yes	yes	yes	yes
Desktop application for Windows and Mac OS	yes	yes	yes	yes
Desktop application for Linux	no	no	yes	no
Multilevel authentication	yes, it can be set	Yes but only in the Personal Valut	yes, it can be set	individually adjustable for individual folders
Maximum file upload size	5 TB	100 GB From 14. 1. 2021 250 GB	not specified	100 GB
Daily upload limit	750 GB	not specified	not specified	not specified
Price	10,33€ user / month as part of G Suite/9,99 € as a separate storage space	4,20 € user / month / 10,50 as part of Office 365	12 € user / month	10.50 € user / month as part of Office 365

Source: author

Based on the available information, we cannot unambiguously determine the best application, even in this case it depends on the preferences and needs of users.

- Google Drive is a good alternative for light users. With its powerful free version and ease of use, it can be a sufficient alternative for small businesses and individuals to share files outside the enterprise. It can become confusing for larger companies. Better alternatives are available for multi-user collaboration within a single repository.
- OneDrive, whether a standalone version or an Office 365 service, is a great tool for Windows 10 users. It offers a clearer web interface compared to Google Drive. A great advantage is also the Personal Currency service or a two-level basket. Due to its limitations, it is not very suitable for cloud backup of large amounts of data. Due to the absence of a desktop application on the Linux OS, it is unusable for its users.
- SharePoint is the right cloud storage for working with files inside and outside the organization. Thanks to its advanced features, it can be sufficient even for demanding users. However, its setup and proper use can be challenging and in many cases requires training, which can be time and money consuming, so it is not suitable for businesses and organizations that do not reach its full potential.
- Drop Box offers the most file sharing options, especially outside the organization. Based on a search of multiple sources, it has the best synchronization of folders and files among the mentioned. It also offers an application for the Linux OS, which is a big plus in companies whose employees also use this operating system. His safety is debatable. Past incidents can discourage people.

We cannot guarantee the security of data against misuse with any cloud storage, so it is advisable to use internal corporate storage when working with sensitive data, and to provide employees with a secure connection to the corporate network even outside the company's premises.

4. Conclusion

In this article, we dealt with the issue of working from home caused by unexpected circumstances. We analyzed communication tools capable of ensuring interpersonal contact even when working remotely. We introduced document-sharing programs as well as tools that employers can use to track the work of their employees. We compared the individual applications and summarized their advantages and disadvantages and a price overview. The portfolio of offered applications is very wide and each offered has its own unique specification, therefore the basis for a good selection of the application package is a good definition of the requirements and goals of the company. It is therefore not possible to unambiguously determine the universally best application.

5. Resources

Reynolds, B. W (2020). Brief History of Microsoft doi.: https://content.dsp.co.uk/a-brief-history-of-microsoft-the-worlds-biggest-software-compan

Serdar S. (2020). Microsoft 365, doi.*: https://docs.microsoft.com/en-us/microsoftteams/teams-overview*

Strutt, A.(2020). The Home Office: An Introduction to its Early History doi.:*https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9299.1961.tb01086.x*

SECURITY ANALYSIS OF CRYPTOGRAPHIC HASH FUNCTIONS USING HARDWARE-ACCELERATED CLOUD COMPUTING

Patrik R. Ujváry[1]

Abstract
While passwords are widely used in authentication, only limited attention is paid to the security of credential storage. Cryptographic hash functions enable storage, retrieval and verification of passwords without explicitly storing the password itself in plaintext. In this paper the security of various password hashing algorithms is compared with the use of high-performance hardware-accelerated AWS EC2 cloud instances. Several hardware and GPU configurations from different process nodes were analysed to demonstrate the effect of generational improvements. This study highlights a common but potentially insecure way of storing hashed passwords and suggests a new hash function specifically designed for secure hashed credential storage.

Key words
cryptographic hash functions, GPU computing, data security, password cost

1. Introduction

User authentication is a common task that many of the services rely on, therefore its security is key. While password authentication may not be the safest solution, it is commonly used due to its easy implementation. The way of credential storage should enable the service provider to identify the user by checking the password entered. This is called verification and is accomplished by a simple comparison. The stored data does not need to be the password itself, it can be any characteristic of the password that is reproducible and collision resistant. The data must be stored in a way that even in the event of a data breach the information could not be misused. Even the biggest companies experienced data breaches when their password databases got released (Kamp, 2012; Thomas et al., 2017). Such intentional or unintentional information release cannot always be prevented.

A possible solution is to store only a cryptographic hash of the password, which improves security by a huge margin. Every time the password needs to be verified, it is hashed again and the hash is compared to the stored value. There must not be an inverse hash function, thus the hash does not reveal the password.

A key feature of cryptographic hash functions is that the password's hash is always the same: it is the same at the time of hashing or when setting a new password or during verification. The probability of a hash collision – i.e. when different passwords result in the same hash – is extremely low. While using a cryptographic hash the leaked password database contains only hashes of the passwords, this can function as the last barrier of protection before an attacker

[1] Óbudai University Doctoral School on Safety and Security Sciences, ujvary.patrik@bgk.uni-obuda.hu

could use the data. The security of the cryptographic hash functions comes into play at this point.

In this paper I will examine a popular but – as shown later – mostly unsecure way of a single-password authentication. Other security measures such as 2FA, server-side detection of suspicions activity, analysing the users' ISP, country, user agent, screen resolution, time of day, usage patterns, etc. are not covered. The present paper suggests a more secure credential storage using a new hash function.

A major factor of security is the rate you can brute-force hashed passwords at, hereinafter referred to as hashrate. First, I examined 5 generations of general-purpose graphics processing units (GPGPUs) from the NVIDIA Tesla lineup. These are specially designed for data center usage (see section 1.3). I benchmarked the systems and compared the results to present the improvements in hashrate and efficiency resulting from the newer manufacturing processes. The security of various password hashing algorithms is compared using AWS EC2 cloud instances. Cost efficiency is also evaluated. The latest-generation Ampere microarchitecture – launched September 2020 – is also included.

1.1 AWS EC2

Five different instance types[2] were used for the benchmarks in the Amazon Elastic Compute Cloud (AWS EC2), each with different NVIDIA Tesla[3] GPUs (Tab. 2). All data presented is measured using the CUDA API.

1.2 Instances

The AWS instances used to hash passwords (Tab. 1) have various specifications, up to 8 GPUs, 96 vCPUs and more than a TiB of memory.

Tab. 1: AWS EC2 instance types

Instance type	Type of GPU	Number of GPUs	vCPUs	Memory [GiB]	GPU Memory [GiB]	$/h
p2.8xlarge	K80	8	32	488	96	7.20
g3s.xlarge	M60	1	4	30.5	8	0.75
p3.8xlarge	V100	4	32	244	64	12.24
g4dn.xlarge	T4	1	4	16	16	0.53
p4d.24xlarge	A100	8	96	1152	320	32.77

1.3 NVIDIA Tesla DC GPU generations

I tested 5 different instances with 5 different GPUs (Tab. 2) from 2014 to the latest of 2020.

[2] AWS refers to VMs as instances.

