International Conference

Big Data, Knowledge and Control Systems Engineering

BdKCSE'2017

Sofia, Bulgaria
7 December 2017

Institute of Information and Communication Technologies - Bulgarian Academy of Sciences

John Atanasoff Society of Automatics and Infomatics

PROCEEDINGS

International Conference on

Big Data, Knowledge and Control Systems Engineering - BdKCSE'2017

7 December 2017
108 G. S. Rakovski Str., Hall 105A, 1000 Sofia, Bulgaria

Institute of Information and Communication Technologies of the Bulgarian Academy of Sciences
"John Atanasoff" Union on Automatics and Informatics, Bulgaria

Editor:
Rumen D. Andreev

Cover Designer: Ivan Panayotov
Prepress: Ivan Panayotov

Department of Communication Systems and Services
Institute of Information and Communication Technologies - Bulgarian Academy of Sciences
Bl. 2, Acad. G. Bonchev Str., 1113 Sofia, Bulgaria

Conference scope

The International Conference "Big Data, Knowledge and Control Systems Engineering" (BdKCSE'2017) aims to provide an open forum for the dissemination of the current research progress, innovative approaches and original research results on all aspects of Big Data Management, Technologies, and Applications. Organizer of the BdKCSE'2016 Conference is the Institute of Information and Communication Technologies of the Bulgarian Academy of Sciences, and co-organizer is the "John Atanasoff" Union of Automatics and Informatics, Bulgaria.

Big Data Management, Technologies, and Applications discuss the exponential growth of information size and the innovative methods for data capture, storage, sharing, and analysis. Modern technologies continue to become more complex as do the applications. The integration of technologies, complex relationships of applications and the accelerated technological changes are new challenges to technology management.

Topics such as product development, innovation management, and research and development management have become very popular. Big data spans dimensions as volume, variety, velocity, volatility and veracity, steered towards one critical destination – value. Following from these, the conference is devoted toward improving the understanding, systems engineering, human cognition and modeling, and data.

The conference will help the research community identify the novel important contributions and opportunities for recent research on the different intelligent methodologies and techniques in the field.

Table of contents

Organized by:

Institute of Information and Communication Technologies
- Bulgarian Academy of Sciences

"John Atanasoff" Union on Automatics and Informatics, Bulgaria

Our conference will take place at the
Federation of the Scientific Engineering Unions /**FNTS**/, Bulgaria

Program committee

Conference Chairs

Chairman	Assoc. Prof. Rumen Andreev	Bulgarian Academy of Sciences
Vice chairman	Assoc. Prof. Lyubka Doukovska	Bulgarian Academy of Sciences
Vice chairman	Assoc. Prof. Yuri Pavlov	Bulgarian Academy of Sciences

Program Committee

Abdel-Badeeh Salem	Ain Sham University	Egypt
Chen Song Xi	Iowa State University	USA
Dimiter Velev	University of National and World Economy	Bulgaria
František Čapkovič	Slovak Academy of Sciences	Slovakia
George Boustras	European University	Cyprus
Georgi Mengov	University of Sofia	Bulgaria
Jacques Richalet		France
John Wang	Montclair State University	USA
Kosta Boshnakov	University of Chemical Technology and Metallurgy	Bulgaria
Krasen Stanchev	University of Sofia	Bulgaria
Ljubomir Jacić	Technical College Požarevac	Serbia
Ljupco Kocarev	Macedonian Academy of Sciences and Arts	Macedonia
Milan Zorman	University of Maribor	Slovenia
Neeli R. Prasad	Aalborg University, Princeton	USA
Olexandr Kuzemin	Kharkov National University of Radio Electronics	Ukraine
	German Academic Exchange Service	Bonn
	North Rhine-Westphalia	Germany
Pedja Milosavljević	University of Niš	Serbia
Peter Kokol	University of Maribor	Slovenia
Petko Ruskov	University of Sofia	
Radoslav Pavlov	IMI, Bulgarian Academy of Sciences	Bulgaria
Rumen Nikolov	UniBIT-Sofia	Bulgaria
Song II-Yeol	Drexel University	USA
Sotir Sotirov	University “Prof. Asen Zlatarov”	Bulgaria
Tomoko Saiki	Tokyo Institute of Technology	Japan
Uğur Avdan	Anadolu University	Turkey
Valentina Terzieva	IICT, Bulgarian Academy of Sciences	Bulgaria
Valeriy Perminov	National Research Tomsk Polytechnic University	Russia
Vera Angelova	IICT, Bulgarian Academy of Sciences	Bulgaria
Vyacheslav Lyashenko	Kharkov National University of Radio Electronics	Ukraine
Wojciech Piotrowicz	University of Oxford	UK
Zlatogor Minchev	IICT, Bulgarian Academy of Sciences	Bulgaria
Zlatolilia Ilcheva	IICT, Bulgarian Academy of Sciences	Bulgaria

Publisher:

“John Atanasoff” Union on Automatics and Informatics, Bulgaria

108 G. S. Rakovski Str., 1000 Sofia, Bulgaria
Phone: +3592 987 6169
e-mail: sai.bg.office@gmail.com
www.sai.bg

ISSN: 2367 - 6450

Big Data Analytics and Genetic Research

Plamenka Borovska

Technical University of Sofia

8, Kliment Ohridksy boul. , Sofia, Bulgaria

pborovska@tu-sofia.bg

Abstract: This paper focuses on the potential of the innovative Big data technologies to facilitate genetic research for the case study of gene mapping. The problem of gene mapping has been formulated on the basis of the computer model of DNA and from the point of view of computer based technologies. Conceptual model of the Big genomic data ecosystem has been suggested and the relevant most popular genomic data platforms revealed. In silico knowledge data discovery pipeline for genome mapping based on promoters has been built up and the functionality of each stage of the pipeline has been defined.

Keywords: Big Genomic Data, in silico technology, in silico knowledge data discovery, bioinformatics.

1 Problem Area

The fundamental scientific studies are revolutionized by the big data files and streams. One of the areas of fundamental science, strongly dependent on big data, is molecular and computational biology [1]. In biological sciences there are well established practices of accumulating data in public open access biological databases, which facilitate scientists all over the world in conducting their research. Modern intensive research in bioinformatics stimulates creating innovative methods for processing and analyzes of biological data. In silico technologies, as well as next generation sequencing, resulted in exponential grow of experimental data, facing the new challenges of Big data technologies. Many scientific research teams predict that most analyses for the period till 2025 will encompass astronomy, molecular and computational biology, medicine and meteorology, as directions of fundamental science, strongly dependent and influenced by big data technologies [2].

Genetics is the branch of science concerned with genes, heredity, and variation in living organisms. It seeks to understand the process of trait inheritance from parents to offspring, including the molecular structure and function of genes, gene behaviour in the context of a cell or organism, gene distribution, and variation and change in populations [3]. One of the major topics of genetic research is gene mapping. Gene mapping may be treated in two major aspects:

(1) investigating the way how each gene works and its function, respectively, and (2) study the role that variations in genes play in disease. The results of such studies will lead to many advances in disease prevention and treatment and will stimulate biotechnologies (for ex. bacteria dissolving CO_2) as well as the design of new types of medicines tailored to an individual's unique genetic profile.