[3] Tesla refers both to the DC (Datacenter) accelerator cards and to an old microarchitecture from 2007-2008.

Tab. 2: Properties of NVIDIA Tesla DC GPUs

GPU	Microarchitecture	Launch year	CUDA cores	f_{base} [MHz]	f_{boost} [MHz]	f_{avg} [MHz]	TDP [W]	BW [GB/s]
K80	Kepler	2014	4992	560	875	718	300	480
M60	Maxwell	2015	4096	899	1178	1039	263	320
V100	Volta	2017	5120	1245	1370	1308	250	900
T4	Turing	2018	2560	585	1590	1088	70	320
A100	Ampere	2020	6912	765	1410	1088	250	1555

TDP (thermal design power) is the heat generated by the card under full load or the amount of power required to achieve the performance. In order to estimate the real-world frequency f_{avg} was calculated as shown by equation (1.1).

$$f_{avg} = \frac{f_{base} + f_{boost}}{2} \tag{1.1}$$

1.4 Software environment

All measurements were made using exactly the same operating system, software, drivers, with the same VM disk image called AMI. Ubuntu Linux 18.04 was chosen as the OS, with the AWS Deep Learning AMI being the base disk image. Only open source software was used except for the NVIDIA GPU Driver (version 450.80.02) and the CUDA Toolkit (version 11.0), which are proprietary.
The benchmark was performed with hashcat (version 6.1.1):

hashcat-6.1.1/hashcat.bin -b > bench-$(curl -s http://169.254.169.254/latest/meta-data/instance-type).txt

The data can be extracted easily with the following multiline PCRE regular expression using pcregrep (version 8.39):

'Hashmode: (\d+) - .*\n\nSpeed\.#\d\.*: +(\d+\.?\d*) .?H\/s.*\nSpeed\.#\d\.*: +(\d+\.?\d*) .?H\/s.*\nSpeed\.#\d\.*: +(\d+\.?\d*) .?H\/s.*\nSpeed\.#\d\.*: +(\d+\.?\d*) .?H\/s.*\nSpeed\.#\d\.*: +(\d+\.?\d*) .?H\/s.*\nSpeed\.#\d\.*: +(\d+\.?\d*) .?H\/s.*\nSpeed\.#\d\.*: +(\d+\.?\d*) .?H\/s.*\nSpeed\.#\d\.*: +(\d+\.?\d*) (.?H)\/s.*\nSpeed\.#*\.*: +(\d+\.?\d*) (.?H)\/s.*'

2. Benchmark data

In order to easily compare results from different GPUs and instances I created an index variable denoted by h_{MSSS} that aims to represent the average performance of the devices and machines. Two types of h_{MSSS} were calculated, one for the GPU and one for the machine including the sum of the performance of the GPUs in the system. With machines containing only one GPU, the two values are the same. This mean hashrate h_{MSSS} is calculated by forming the arithmetic mean of 4 common conventional hash functions, which were the MD5, the SHA1, the SHA2-256 and the SHA2-512 algorithm, shown by equation (2.1).

$$h_{MSSS} = \frac{1}{M}\sum_{j=1}^{M}\frac{1}{N}\sum_{i=1}^{N} h_{ij} = \frac{1}{MN}\sum_{j=1}^{M}\sum_{i=1}^{N} h_{ij} \quad \textbf{(2.1)}$$

2.1 GPU performance

Tab. 3 summarizes the performance measured by the individual GPUs.

Tab. 3: Hashrate h *for various hashes in MH/s*

GPU	MD5	SHA1	SHA2-256	SHA2-512	h_{MSSS}
K80	4 353.8	1 865.8	814.3	221.4	1 813.8
M60	12 107.0	4 156.9	1 478.5	513.2	4 563.9
V100	56 107.8	17 730.6	7 692.3	2 402.4	20 983.3
T4	21 895.6	7 629.5	3 265.6	1 040.2	8 457.7
A100	65 858.8	22 053.0	9 555.3	3 126.9	25 148.5

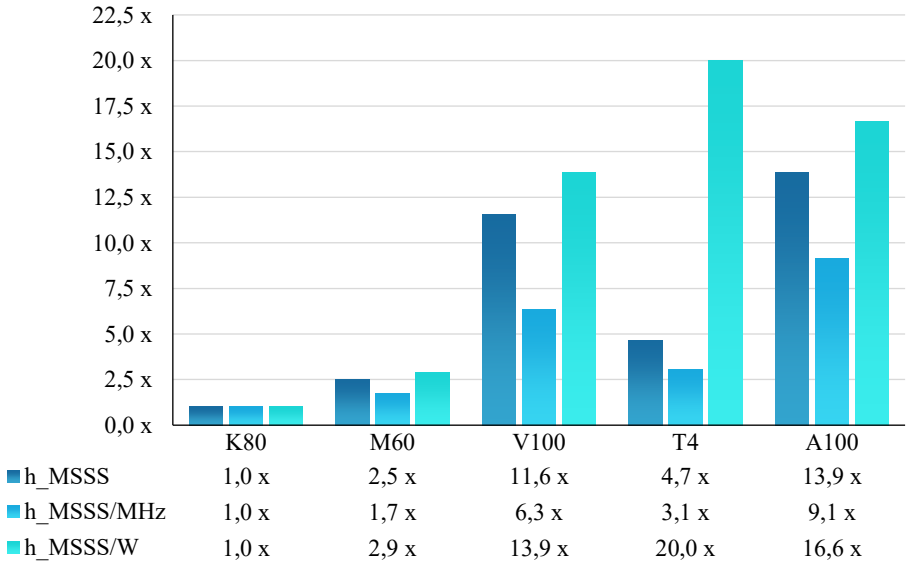
The raw data in Tab. 3 is presented in Fig. 21, where h_{MSSS} shows the relative hashrate of a single GPU, and $h_{MSSS/MHz}$ and $h_{MSSS/W}$ show the relative hashrate per MHz and per Watt respectively compared to the K80 GPU.

$$h_{MSSS/MHz} = \frac{h_{MSSS}}{f_{avg}} \quad \textbf{(2.2)}$$

$$h_{MSSS/W} = \frac{h_{MSSS}}{P_{TDP}} \quad \textbf{(2.3)}$$

Relative values are calculated as shown by equation (2.2) and (2.3) and are intended to emphasize the increase in efficiency.

Fig. 21: Relative hashrate and efficiency compared to the K80 GPU



Despite the fact that the T4 card is newer than the V100, it fell short of the older version's performance. The reason is the significant difference in TDP: while the V100 is a 250W card, the T4's power usage is only 70W, so it has a way higher power efficiency measured in hashes/Watts. Interestingly neither the most recent nor the fastest card is the most power-efficient. It is notable that the raw hashrate has increased almost 14-fold in a short period of just 6 years. Salting might slow down guessing a hashed password, but only in some cases.

2.2 Aggregated GPU performance

In Tab. 4 the aggregated GPU performance is shown, which is the sum of all cards' hashrates (equation (2.4)). The g3s.xlarge and the g4dn.xlarge include only one GPU per instance, therefore their h_{MSSS} in Tab. 4 equals their h_{MSSS} in Tab. 3.

$$h_{MSSS} = \sum_{j=1}^{M} \frac{1}{N} \sum_{i=1}^{N} h_{ij} = \frac{1}{N} \sum_{j=1}^{M} \sum_{i=1}^{N} h_{ij} \tag{2.4}$$

Tab. 4: Aggregated GPU performance in MH/s

instance	GPUs	MD5	SHA1	SHA2-256	SHA2-512	h_{MSSS}
p2.8xlarge	8×K80	34 830.0	14 926.1	6 514.1	1 771.4	14 510.4
g3s.xlarge	1×M60	12 107.0	4 156.9	1 478.5	513.2	4 563.9
p3.8xlarge	4×V100	224 400.0	70 922.6	30 769.3	9 609.7	83 925.4
g4dn.xlarge	1×T4	21 895.6	7 629.5	3 265.6	1 040.2	8 457.7
p4d.24xlarge	8×A100	526 900.0	176 400.0	76 442.8	25 015.4	201 189.6

Regarding costs, apart from the pay-as-you-go price all other costs are covered by the cloud service provider.

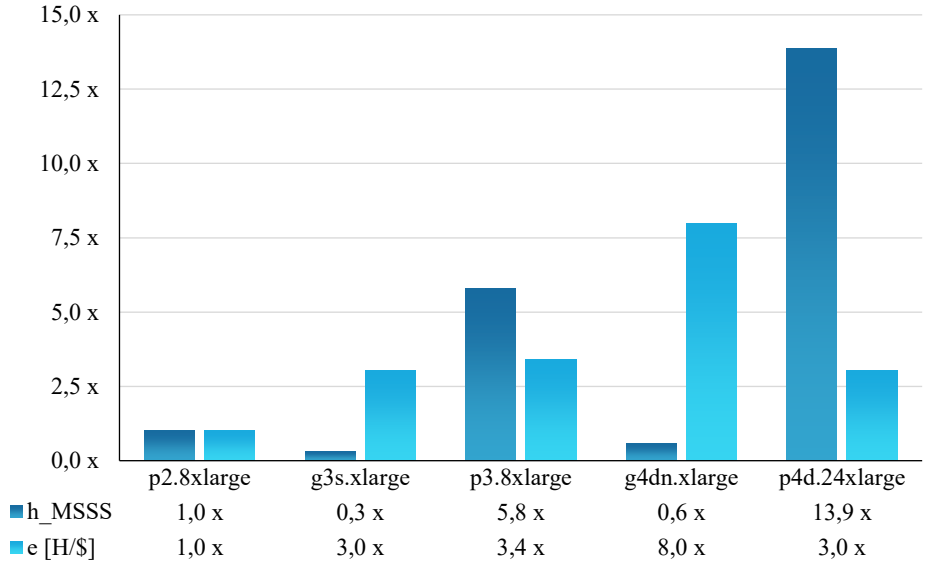
This type of password guessing uses only the GPUs and leaves other system resources virtually unutilized. I assumed that the rented EC2 instance is not being used for any other purposes. Finding a secondary use for the (in some cases) huge amount of unused resources could decrease costs. In case of the p4d.24xlarge instance 8TiB of fast NVMe SSD space, 1152GiB of RAM and 96 vCPUs are unused.

There are no upfront costs of purchasing expensive equipment, no hardware maintenance costs, no electricity costs, and no costs related to having a DC you can deploy hardware to.

Price efficiency e represents the number of hashes that can be computed per one dollar (Fig. 22) and should be evaluated on the instance level, since the hourly price applies to a full instance with possibly multiple GPUs – see equation (2.5).

$$e\left[\frac{\mathrm{H}}{\$}\right] = \frac{h\left[\frac{\mathrm{H}}{\mathrm{s}}\right]}{p\left[\frac{\$}{\mathrm{h}}\right]} = \frac{h}{p}\frac{[\mathrm{H}\cdot\mathrm{h}]}{[\$\cdot\mathrm{s}]} = 3600\cdot\frac{h}{p}\left[\frac{\mathrm{H}}{\$}\right] \tag{2.5}$$

Fig. 22: Relative hashrate and price efficiency compared to the K80 GPU



2.3 A note on Moore's law

Gordon Moore's prediction was that the number of transistors on an integrated circuit would double every 24 months (Moore, 1965). This leads to an exponential growth in performance over time and it seems that his observation still holds.

2.4 Coefficient of variation

As Tab. 2 illustrates several instances have multiple GPUs, such as the p2.8xlarge, the p3.8xlarge and the p4d.24xlarge. The average relative standard deviation (dimensionless) or coefficient of variation c_v of the individual GPU's hashrates is shown by equation (2.6) and (2.7).

$$\overline{c_v} = \frac{1}{M}\sum_{j=1}^{M}\frac{\sigma_j}{\mu_j} \tag{2.6}$$

$$\sigma_j = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(h_{ij}-\mu_j\right)^2}\ ;\ \mu_j = \frac{1}{N}\sum_{i=1}^{N}h_{ij}. \tag{2.7}$$

where h_{ij} is the hashrate of GPU i for hash j,

μ_j is the mean hashrate for hash type j,

N is the number of GPUs,

M is the number of hashes tested,

σ_j is the standard deviation of hash type j.

Tab. 5: Coefficient of variation

GPU type	Number of GPUs	c_v
K80	8	2.74%
V100	4	0.28%
A100	8	0.17%

c_v in Tab. 5 shows that there is only a small (under 5%) deviation in the individual GPU's hash power among the K80's and it is even smaller among the V100's and A100's.

2.5 OpenCL

Apart from CUDA, OpenCL measurements were also carried out. The difference between CUDA and OpenCL hashrates were calculated by using the formula shown by equation (2.8),

$$r = \left[\sum_{j=1}^{M} \sum_{i=1}^{N} h_{ij,OpenCL} \cdot \left(\sum_{j=1}^{M} \sum_{i=1}^{N} h_{ij,CUDA} \right)^{-1} \right] - 1 \qquad \textbf{(2.8)}$$

where r is the relative difference in hashrate displayed in percentage, and N and M the total number of GPUs and hashes tested respectively.

Tab. 6: OpenCL measurements[4]

GPU	Microarchitecture	year	r
K80	Kepler	2014	-4.31%
M60	Maxwell	2015	-5.05%
V100	Volta	2017	0.49%
T4	Turing	2018	0.47%

Tab. 6: OpenCL measurements explains that the relative difference is an order of magnitude smaller with the newer generation GPUs (V100, T4). Hashing with the use of the OpenCL API was slower with older GPUs and slightly faster with the newer ones. The most significant difference could be measured in bcrypt – 31% slower than CUDA, and in the RAR5 algorithm – 38% faster than CUDA both while using the T4 (Turing) card.

3. Cost of a password

Calculating how long it takes to guess a k character password is not the best metric, as you can always add more instances in parallel as long as there is enough capacity available. A better metric would be to specify how much it would cost – or in other words – how much an individual password is worth, given a certain length and complexity, expressed in USD.

[4] OpenCL data for the A100 is not available.