The goal of this paper is to investigate the relationships of Big data technologies and genetic research and the potential of big data analytics methods and software tools to help solving the problem of gene mapping in the field of genetic research.

2 The Computational Paradigm for Biology Research and the Problem of Gene Mapping

The computational paradigm for biology research is shown in Figure1. Exploring the evolution of biology experimentation we can distinguish throughout time several stages such as in vivo, in vitro and in silico experimentation. In silico biology involves developing and implementing computer models and conducting computer based simulations in order to investigate the most important aspects of living organisms and explore biological phenomena. In silico technologies save time and finances as well as the lives of millions of animals in developing and testing new drugs and therapies. The wide spectrum of modern high-performance experimental technologies (genomic, transcriptomic, proteomic and metabolic) generate huge amounts of data facilitating research and innovations.

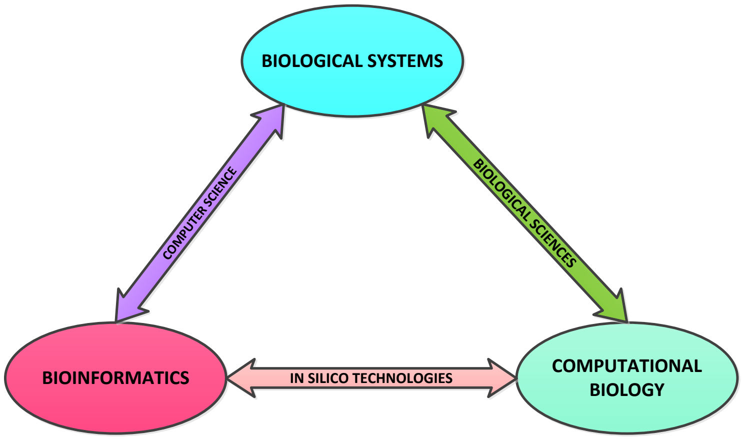


Figure 1. The computational paradigm for biology research

The first fundamental dataset in biology comprise molecular sequences. The advent of the DNA small sequence alphabet (4 nucleotides, as compared to 20 amino acids) stimulated and resulted in the comlete automation of genetic research. DNA is built up of 4 components, called nucleotides, denoted by the symbols A, C, G, T, respectively (Figure 2).

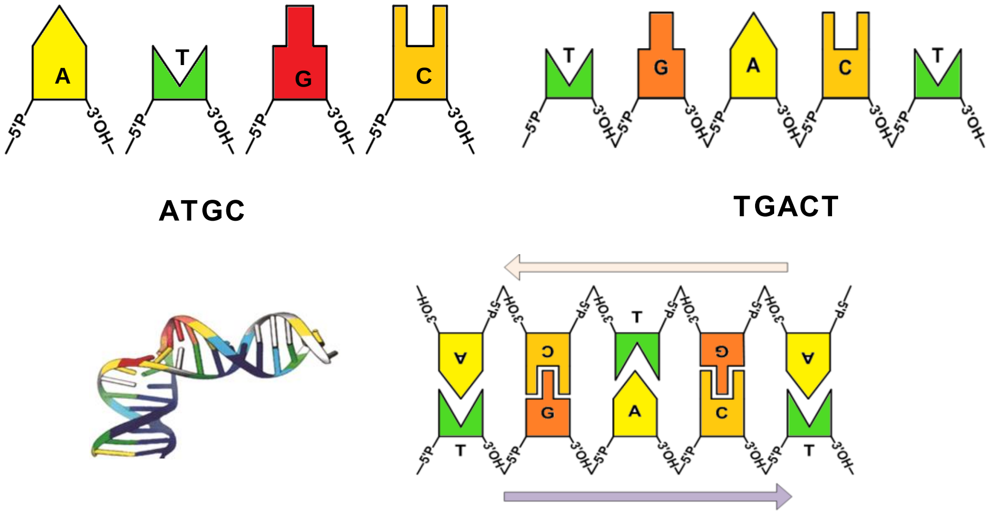


Figure 2. The computer model of DNA

The first task of functional genomics is to identify and map the genes in new sequenced genome. The human genome was sequenced in 2000, but so far only 10-15% of the genes have been identified and the function of 99% of the human DNA remains unknown [4]. To a greater extend this is valid to thousands of fully sequenced genomes of other organisms. One reason for the backwardness of the functional genomics in comparison with the structural is the lack or the shortage of reliable and efficient methods for analysis of raw genomic information. Such methods can be developed on the basis of genetic regulatory elements. They are perfect base for identification and mapping of genes as they have characteristic structures (nucleotide composition and primary structure) and are directly related to the functions of the genes [5]. The section of DNA that controls the initiation of transcription is called a gene promoter, or a promoter in the field of genetics.

Unknown gene identification and mapping in sequenced genome is fundamental task of functional genomics. Important task of the scientific research in the area of molecular biology is identifying regulatory genetic elements out of sequenced genomes, which will be used for the purpose of unknown genes identification and mapping. The initial phase of genomics aims to map and sequence an initial set of entire genomes in order to know all the genes in a genome, and the sequence of the proteins they encode. The fact that much DNA in large genomes is non-coding is a complicating circumstance. We must have in mind that non-coding DNA include introns in genes, regulatory elements of genes, multiple copies of genes, including pseudo- genes, inter-genic sequences and interspersed repeats.

As far as coding regions are of concern there are several kinds of exons: initial coding exons, internal exons and terminal exons.

The problem of gene mapping can be summarized as follows: considering the DNA coding regions we have to identify the start and the stop codon of each gene within a genome, having in mind that the start codon of a gene is "ATG" and the stop codon has 3 options – "TAA", "TAG" or "TGA" (Figure 3).

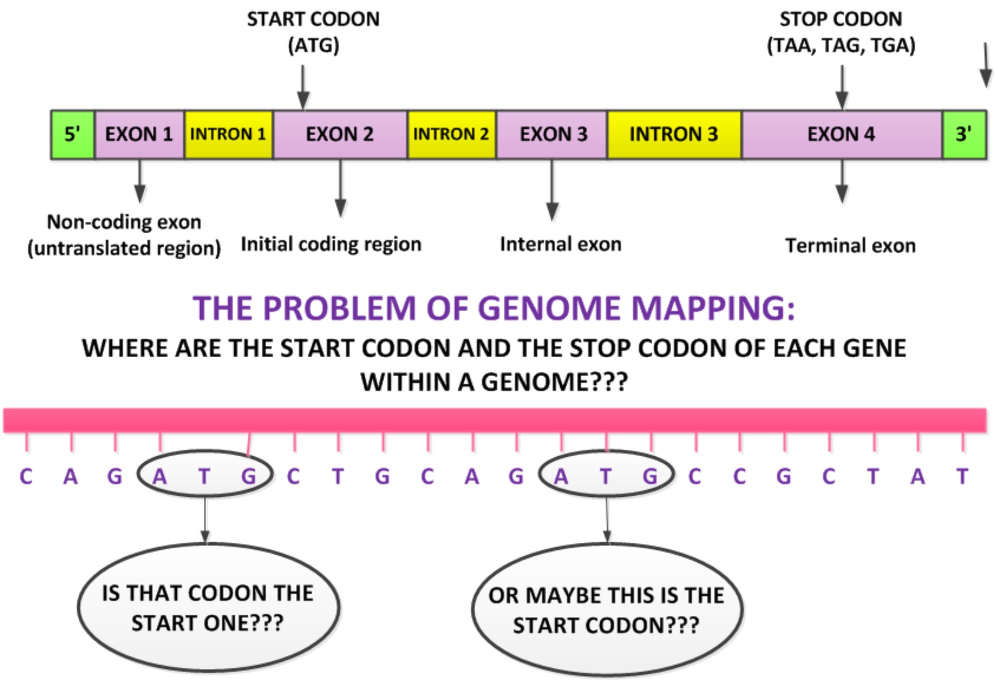


Figure 3. The problem of gene mapping

The important biological questions to answer are [6]:

(1) Which portions of DNA actually do something?