Currently the least expensive AWS regions were us-east-1 (N. Virginia), us-east-2 (Ohio) and us-west-2 (Oregon), each priced equally.

Tab. 7: Price efficiency

instance	GPUs	AWS cost [$/h]	MD5 [H/s]	h/p [H/$]
p2.8xlarge	8×K80	7.20	34 830 000 000	17 415 000 000 000
g3s.xlarge	1×M60	0.75	12 107 000 000	58 113 600 000 000
p3.8xlarge	4×V100	12.24	224 400 000 000	66 000 000 000 000
g4dn.xlarge	1×T4	0.53	21 895 600 000	149 855 817 490 494
p4d.24xlarge	8×A100	32.77	526 900 000 000	57 878 837 809 634

The g4dn.xlarge with the T4 GPU proved to be the most efficient, it can calculate the MD5 hash of 21 895 600 000 password candidates per second, costing $0.53 per hour, resulting in an efficiency of 149 855 817 490 494 passwords per USD (Tab. 7). The cost of a password was calculated based on the highest H/$ value – i.e. the most efficient instance. A worst-case-scenario is modelled below assuming that only the common MD5 hashing algorithm is used.

A data breach that happened at the RockYou.com website in 2009 resulted in over 32 million leaked passwords in plain-text format. Analysing these passwords revealed interesting information about passwords (or about humans).

Tab. 8: RockYou password patterns (Lowercase, Uppercase, Digit, Special character)

Pattern	length	% of total	% cumul.	possible combinations	cost	cumul. cost
LLLLLL	6	12.2	12.2	309	0.00	0.00
LLLLLLL	7	8.4	20.6	8 032	0.00	0.00
LLLLLLLL	8	7.6	28.2	208 827	0.00	0.00
DDDDDD	6	7.0	35.2	1	0.00	0.00
LLLLLLLLL	9	4.2	39.4	5 429 504	0.04	0.04
LLLLL	5	2.9	42.3	12	0.00	0.04
LLLLLLDD	8	2.8	45.1	30 892	0.00	0.04
LLLLLLLLLL	10	2.7	47.8	141 167 096	0.94	0.98
DDDDDDDD	8	2.5	50.3	100	0.00	0.98
LLLLLDD	7	2.3	52.6	1 188	0.00	0.98
LLLLLLD	7	2.2	54.8	3 089	0.00	0.98
DDDDDDD	7	2.0	56.8	10	0.00	0.98
LLLLDD	6	1.9	58.7	46	0.00	0.98
LLLLLLLD	8	1.7	60.4	80 318	0.00	0.98
LLLLLD	6	1.7	62.1	119	0.00	0.98
DDDDDDDDDD	10	1.7	63.8	10 000	0.00	0.98
LLLLLLLDD	9	1.5	65.3	803 181	0.01	0.99
LLLLLLLLLL	11	1.4	66.7	3 670 344 487	24.49	25.48
DDDDDDDDD	9	1.4	68.1	1 000	0.00	25.48
LLLLLLLLD	9	1.2	69.3	2 088 271	0.01	25.49

This data is more than a decade old, its relevance continues to decrease as it is more and more common by service providers to implement stricter password policies. The power of the

RockYou data lies in the fact that it was released in plain-text format, 100% of the data could be processed to extract the most common password patterns (seen in Tab. 8: RockYou password patterns (Lowercase, Uppercase, Digit, Special character)), without any extrapolation. This might not be true for newer data, but if RockYou leaked MD5-hashed passwords today, you could recover at least 69.3% of the passwords with guessing only alphanumerical characters (only lowercase characters and digits). In fact, 69.3% of the passwords are composed of only 20 patterns (see Tab. 8: RockYou password patterns (Lowercase, Uppercase, Digit, Special character)). Excluding[5] the 18th most common pattern (11 lowercase letters) results in 67.9% for just $1.00 lasting 114.1 minutes. Preferring speed over efficiency this operation would take approximately 285 seconds with a single p4d.24xlarge instance costing $2.59. People tend to bias towards certain common password patterns, which drastically reduces the total number of possible character combinations (Leininger, 2014), resulting in a very poor security when combined with weak cryptography.

4. Suggestions, recommendations

Conventional hash functions like the MD5 or the SHA1 algorithm are not suitable for password hashing purposes. They are fast and this fact makes them weak from a security point of view. These functions are still appropriate for fingerprinting files, identifying duplicate entries in a database, verifying the integrity of data after it was transmitted through the network with a checksum. The MD5 algorithm's (Rivest, 1992) original specification was published in 1992 and SHA1's in 1993 (NIST, 1993). These hash functions should not be used in any cryptographic application anymore.

Eventually, the security of a stored password depends only on two factors: the number of possible combinations and the number of guesses you can make per unit of time (the hashrate). Increasing the first factor results in too complex password policies thus hard-to-remember passwords, so the solution has to come from the second factor: the hashing process must be slowed down.

Testing a password candidate for a conventional hash function requires one hash operation. In case of an adaptive hash function with n iterations the attacker has to perform n operations for each password they want to test, significantly increasing the time of an attack. There are so-called password-based KDFs purposefully designed to be computationally intensive, in order to increase the time spent hashing. Most KDFs are adaptive, they can be tuned: the iteration count can be increased to make the computation intentionally slower, so that the hash function remains resistant to brute-force attacks even with increasing computational power. Adaptive hash functions' performance depends on the iteration count n (and other tunable variables), they can be adjusted to the user's needs, therefore no benchmark data is shown for these types of hash functions. Such hash functions are: PBKDF2, BCRYPT, SCRYPT or Argon2.

I personally recommend using Argon2 (Biryukov et al., 2015) for password storage: it has all of the important features a password hashing algorithm should provide. For instance, Argon2 is designed to be GPU/FPGA/ASIC-unfriendly, memory-hard, etc., it is included in many software libraries, it has wide software support. Furthermore, Argon2 is the winner of the Password Hashing Competition organized in July 2015.

[5] The 18th most common pattern covers only 1.3% of the total number of passwords but costs $55.61 to exhaust its key space.

Using an appropriate hash function is key to good security.

5. Conclusion

According to Moore's law computational performance increases exponentially over time – so does the speed of guessing passwords. A single general-purpose GPU today can try approx. 65 billion password candidates per second. This means that conventional hash functions do not provide enough protection against brute-force attacks. Although brute-force is now viable option, applying more advanced statistical processing and mathematical analysis provides even better, more efficient ways of guessing passwords. The combination of constantly growing computational performance, mathematical statistics and data analysis, smarter tools and special hardware indicate a worrying trend for data security.

The cloud provides a high degree of scalability, flexibility, quick deployment, and pay-as-you-go billing. Leveraging the power of parallelization and automation brings high efficiency without upfront investments.

People tend to create certain common password patterns, almost 70% of the RockYou database consists of only 20 of these patterns. Even stricter password policies will probably lead to hard-to-remember passwords and shift the new passwords to other common patterns. Therefore, alongside of stronger passwords a more secure way of storing credentials is needed.

New hash functions such as Argon2 are purposefully designed to be suitable for password storage. To ensure the users' security software developers and service providers should implement advanced password protection methods that meet today's standards.

6. Symbols and abbreviations

Symbol	Description
2FA	Two-factor authentication
AMI	Amazon Machine Image
API	Application Programming Interface
AWS	Amazon Web Services
BW	Memory Bandwidth
$\overline{c_v}$	Coefficient of variation
CUDA	Compute Unified Device Architecture
DC	Data center
e	Price efficiency h/p
EC2	Elastic Compute Cloud
f_{avg}	Average clock
f_{base}	Base clock
f_{boost}	Max boost clock
FPGA	Field-programmable gate array
GPGPU	General-purpose computing GPUs
GPU	Graphics processing unit
H	Total number of hashes
h	Hour

Symbol	Description
h_{MSSS}	Average hashrate for MD5, SHA1, SHA2-256 and SHA2-512
i, j	Iterator
IP	Internet Protocol
ISP	Internet Service Provider
M	Number of hashes tested
MD5	Message-Digest algorithm 5
μ_j	Mean hashrate for hash type j
N	Number of GPUs
NVMe	Non-Volatile Memory Express
OpenCL	Open Computing Language
OS	Operating system
p	Price/hour
P_{TDP}	TDP in Watts
PCRE	Perl Compatible Regular Expressions
r	Difference in hashrate
RAM	Random Access Memory
SHA1	Secure Hash Algorithm 1
SHA2	Secure Hash Algorithm 2

$h_{i,CUDA}$	Hashrate of GPU i using CUDA	σ_j	Standard deviation of hash type j
$h_{i,OpenCL}$	Hashrate of GPU i using OpenCL	SSD	Solid State Drive
h_{ij}	Hashrate of GPU i for hash j	TDP	Thermal Design Power
$h_{MSSS/X}$	Hashrate per X (X: MHz / X: W)	USD, $	U.S. Dollar

7. Resources

BIRYUKOV, A., DINU, D., & KHOVRATOVICH, D. (2015). Argon2: The memory-hard function for password hashing and other applications. 18.

KAMP, P.-H. (2012). LinkedIn Password Leak: Salt Their Hide—ACM Queue. https://queue.acm.org/detail.cfm?id=2254400&doi=10.1145%2F2246036.2254400

LEININGER, H. (2014, June). PathWell: Password Topology Histogram Wear-Leveling. https://korelogic.com/Resources/Presentations/bsidesavl_pathwell_2014-06.pdf

MOORE, G. E. (1965). Cramming more components onto integrated circuits. Electronics, 38(8).

NIST. (1993). Secure Hash Standard (SHS) (Federal Information Processing Standard (FIPS) 180 (Withdrawn)). U.S. Department of Commerce. https://doi.org/10.6028/NIST.FIPS.180

RIVEST, R. (1992). The MD5 Message-Digest Algorithm (RFC 1321). Article RFC 1321. https://www.rfc-editor.org/info/rfc1321

THOMAS, K., LI, F., ZAND, A., BARRETT, J., RANIERI, J., INVERNIZZI, L., MARKOV, Y., COMANESCU, O., ERANTI, V., MOSCICKI, A., MARGOLIS, D., PAXSON, V., & BURSZTEIN, E. (Eds.). (2017). Data breaches, phishing, or malware? Understanding the risks of stolen credentials.

THE DUALITY OF RISK AND TRUST AT USAGE OF ONLINE DONATIONS

Cyntia Valociková[1]

Abstract
Why do we help strangers on the Internet? We anticipate trust while voluntarily taking risks due to becoming vulnerable by a foreign party. Practicing the help activity via the Internet is now just a click away. Social networking sites like Facebook have so-called "click-to-donate" interface where donations can be practiced easily and quickly. However, the opportunity provided by the Internet is not only exploited by non-profit organizations, users can also initiate donations for their own purposes. Helping activities have become the target of Internet fraud, building on user manipulation, exploiting human vulnerabilities to achieve the behaviour desired by attackers. One of the aims of the study is to examine the different types of money transfer scams and strategies of attackers. Another goal is to map the risks inherent in users' advanced trust, and analyse the behaviour and habits of benefactors through a literature review.

Key words
Advanced trust, online donation, money transfer scam

1. Introduction

Altruism is a prosocial behavior that can also be interpreted as a helping lifestyle that serves the interests of both parties and is encouraged by a helping tendency. However, altruism is a much more complex concept and varies depending on the disciplines (Hámori, 2003). Charitible act can be practiced online via several internet platforms (Facebook, Crowdfunder, online interfaces of non-profit organizations, etc.). According to a representative survey of Nonprofit Tech for Good (2019) explored in 119 countries, 54% of respondents (n = 6057) prefer donating online (Nonprofit Tech for Good, 2019). However, charities has become the target of Internet fraud, building on user manipulation, exploiting human "*weaknesses*" (selflessness, compassion, openness, etc.) to achieve the behavior desired by fraudsters (Holt, Smirnova, Chua, & Copes, 2015). This study provides an introduction into the world of scams and frauds through specifying the concepts, and demonstrating the taxonomy of scams. Then it will take a closer look on a special type of money transfer scams, more precisely the *charity scams* with analyzing the strategies, habits and impact of scammers on charities. The study focusing mainly on online donations (included e-mails, social media sites), and in same cases at the use of text messages. As trust is a vital factor of altruistic behaviour, the study will explore several approaches on trust-based theories to highlight the inherent risks. By understanding the operation of scammers, we can get a clearer picture about advanced trust toward strangers. The study relies on existing literature, by looking at several reports and researches in current subject providing an overview and preparing further research.

[1] Obuda University / Doctoral School on Safety and Security Sciences, Banki Donat Faculty, Becsi ut 96/B 1034 Budapest- Hungary, valocikova.cyntia@uni-obuda.hu

2. The world of frauds and scams

Internet fraud is one of the most prominent problems in the world today. The strong presence of the online society has also changed payment habits as the role of credit card payments and electronic transfers is becoming more valuable and cash payments are increasingly pushed into the background. This provides an excellent opportunity for Internet fraudsters to take advantage of modern information and communication technology to gain important values such as money, access to different systems or information (Wiggen, 2020).