(2) Which portions of DNA actually code for protein or some other product?

(3) Which portions of DNA regulate expression?

(4) Which portions of DNA are used in replication?

3 The Big Genomic Data Ecosystem

Since 2000 medical and biological sciences entered the post-genomics era, which is associated with the emergence and intense development of the new science genomics. The prerequisites for the emergence of genomics worldwide encompass the new technology of complete genomic sequencing, i.e. determining the nucleotide sequence (primary structure) of the entire genomes. This resulted in rapid sequencing of tens of thousands of prokaryotic and thousands of eukaryotic genomes, including the human genome. DNA databases were deployed and their capacity is exponentially growing for the last decade. Genomic data has doubled up every 5 month for the last 8 years [7].

The conceptual model of big genomic data ecosystem is presented in Figure 4. The major sources of genomic data acquisition are the new DNA sequencing technologies, "omics" data generation, in silico technologies, generating huge amounts of in silico experimental data, genome databases, cancer genome databases, and the related technologies Internet of medical Things (IomT) and cloud technologies. Consequently, there is a new challenge - the capacity of genomic databases is growing up faster than the capacity of analytic tools.

Precision medicine is a hot topic nowadays. It starts with genomics and relies on the omics platforms for the analysis and interpretation of multi-scale data. Nowadays, there is a wide spectrum of Big genomic data platforms such as Google Genomics, IBM Reference Architecture for Genomics, SAP® Connected Health platform, etc. These platforms unite the efforts of developers, researchers and healthcare organizations to innovate patient-centered solutions for improving healthcare, reducing costs and providing connected healthcare services.

Google Genomics [8] offers Infrastructure as a Service (IaaS) available through their Cloud Platform, to run large-scale workloads on virtual machines on the pay-per-use principle. The Google web-based Genomics API serves as a gateway to access and use software tools from the Google portfolio of solutions. The software tool BigQuery enables very fast SQL-like queries of massive biological and medical data sets. Based on the MapReduce programming model the platform provides opportunities to discover co-relationships in genomic data through machine learning and other Knowledge Data Discoveries (KDD) methods.

Amazon's Genomics in the Cloud offers infrastructure, software tools and data sets for genomic analyses with the aim of facilitating personalized medicine [9].

IBM® end-to-end reference architecture defines the most critical capabilities for genomics computing: Data management (Datahub), workload orchestration (Orchestrator), and enterprise access (AppCenter). It can be deployed with various infrastructure and informatics technologies obeying 3 main principles: software-defined, data-centric and application-ready [7].

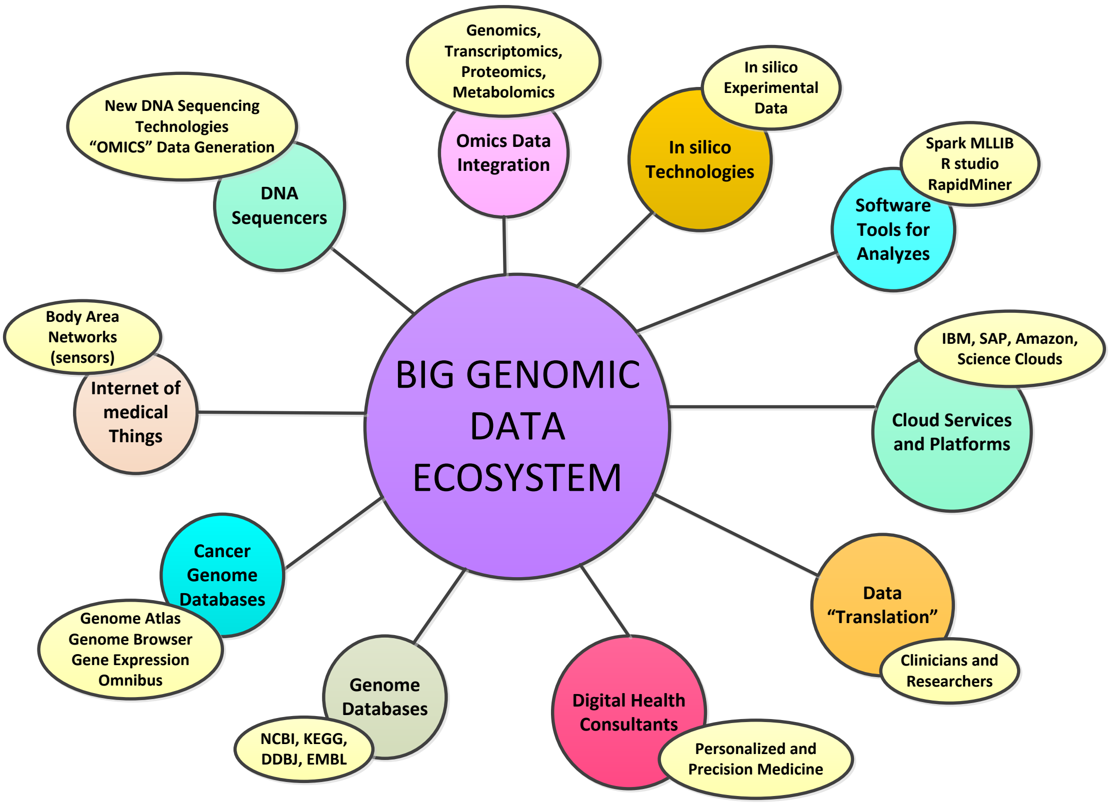


Figure 4. Conceptual model of the big genomic data ecosystem

In 2015, Intel and the Oregon Health and Science University launched a joint project - the *Collaborative Cancer Cloud* [10], actually a high-performance analytics platform accumulating and integrating private medical data targeted for cancer research. Intel intends to establish federated cloud network to other institutions, extending research to Parkinson's disease.

4 In Silico Knowledge Data Discovery for the Case Study of Gene Mapping

The major goal is to conduct scientific research in the area of molecular biology for identifying regulatory genetic elements (promoters) out of sequenced genomes, which will be used for the purpose of unknown genes identification and mapping. The idea is to extract knowledge out of genomic data of a specific organism for identifying the gene promoters and, consequently, to be able to map the genes within the genome [5].