Fraud is a complex activity, a concept what is hard to define with only one definition, because it can be described from many perspectives, from legal to scholar, the way of perpetration or applied techniques (Vasiu, Warren, & Mackay, 2003). According to Kevin Paul Zervos, an acknowledged lawyer stated, "*In simple terms, fraud is the art of deception for gain. Dishonesty is an essential ingredient. Fraud varies in type, size and complexity. It is encountered in many different contexts. It is very much a creature of its time; it changes as society changes, with all its different attitudes and technological advancements"* (Zervos, 1991, old.: 199). This definition does not completely portray the concept of fraud, however it shows, that fraud is evolving with time and technology adjusting to targets' habits, behaviour or personality. At this time, the concepts of fraud and scams will be used together, as many authorities, especially the European Commission doesn't make vital distinctions between these two phrases. According to the European Commission report (2020) about "*Scams and Fraud Experienced by Consumers*" 56% of Europeans experienced some kind of scam or fraud (at least once) in the past two years, and the most common type of scams and fraud was "*monetary fraud*" (39%) (European Comission, 2020). Monetary frauds or money transfer scams are a common type of frauds, which including "*moving money electronically or physically from a specified account or person to another specified account or person*" (Financial Intelligence Centre, 2020, p. 2). At this case, scammer trick the victim with a fake promise that in exchange of money the scammer will provide a product or a service as a one-time offer, what cannot be refused. However, the victim will never perceive any of this promise as it is non-existent. Money transfer scam is also known as Nigerian scam or advanced fee fraud, which rise became in 1990s, and originally spread with e-mail or letters. This type of scam also improved with the development of technology, act upon the habits and characteristics of targets (Atkins & Huang, 2013). The taxonomy of scams can differ at several countries, institutions and even researchers. According to Jakobsson (2016) there are three types of scams depending whether the attack is non-targeted, targeted or both (see *Tab. 1*). At non-targeted scams, scammers are not picking on one specific individual or group, they send out as many scams as possible, contrarily at targeted scams, scammer aim on a specific individual or group with definite attributes. Some types of scams can be both depending on the situation (Jakobsson, McCoy, Park, & Shi, 2016).

Tab. 1: Jakobsson scam taxonomy

TYPE	CATEGORY	SUBCATEGORY
NON-TARGETED	Authority	*Bank*
		Goverment, organization
	Loan	
	Lottery	
	Money Transfer	*Charity, dying person*
		Business, commodity
		Next of kin
		Widow, orphan, refugee

TARGETED	**Business E-Mail Compromise**	
	Rental	
	Romance	
BOTH TARGETED AND NON-TARGETED	**Employment**	
	Sales	
	Phishing	

Source: (Jakobsson, McCoy, Park, & Shi, 2016, old.: 9)

In a joint research of the Financial Fraud Research Center at the Stanford Center on Longevity and the FINRA Investor Education Foundation (2015) a classification system were created in order for a better and more transparent understanding of frauds. The structure is organized on five different levels depending on the key attributes of their results. The first level of the taxonomy including two categories based on the target of fraud: Fraud against an individual and Fraud against an organization. Based on categorization of Jakobsson in behalf of better understanding the taxonomy of the research by Stanford Center and FINRA is seen on *Tab. 2* (non-exhaustive list) (Stanford Center on Longevity, 2015).

Tab. 2: Stanford Center on Longevity and FINRA scam taxonomy

TYPE	CATEGORY	SUBCATEGORY
FRAUD AGAINST AND INDIVIDUAL	*Consumer Investment Fraud*	Securities fraud
		Commodities trading fraud
	Consumer Products and Services Fraud	Worthless or non-existent products or services
		Unauthorized billing for products or services
	Employment Fraud	Business opportunities fraud
		Work-at-home scam
		Government job placement scam
	Prize and Grant Fraud	Prize promotion/Sweepstakes scam
		Bogus lottery scam
		Nigerian letter fraud
	Phantom Debt Collection Fraud	Government, lender or business debt collections scam
	Charity fraud	Bogus charitable organization
		Crowdfunding for bogus cause
	Relationship & Trust Fraud	Romance scam/Sweetheart scam
		Friends or relatives imposter scam
FRAUD AGAINST AN ORGANIZATION	*Fraud against government agencies, programs, regulations, and society*	Government Programs and Regulations
	Fraud against nongovernmental businesses or organizations	Occupational Fraud (committed by internal perpetrator

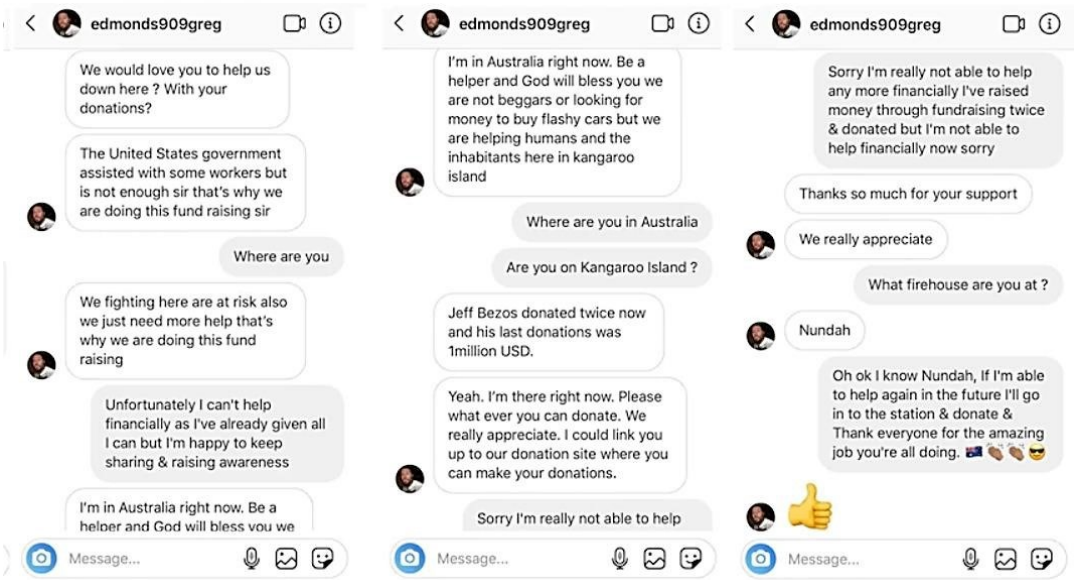
Source: (Stanford Center on Longevity, 2015, old.: 39-40)

Both categorization is valid approached from different angles, although there are several similarities in type of attacks, also most of the categories including financial transaction and cause financial loss for the victim. It is worth to take a closer look on *charity fraud or scams* as a vital type of money transfer scams, however the overview of different frauds and scams was necessary toward better understanding the "*task*" of fraudsters.

2.1 Charity fraud and the most common strategies

In case of charity fraud, attackers exploit human generosity by asking for donations for fake cause. In this case, there is no expected product or service from the business, however the joy of giving and the altruistic act is enough for the victim of its own. The fraudster ask for financial help in urgent disaster or emergency playing on human emotions. Scammers also take advantages of real disasters what cause fear and desperation in people. They often made fake profiles with fake IDs and pretending to work for a charity organization. Most of these cases are happening via e-mail our through social media sites. Several authorities and organizations (US Charity Navigator, AUS Scamwatch, UK GOV.UK, Interpol, Europol etc.) all over the world warning internet users for wariness and consciousness to avoid charity scammers (Bitaab, Cho, Oest , & Zhang, 2021). The COVID-19 pandemic changed the life of every human being in the world, affected everyone. The usage of Internet increased even more at the time of pandemic, and the unstable emotions caused by the situations made users more vulnerable for scammers. From the outbreak of pandemic (2019), the number of cyber threats like phishing, scam and fraud increased 59% all over the world according to Interpol Cybercrime Report, and COVID-19 related charity and donation requests became a common threat type (Interpol, 2020). However, not only the pandemic can be used by fraudsters as a good opportunity to cheat donation out of victims. The Australian bush fire in 2019/2020 was also a potential target for charity scammers, as according to the Australian Competition and Consumer Commission report, fake crowdfunding platforms made $4500 losses in 2019 (Australian Competition and Consumer Commission, 2020). A typical example of a charity scam can see in *Fig. 1,* where the scammer releases himself as a representative of the government asking for donation.

Fig. 1: Example of a charity scam through iMessage during the Australian bushfire 2019-20



Source: (Elsworthy, 2020)

The most common Charity scams nowadays, what are targeting the public are fake charities impersonating charities; an individual scammer sending person-in-need scams seeking for financial help for a friend or relative; vaccine or testing related scams offering fake treatments for COVID-19; government checks, when the scammer charge fake fees on behalf of the government; or bank related scams, when a scammer personate a representative from a bank asking for bank account details (Australian Competition and Consumer Commission, 2020) (Wiggen, 2020).

Charity organizations are important part of government, society and also economy, as it is done for the public benefit by reducing poverty, fighting for well-being and helping those who are in need. Charities have to demonstrate trustworthiness, to serve faithfully selfless values and proving that the donations are reaching the cause. Research of Populus (2018) among UK citizens (n=2059) shows that decreased trust in charities has been direct impact on the numbers of donations. When supporting charities, benefactors does not get any substantive product or service in exchange of support, except emotional consolidation, like the joy of selflessness. Being trusted is a vital factor for charities to raising donations and became honourable. Most important factors what makes charities trusted are transparency and the importance of cause. Donation behaviour is also increasing with trustworthiness, as benefactors are more likely to make donations in the future if the trust is steady. Threats like scams and fraud can stagger benefactors' trust and jeopardize the works of real charity organizations (Populus, 2018). Therefore trust in advance can be risky and doubtful, and to understand advanced trust toward stranger it is worth to examine the concept and models of trust and altruism.

3. Trust is key

The concept of trust is not steady in the literature. There are several approaches depending on the scientific discipline. At this time, the risk-based approach is the closest to the goal of this study. In a level of risk management, the purpose is to provide the highest possible level of safety, based on the ranking and identification of risk. Risk is an uncertain activity or event, what existence may affect the expected result in a negative way (Molnár, 2019). Defining trust shows the relationships between trust and risk. However, nor in this case this study can provide a one and only concurrent definition, rather it can provide different scientific approaches.

According to Mayer, Davis, and Schoorman trust is *''the willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party''* (Mayer, Davis, & Scoorman, 1995, p. 712). The researchers hint at trust as an action where outcome is uncertain because the trustee cannot control this interaction. The authors draw a distinction between trust and trustworthiness. At trust, it can be taken that the trustees have outlined and also aware of the possible outcomes of the events, still they take a voluntary risk, even though they become vulnerable by the other participant. On the other hand, trustworthiness is the confidence of the other participant's responsible behaviour what based on experiences. The essential difference is that while trust is an expectation, trustworthiness is a concrete certainty connected with experience (Mayer, Davis, & Scoorman, 1995).

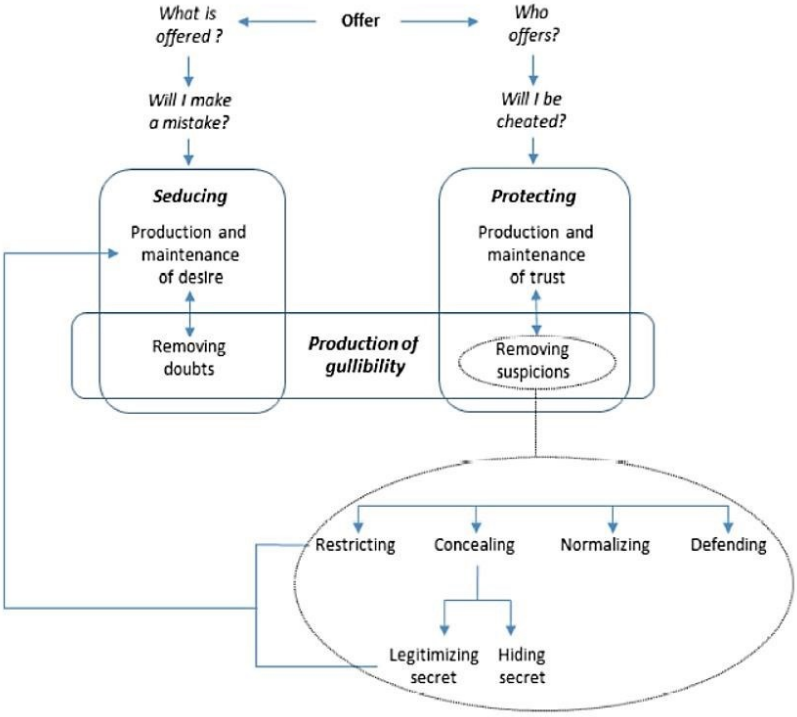
In other approach, Das & Tang in their joint work revealed *"at the subjective trust level, trust is not a subclass of risk but rather a mirror image of risk. […] In fact, both subjective trust and perceived risk represent the assessment of outcome probabilities of the same event. The crucial difference is that, whereas subjective trust portrays the assessment in a positive light — that is, the probability that the outcome will be what is desired — perceived risk describes the situation in disquieting hues — the probability that the outcome will be what is feared. In*

actuality, they are the mirror image of each other, evaluating the same situation from two distinctly different perspectives of hope and concern" (Das & Teng, 2004, p. 110). In this case, risk and trust are not highlighted as obvious contrast pairs, but these two completely different constructs were identified as theoretical opposites. While we hope for a positive outcome in event of trust, the outcome for risk can be negative. Risk-based approaches suggest that trust is relevant where risk typify the relationship between the participants. In a Hungarian study by Gelei & Dobos demonstrated that in cooperative relationships, the level of trust influences risk appetite. The stronger the trust between partners, the greater the willingness to take risks in risky situations (Gelei & Dobos , 2012). Another interesting finding, that the level of altruistic behaviour increases in parallel with the level of advanced trust (Fehr & Fishbacher, 2003). Although trust is risky, while these are working as mirror images. When trust is low between participants, risk is observed as high and backwards. The benefit of high trust should indemnify for the risk taking (Das & Teng, 2004).

When we are willing to donate money for a cause, we barely know the other party, practically we anticipating trust to a stranger (either is it a person or organization). Group and individual selection models provide possible responses to selflessness toward strangers. According to them, certain forms of human selflessness could have emerged during evolution because they increased group integration and thus helped it survive in competition with other groups. Collaborating with teammates increases reputation and those who violate the norms of collaboration are being punished and shutting out selfish individuals from the group. Some game theory experiments have shown that a proportion of people (30-50%) are willing to share some of their resources with strangers and punish the offenders in the best interests of the group. In this case, altruists do not receive reciprocal service directly from the beneficiary, but from others, mostly other members of the group, which increases their reputation and authority, leading the benefactor to expect in the future for the trust and support of others. The short-term costs of investing in donations for those in need are thus outweighed by the long-term gains from positive reputation, that the altruist will be treated differently, and better within the group. Also the willingness of donating is increasing if the noble act happening in public (Bereczkei, Birkás, & Kerekes, 2007) (Sargeant & Lee, 2004).