The insilico knowledge data discovery pipeline for genome mapping based on promoters is shown in Figure 5. The first stage of the knowledge discovery process is descriptive analysis (data mining) as a result of which 2 fundamental sets are being built up: training set and validating set. After the stage of the diagnostic analysis (identifying the promoters) we come to the knowledge discovery – the actual gene mapping within the genome based on the identified promoters. Then there is the stage of predictive analysis – gene function prediction – that builds up the hypothesis followed by the stage of visualization and interpretation.

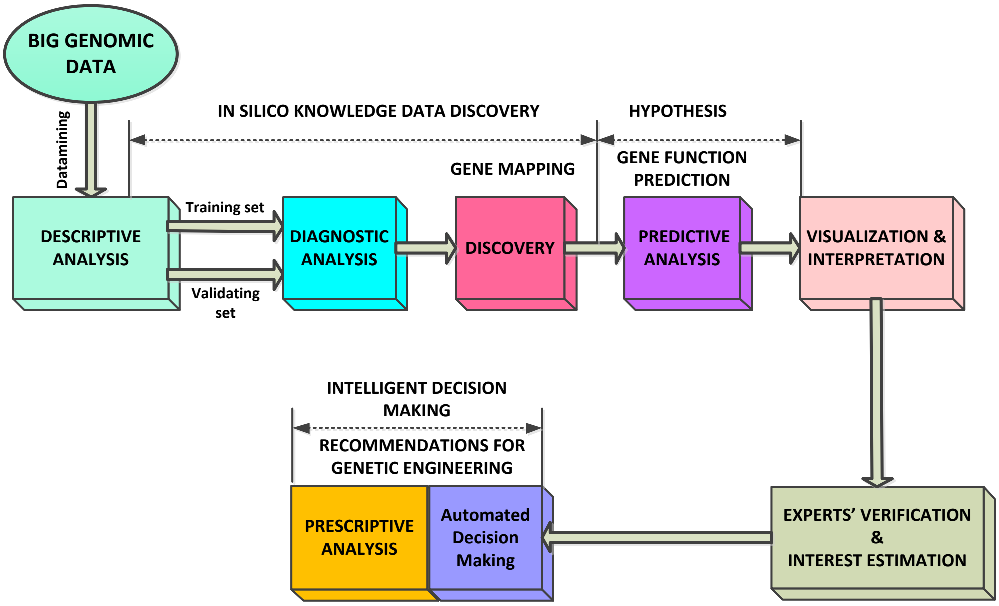


Figure 5. In silico knowledge data discovery pipeline for genome mapping based on promoters

The results obtained are verified obligatory by experts in molecular biology and genetics and the interest of the discovered knowledge has to be estimated. The final stage of prescriptive analysis involves intelligent methods for automated decision making based on metaheuristics and involves recommendations for genetic engineering.

5 Conclusions and Acknowledgements

In this paper the potential of the innovative Big data technologies to facilitate genetic research for the case study of gene mapping has been revealed. The problem of gene mapping has been formulated on the basis of the computer model of DNA and from the point of view of computer based technologies. Conceptual model of the Big genomic data ecosystem has been suggested and the relevant most popular genomic data platforms revealed. In silico knowledge data discovery pipeline for genome mapping based on promoters has been built up and the functionality of each stage of the pipeline has been defined. Future work will involve implementation of the in silico knowledge data discovery pipeline for genome mapping based on promoters and deployment of the relevant workflows.

This work is part of scientific research project "Intelligent Method for Adaptive In-silico Knowledge Discovery and Decision Making Based on Analysis of Big Data Streams for Scientific Research", ДН-07/24, financed by the National Science Fund, Competition for Financial Support for Fundamental Research – 2016, Bulgaria.

References

[1] Chen P., Zhang C, Data-intensive applications, challenges, techniques and technologies: A survey on Big Data, Journal of Information Sciences, www.elsevier.com/locate/ins

[2] Xiaolong Jin, Benjamin W.Waha, XueqiCheng, YuanzhuoWang, Significance and Challenges of Big Data Research, Big DataResearch2(2015)59–64, www.elsevier.com/locate/bdr

[3] http://www.genetics.org/

[4] https://www.genome.gov/12011238/an-overview-of-the-human-genome-project/

[5] B. Bachvarov, K. Kirilov and I. Ivanov *"Codon Usage and Gene Expression in Bacteria"*. In: *Encyclopedia of DNA Research*, Chapter 1 (Eds: Samuel J. Duncan and Patricia H. Wiley), Nova Science Publishers, Inc. NY, USA. (2012) ISBN 978-1-61324-305-3.

[6] https://www.genome.gov/10000715/genetic-mapping-fact-sheet/

[7] http://www.redbooks.ibm.com/abstracts/redp5210.html?Open

[8] https://cloud.google.com/genomics/

[9] https://aws.amazon.com/health/genomics/

[10] https://www.intel.com/content/www/us/en/cloud-computing/ohsu-precision-medical-analytics-video.html

A Use-case Driven Open Reference Architecture for Cross-domain IoT Open Source Component Based Applications

Alexandre Chikalanov, Roumen Nikolov, Iliya Vukarski, Julian Petkov,
Zlatko Petrov, Alexander Savov
University of Library Studies and Information Technologies
Bul. Tzarigradsko Shosse 119, Sofia, Bulgaria
a.chikalanov@unibit.bg, r.nikolov@unibit.bg, i.vukarski@unibit.bg, j.petkov@unibit.bg,
petrov.zlatko@gmail.com, savov@comicon.bg

Abstract: Future Internet, including Internet of Things (IoT), has to integrate the needs of all stakeholders - citizens, entrepreneurs, public administration and policy makers, in order to meet the main societal and economic needs. The user-centric and business-centric applications are expected to be based on open architectures, open source components, open standards, open innovations and technologies in order to let users share experience, knowledge, innovations and personalised experiences and insights. At the same time, the need to develop and validate multi-domain IoT solutions and methodologies require appropriate reference architectures and suitable integration and verification approaches that combine user centricity with rich functionality and high level security and privacy protection. This paper identifies the above needs and provides a use-case driven approach for identification of the requirements and implementation of a reference architecture for cross-domain IoT solutions for Smart City, such as – energy, ecology, health and agriculture. The model of Fog Computing for analyzing and acting on IoT data is being implemented, which provides the opportunities to analyze the most time-sensitive data close to where it is generated, to fast compute the IoT data and to send selected data to the cloud for further processing and analysis.

Keywords: Internet of Things (IoT), Cloud Computing, Fog Computing, Open Reference Architecture, Smart Place as a Service, Smart City

1 Introduction

The concept of Smart City as a means to enhance the quality of life of the citizen has been gaining increasing importance for technology providers, citizens and policy makers [1, 2]. As an example, there are currently more than 100 cities from 23 countries in Europe, Latin America and Asia-Pacific which have joined the European Open & Agile Smart Cities (OASC) initiative that was launched on March 2015 (oascities.org). There is a clear link between the OASC initiative and the European Strategy for Digital Single Market. OASC promotes interoperability of systems based on the free flow of data, between cities and within cities, by adopting a shared set of simple, widespread, open and freely available mechanisms. This enables the development of better and more effective smart city applications and solutions which can reach many cities once created. The vision of the OASC initiative is to create an open smart city market based on the needs of cities and communities. OASC supports development of efficient services that avoid dependence of a concrete vendor and catalyse creation of a city-based innovation ecosystem oriented towards innovative SMEs and digital entrepreneurs. OASC boosts development of solutions based on open innovation, open standards, open source and open data and thus avoiding vendor locked-in solutions. Citizens participate in co-creation and experimentation processes for the development of technologies and services and thus achieving the necessary skills and competences needed to be citizens of a Smart City [3]. One of the technology platforms that are used as a backbone of the OASC initiative is FIWARE (www.fiware.org) which is the core of an independent open community whose members are committed to implement the FIWARE mission, namely “to build an open sustainable ecosystem around public, royalty-free and implementation-driven software platform standards that will ease the development of new Smart Applications in multiple sectors”. The presented bellow use-cases are based on the FIWARE platform and enablers and follow the above mentioned principles.

A Smart City must have suitable infrastructure, human capital and information. The digital dimension has a strong relationship with intelligence and innovativeness [4]. The linking component of Big Data and Smart Cities is the worldwide network of interconnected uniquely addressable objects (Internet of Things) based on standard communication protocols. The main characteristics of Smart Cities are divided into three forms of Intelligence [5]:

- Orchestration intelligence: where cities establish institutions and community-based problem solving and collaborations;
- Empowerment intelligence: cities provide open platforms, experimental facilities and smart city infrastructure in order to cluster innovation in certain districts;
- Instrumentation intelligence: where city infrastructure is made smart through real time data collection, with analysis and predictive modeling across city districts.

This paper provides a use-case driven approach for identification of requirements and implementation of a reference architecture for cross-domain IoT solutions in a Smart City, such as – energy, ecology, health and agriculture.

2 High Level Architecture

A central objective of the paper is to describe the basic principles of design of a reference architecture - High Level Architecture (HLA), that will enable interoperability (including – at semantic level) of cross-domain application developments. This will support reusability, scalability, and interoperability of the solutions. The reference architecture will apply use-case driven approach with four use-cases from different areas with a specific focus on cross-domain use cases and cloud/fog computing solutions which will allow building digital ecosystems and large scale experimental facilities of IoT solutions with open re-usable building blocks for heterogeneous wireless sensor and actuator networks. The reference architecture design will apply the principle of Privacy by Design [6] and Security by Design [7] and thus enforce trustiness, security, safety and privacy, including the requirements of the new European Union General Data Protection Regulation [8].

The reference architecture will be in line with the requirements of big data analytics of data collected from from different network domains. The reference architecture will be also aligned with the current standardisation developments at European and international level. For instance, the reference architecture will rely on designing a proper wireless sensor and actuator networks and communication that are fully compliant with ISO 29182 (Sensor Network Reference Architecture) which provides guidance to facilitate the design and development of sensor networks, improve interoperability of sensor networks, and makes sensor networks plug-and-play, so that it becomes fairly easy to add/remove sensor nodes to/from an existing sensor network [9]. The term reference architecture in the context of sensor networks can be defined as a "framework that provides common features collected from different types of sensor networks not only to provide developmental guidelines and reuse but also to describe the interrelations and interactions among the entities in a sensor network and possibly between sensor networks" [10]. This includes definition of functionalities and services, interface protocols, interoperability and analysis of the information flow.

The concrete solutions will be based on wireless transmission with improved interoperability, security, trust and safety and will catalyze variety of applications development. This will be ensured be using open standards, interfaces and protocols for communication between the different components for use cases in different domains. This will lead to a common framework for development and technology components which will allow developers (including – third parties) to follow a standardized approach.

In order to achieve the above said, a careful analysis of the architectural requirements and technology components for each use case was done in order to design the infrastructure and the adequate relations of of entities and functionalities that ensure interoperable, standardized, and secure wireless solutions in different applications. At the same time, the reference architecture will provide different perspectives of the architecture adapted to the main requirements of the different use cases. This will be achieved by a platform for cross-domain application development, interoperability and secure cross-domain data exchange based on open standards for access to the smart infrastructure. In addition, cloud and fog integration management solutions with cross-domain authentication and authorization solutions will be demonstrated. Thus the reference architecture will bridge the gap between the local wireless communication and computing, and global cloud operation and infrastructure. A repository of open source middleware tools and application components available to external developers will be created, by following the FIWARE approach. A common component integration will be applied, e.g. agile and API-based approach to IoT integration. The solutions may involve a variety of technologies, including - Integration Platform as a Service (iPaaS). They will follow the developed by AIOTI WG3 HLA for IoT that should be applicable to AIOTI based Large Scale Pilots [11]. A key recommendation is that architectures should be described using the ISO/IEC/IEEE 42010 standard which presents an architecture in terms of multiple views in which each view adheres to a viewpoint and comprises one or more architecture models (Figure 1).

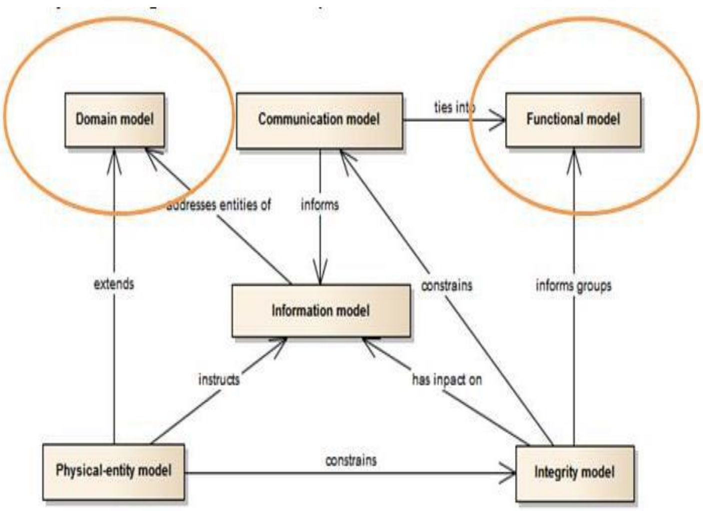


Figure 1. Architectural Models based on ISO/IEC/IEEE 42010

The functional model of AIOTI is composed of three layers:

- Application layer: contains the communications and interface methods used in process-to-process communications;

- IoT layer: groups IoT specific functions, such as data storage and sharing, and exposes those to the application layer via interfaces commonly referred to as Application Programming Interfaces (APIs). The IoT Layer makes use of the Network layer's services;
- Network layer: the services of the Network layer can be grouped into data plane services, providing short and long range connectivity and data forwarding between entities, and control plane services such as location, device triggering, quality of service or determinism.