Excessive trust from another approach can disfigure judgment and lead to erroneous or costly decisions, therefore the trustee receiving a "*blinders*". This behaviour easily leading to gullibility, what can be defined *"as a tendency to believe items of information received from a third party and to take them at face value without looking for further evidence, and without critical analysis"* (Laroche, Steyer, & Théron, 2019, p. 643). According to Laroche, Steyer & Théron findings, gullibility is more common when social pressure is forceful. Individuals can show gullibly behaviour when scammer takes effect on proximity (time or space), and the excellence of the reward. Scammers aim is to create a situations, where they amplify the victims' gullibility and reduce the possibility of scepticism to make them act in a desired way. The exact way of creating gullibility by the scammers is seen on *Fig. 2.*

Fig. 2 The scammers production to reach the desired behaviour



Source: (Laroche, Steyer, & Théron, 2019, old.: 647)

The scam is beginning with an offer, what does not exist, however the scammers aim is to give the target to believe that the offer is real. To reach their goal, they use several tactics, like concealing, which means hiding the truth and pretending their act is legitimate. By normalizing, the scammer provides familiar characteristics of the offer usually via documents. Then scammer monitoring the parameters of the target (restricting) to reduce the risks and avoid sceptical individuals. Finally scammer using the tactic of defending, to avoid rejection usually with threatening or emotional pressure (Laroche, Steyer, & Théron, 2019). As it can be seen, scam is an elaborated operation, with complex and diversified responses. Scammers always plays on humans weaknesses, and if any individual once experiencing a threat, it is hard and long process to rebuilt trust and credibility, especially in such a fragile community as charities (Populus, 2018).

4. Conclusion

For charities the trust of benefactors is necessary to achieve their goals. Online donating became an important part for raising donations as the rapid development of Internet usage provide many opportunities. Scammers take advantages of this development, and as a thoroughly organized group of fraudsters, they adjust their techniques and strategies for the constantly changing habits and behaviours of users. As the benefactors does not get any exchange product in return of their act from individuals or organizations they did not know in person, advanced trust is a major factor to rely on. However, trust and risk can work as theoretical opposites, so the benefactor can get out of the situation in a positive and negative way as well. The level of trust affects the level of risk taking, although this level of trust is depending on the relationship between the two parties (Gelei & Dobos , 2012). In case of parties without any previous acquaintance the need of belonging on a social group can affect the

altruistic act, and trust is a determinate factor of altruism (Hámori, 2003) (Sargeant & Lee, 2004). Although, over trust can lead to gullibility what offer scammers a great opportunity to act.

As a preliminary research, it gave an interesting base for further researches. However, this study does not completely cover the complex system of trust and risk, it provided some interesting results from the literature. It is require further researches, first, does the need of belonging into a social group affects advanced trust. Second, whether there is any significant relationship between risk and gullibility. Finally, in what level of risk are benefactors are still willing to donate considering factors like the amount of donations, previous experiences with donations or the extant level of trust towards charities.

5. Resources

Atkins, B., & Huang, W. (2013). A Study of Social Engineering in Online Frauds. Open Journal of Social Sciences, 1(3), 23-32.

Australian Competition and Consumer Commission. (2020). Targeting scams 2019 A review of scam activity since 2009. Canberra, Australia: Commonwealth of Australia.

Bereczkei, T., Birkás, B., & Kerekes, Z. (2007). Selflessness to strangers - an experimental test of an evolutionary model. Magyar Pszichológiai Szemle, 62(4), 449-473.

Bitaab, M., Cho, H., Oest , A., & Zhang, P. (2021). Scam Pandemic: How Attackers Exploit Public Fear through Phishing. Cryptography and Security, 1-10.

Das, T., & Teng, B. (2004). The risk-based view of trust: a conceptual framework. Journal of Business and Psychology, 19(1), 85-119.

Elsworthy, E. (2020, February 7). *This is how you can tell if you donated to a scam this bushfire season*. ABC News. Retrieved December 29, 2021, from https://www.abc.net.au/news/2020-02-07/australia-fires-sees-spike-in-fraudsterbehaviour/11923174

European Comission. (2020). Survey on "Scams and Fraud Experienced by Consumer". Brussels: European Commission.

Fehr, E., & Fishbacher, U. (2003). The nature of human altruism. Nature, 425, 785-791.

Financial Intelligence Centre. (2020). Money Transfer Scams. Windhoek: Republic of Namibia.

Gelei, A., & Dobos , I. (2012). Trust and risk in relationships - the results of an experiment. Budapest: BCE Versenyképesség Kutató Központ.

Hámori, B. (2003). Emotion Economics. Budapest: Kossuth Kiadó.

Holt, T. J., Smirnova, O., Chua, Y. T., & Copes, H. (2015). Examining the risk reduction strategies of actors in online criminal markets. Global Crime, 16(2), 81-203.

Interpol. (2020). Cybercrime: COVID-19 Impact. Lyon, France: INTERPOL General Secretariat.

Jakobsson, M., McCoy, D., Park, Y., & Shi, E. (2016). Identifying Scams and Trends. In M.

Jakobsson, Understanding Social Engineering Based Scams (pp. 7-19). New York: Springer.

Laroche, H., Steyer, V., & Théron, C. (2019). How Could You be so Gullible? Scams and OverTrust in Organizations. Journal of Business Ethics, 160(3), 641-656.

Mayer, R. C., Davis, J. H., & Scoorman, F. D. (1995). An integrative model of organizational trust. Academy of Management Review, 20, 709-734.

Molnár, J. I. (2019). Risk analysis, protection of property and property security from the aspect of architectural crime prevention. Hadmérnök, 14(1), 17-31.

Nonprofit Tech for Good. (2019). 2018 Global Trends in Giving Report. Long Beach, CA: Nonprofit Tech for Good.

Populus. (2018). Trust in Charities, 2018 How the public views charities, what this means for the sector, and how trust can be increased. London, UK: Charity Comission for England and Wales.

Sargeant, A., & Lee, S. (2004). Donor Trust and Relationship Commitment in the U.K. Charity Sector: The Impact on Behavior. Nonprofit and Voluntary Sector Quarterly, 33(2), 185202.

Stanford Center on Longevity. (2015). Framework for a Taxonomy of Fraud. Stanford, CA, USA: Stanford Center on Longevity.

Vasiu, L., Warren, M., & Mackay, D. (2003). Defining Fraud: Issues for Organizations from an Information Systems Perspective. Proceedings of the Seventh Pacific Asia Conference on Information Systems (pp. 971-979). Adelaide, South Australia: University of South Australia.

Wiggen, J. (2020). The impact of COVID-19 on cyber crime and state-sponsored cyber activities. Berlin, Germany: Konrad-Adenauer-Stiftung e.

Zervos, K. P. (1991). Responding to Fraud in the 1990s. In P. Grabosky (Ed.), Complex Commercial Fraud: Proceedings of a Conference (pp. 199-209). Canberra: Australian Institute of Criminology .

Notes

ISSN: **2729-8493**