The platform should have following merits – to support essential functionality of an IoT platform, to support real time data acquisition and processing with capability to store tens of billions data records, to endure high computation load by applying a parallel processing for conducting analyses over Big Data. The proposed architecture is based on the OpenFog Architecture model [12]. The OpenFog Architecture is a system-level architecture that extends elements of computing, networking and storage across the cloud through to the edge of the network. This approach is particularly suited to IoT systems and accelerates the decision-making velocity. OpenFog should be thought of as complementary to, and an extension of the traditional cloud based model where implementations of the architecture can reside in multiple layers of a network's topology. The goal of the OpenFog architecture is to facilitate deployments which highlight interoperability, performance, security, scalability, programmability, reliability, availability, serviceability, and agility.

Based on such an architecture the research presented in the paper is focused on finding usable open source components which can be used for building such an architecture. They comprise the set of functionalities starting with data brokers close to hardware and reaching systems for parallel processing.

Fog computing is a system-level horizontal architecture that distributes resources and services of computing, storage, control and networking anywhere along the continuum from Cloud to Things [12]. It is a:

- horizontal architecture: Support multiple industry verticals and application domains, delivering intelligence and services to users and business;
- Cloud-to-Thing continuum of services: Enable services and applications to be distributed closer to Things, and anywhere along the continuum between Cloud and Things;
- system-level extend from the Things, over the network edges, through the Cloud, and across multiple protocol.

For the future developments some characteristics of the Mobile Edge Computing (MEC) Framework and Reference Architecture [13] built on recent advances in mobile cloud computing will be harnessed as well. MEC enables the implementation of mobile edge applications as software-only entities that run on top of a virtualisation infrastructure, which is located in or close to the network edge.

The HLA is centered on use-cases as it is presented on Figure 2.

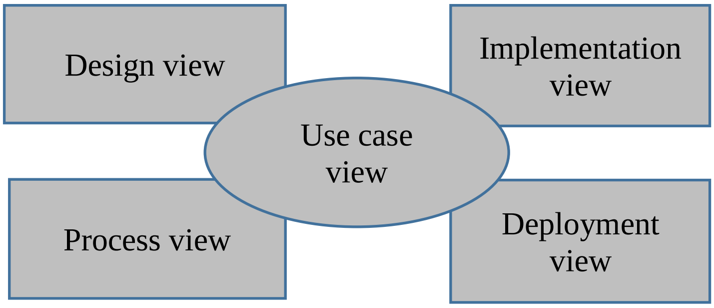


Figure 2. Use case centric architecture

The software architecture architecture used in presented four use cases is built according to a multi-tier paradigm (see Figure 3). The Cloud level consists of one or more databases and analytics tools. It can be hosted in the environment of an Open Stack cloud environment where a number of different databases coexist, such as: Craphite, HBase, MongoDB, etc. The usage of each database is related to a specific application. For similar reasons a set of coexisting different analytics tools is also envisaged, such as: Grafana, MapReduce, Apache Storm, Apache Spark, etc.

The software components positioned in the Fog tier are mostly related with data transition from sources to databases and with access validation and security. For data transmission two possible protocols are proposed – MQTT and AMPQ. Data access validation will be build according to the OAuth2.0 standard (oauth.net/2/) and a possible solution is the FIWARE component KeyRock. The Secure MQTT (mqtt.org/) can be used for secure data transmission. The role of the Gateway level is to transform the data coming in multitude formats to unified ones. The Gateway tier can be built as a stack of gateways as it is shown in Use-Case 4. The Things Tier is modeled as an abstract base-class which encapsulates all possible end points like sensors, actuators or even very complex hardware or software components, such as smart multisensory devices and software applications (e.g. chat bots as is presented in Use-Case 3).

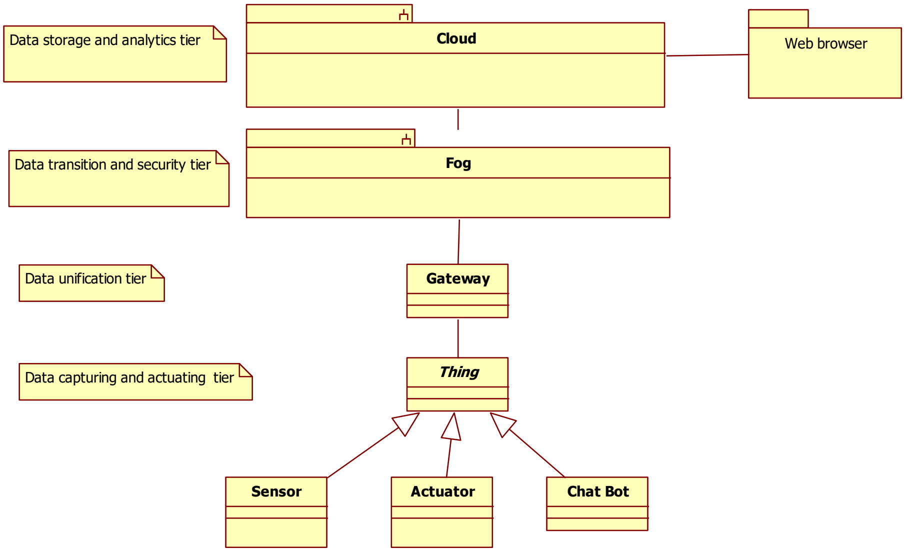


Figure 3. High level multi-tier architecture of proposed application

3 Use-Case 1: Smart Place as a Service

The general objective of the Smart Place project is to develop a model and a Future Internet (FIWARE) based platform and services for energy efficiency and user comfort monitoring and management in a Smart City building environment by using a combination of Big Data processing and simulation software [14]. Smart Place has got seed funding and support from the FIWARE Finodex accelerator (speedupeurope.eu) and from the Operational Program for Innovation and Competitiveness (www.eufunds.bg). It includes development of an integrated environment, which consists of three components:

- integrated solutions for smart management of resources and comfort: using variety of smart sensors; appliances and other devices; registered useful models for lighting and heating; development of the technology infrastructure for data analytics, mobile applications, and user interfaces (Android/iOS);
- platform for sharing of: projects, algorithms, a constructor with graphical interface for smart solutions dedicated to semi-professional/professional end-users for management of smart places;
- cloud infrastructure and services for Big Data analytics and support for providing a Smart Place as a Service (SPaaS) functionality. The envisaged products and services have innovative character in a global reach.

The integrated solutions for smart management of resources and comfort of inhabited places consist of a a variety of intelligent sensors and actuators, appliances; a set of appropriate models for lighting systems and heating; appropriate technology infrastructure and tools for Big Data analytics, mobile applications and user interfaces (Android/iOS). The model of Fog Computing is being used for analysing and acting on IoT data. This allows to analyse the most time-sensitive data at the network edge, close to where it is generated instead of sending vast amounts of IoT data to the cloud. The system acts on IoT data in milliseconds and sends selected data to the cloud for historical analysis and longer-term storage [15].

The main outcome of the first phase will be a well calibrated and evaluated (in a real environment with real users) prototype of a smart place (e.g. apartment), which includes a system for monitoring of energy consumption and comfort. This will lead to improved indicators for energy consumption and level of comfort and wellbeing. This prototype will be in the core of service for an integrated solution for a smart place, which (depending on the concrete needs and technology competence of the user) will include a complete product (software, sensors, actuators, controllers) with basic functionality, or – a service (design and implementation of an integrated solution, including a sensor network, appliances, etc.). The implemented smartphone application will allow monitoring the indicators of energy efficiency and comfort, providing control of the smart cyber-physical system, as well as to control different components. In addition, by using the concepts of Fog Computing and Cloud Computing, some functionality of data analytics and predictions will be implemented, e.g. behavior monitoring and predictions, creation of user profiles, etc. This will allow better personalisation of the application based on the evaluation of the user experience. The application will use also instruments for learning, e.g. through serious games and gamification scenarious. Different award schemes will be promoted through specialized or general social networks aiming to stimulate users to reduce energy consumption and CO_2 emissions. The prototype is based on open source components and will provide services to individual users. These services will be available through a specialized Smart Place platform and dedicated to different types of inhabited places, such as apartments and homes. Remote management of the resources and comfort will be achieved by dedicated smartphone applications and specialized data analysis and recommendations for behavior change.

Smart Place as a Service counts on an intelligent self-learning platform, which is not just storing Big Data on the cloud, but also - analysing behavior patterns and harnessing sofisticated algorithms for self-learning and optimization. (Figure 4 & Figure 5).

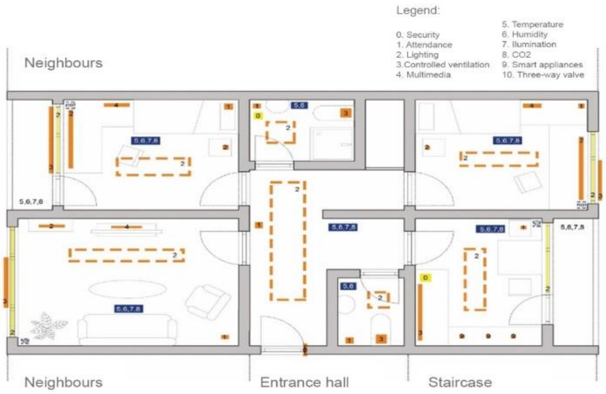


Figure 4. Scheme of a typical apartment with sensors and appliances

Figure 5. A prototype of a smartphone user interface

4 Use-Case 2: Smart Forest Ecological Management System

The Use-Case "Smart Forest Ecological Management System (ForEco)" embeds an expert system aiming at assessing conditions of forests and providing services and ground for informed management decisions to different end-users [16]. ForEco will integrate a set of models to support the adaptive management of forest ecosystems in order to achieve their sustainable development and operation as a natural resource. ForEco is based on an analysis of a large quantity of open data from different international initiatives for satellite data exploitation and will support the decision-making processes of a variety of end-users. ForEco aims at providing an adaptive and environmentally sound management of forest areas, as well as a large set of materials that are closely related to the sustainable development of regions. ForEco has got seed funding and support from the FIWARE Finodex accelerator (www.finodex-project.eu).

The reference architecture of the system is developed by extensive usage of the FIWARE middleware framework. Main advantages of such an approach are, among others, rapid development, reliability, heterogeneity, and compliance with similar external systems. Presented reference architecture is foreseen for ForEco implementation running in a Cloud and Fog computing environment.

The implementation of ForEco is based on development of following three mathematical models: asset, supply and services assessment and valuation - Ecosystem Unit Valuation (EUV) [17]; Biophysical Forests Assessment (BFA-SPPAM) [18] and a mathematical model of decision support systems - Value Based Model (VBM) [19].

In order to describe the system architecture, it is necessary to outline the main interacting actors and the ways they do it (see Figure 6). All presented users have same basic functionality. The introduction of an Abstract user is a common technique to outline same functionality for a set of users. The users also have additional specific functionality (use cases). The common use cases are: a) Sign in. The users enter their credentials and the system authenticates and authorizes them. Such users can be either humans or external applications. In case of external applications, the procedure follows the OAuth2.0 requirements for identification (oauth.net/2/); b) Sign out (users are no more allowed to interact with the system); c) Register (users enter requested identification information and additionally are approved by the System administrator).

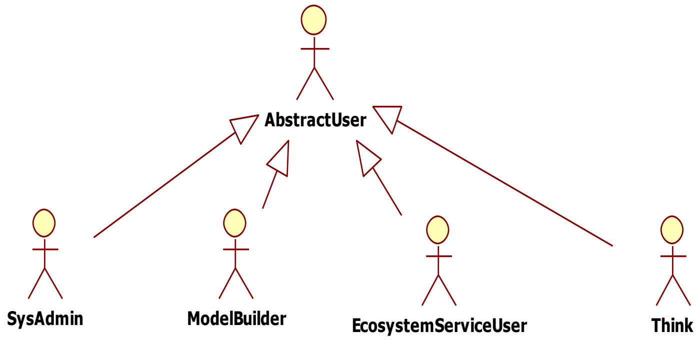


Figure 6. Main actors of Ecological Management System

The full description of all use cases defined for ForEco goes beyond the limits of present paper. In the next lines, the very specific to ForEco case – creation of the Ecosystem service models, their storage, and usage are described. The interactions between the three main actors – ModelBuilder, ForEco User, and Model repository are shown on Figure 7.

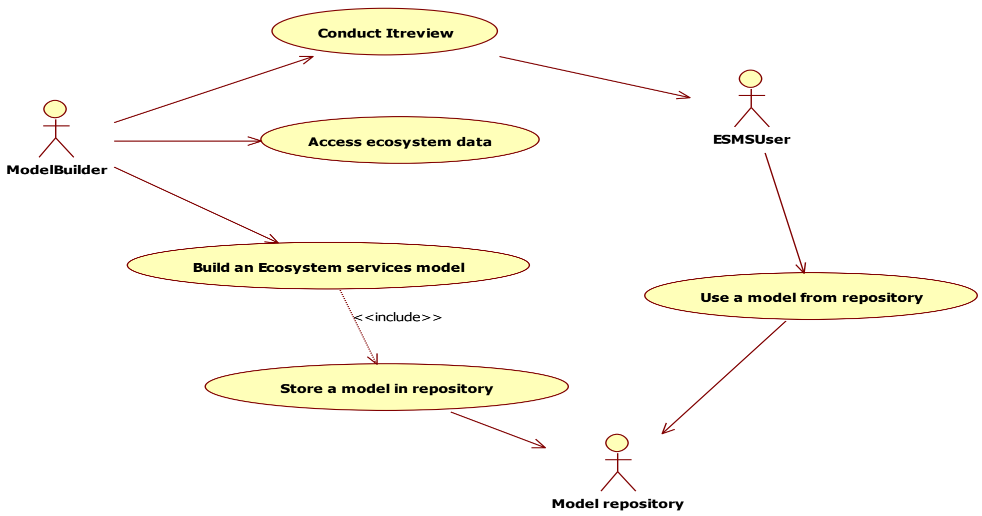


Figure 7. Use cases describing creation and usage of ESMS related models

In the presented approach the VBM models are built after a set of interviews with a number of experts in Ecosystems services denoted as ESMSUser actor and VBM experts denoted as the ModelBuilder one. After creation of a VBM model based on expert's knowledge, that model is stored into the system repository for future use by the ESMSUser actors. The model repository is denoted as Model repository actor. Usually the first two actors are humans and the third one is a database. In the use case diagram, the actor ESMSUser can be either an expert, or an Ecosystem services manager (or both at the same time).

5 Use-Case 3: Diabetics Patients Support

Diabetes Mellitus (DM) is one of the most common chronic diseases affecting humans and is rapidly expanding in prevalence worldwide. It is the most expensive disease today, placing an enormous burden on the global economy, national health systems and families. A person with DM has high blood glucose levels either because they are not producing enough insulin, or because the body does not respond properly to insulin.

There are three main types of diabetes: type 1 diabetes (T1DM), type 2 diabetes (T2DM), and gestational diabetes (GDM). At present, no cure for DM is known and self-management plays a vital role over the course of one's lifetime. Experience and forecasts suggest that DM will remain among the major global and European challenges in the upcoming decades. Technological advancements provide opportunities to remove, to some extent, the burden of self-management from the patient and to improve care plan efficiency.

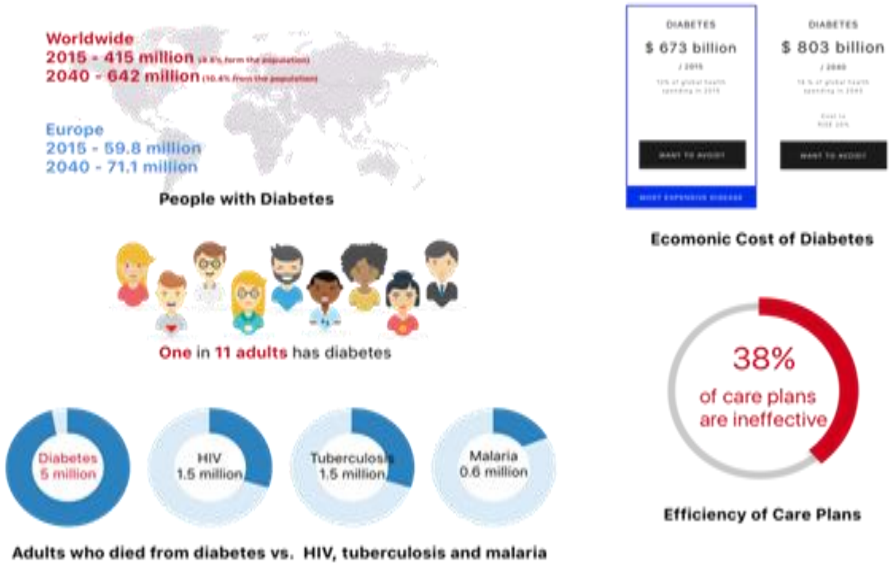


Figure 8: Key Facts about Diabetes Mellitus. Source: IDF Diabetes Atlas, 7th Ed., 2015

The NG DiaMgmt project plans to bring to the market a virtual coaching solution that will offer an individualized T2DM management recommendation system. The virtual coach builds upon strengths of human and machine intelligence to deliver an augmented intelligence solution for an effective glycemic control. The center of gravity in the solution is around actions – diabetes people are to be empowered to take timely actions and make well-informed decisions. The NG DiaMgmt introduces "machine intelligence" to help building habits for timely actions - leveraging the carrot and stick concept - reward when action is taken on time and remind when action is overdue (Figure 9). The NG DiaMgmt introduces "machine intelligence" to help better understand the current life-style of a patient and their response to

Figure 9. NG DiaMgmt solution

prescribed life-style changes and healthcare plan. The NG DiaMgmt collects and analyses data, finds correlations amongst the data, and comes up with an individualized model, as depicted in Figure 10. Data insights are set to provide vital information to the system when reflecting on prescribed care plan. Data-driven decisions are expected to help quickly revisit or refine care plans and achieve an effective glycemic control.

The NG DiaMgmt solution is based on three pillars:

- Pillar 1: a new generation wearable device that integrates several sensors for measuring data that can impact blood glucose levels.
- Pillar 2: a mobile app that delivers the virtual coach experience to prospective customers.
- Pillar 3: Software as a Service (SaaS) that provides the execution of complex algorithms, such as artificial intelligence, by meeting (near) real-time constraints. It integrates high-performant processing nodes, such as GPGPUs and FPGAs that are optimized for algorithms dominated by Multiply&Accumulate (MAC) operations (e.g. neural networks).

  Similar solution is implemented in an IoT Smart Health Cardio Belt system as well [19].

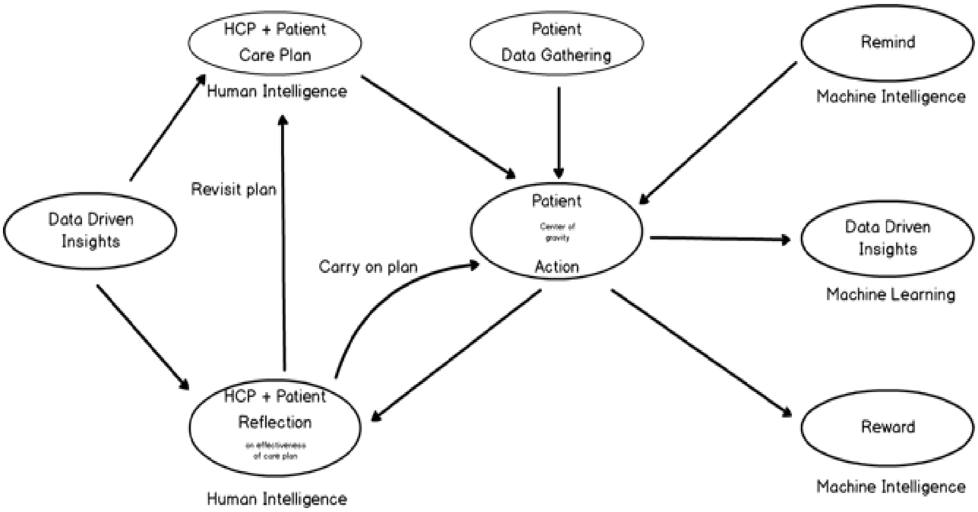


Figure 10. Model of data-driven insights and recommendation system

6 Use-Case 4: Vegetable Garden Monitoring

This use-case illustrates an approach for unifying some of the mostly used sensor protocols like ZigBee, EnOcean and MODBUS/TCP by applying gateways at different level of abstraction. A multi-tier architecture implemented for management of a vegetable garden is presented at Figure 11. The use-case defines the following levels (tiers) of abstraction:

- MODBUS/TCP level. At that level we cast sensors and actuators as MODMUS devices. Main advantage of such approach is that we consider the multitude of different IoT hardware endpoints as MODBUS devices. Additionally, MODBUS is well established and accepted industry standard and there exist gateways between Zigbee and EnOcean from one side and MODBUS controllers on the other side
- The next level of abstraction is the MQTT broker level. MQTT is wide spread protocol for data brokerage. In presented implementation we have developed a Java written gateway between MODBUS/TCP controller and Mosquito MQTT broker. The Java gateway is implemented as a MQTT publisher. The MOTDBUS/TCP gateway and MQTT publisher run on the same Raspberry PI computer. This level of abstraction is implemented on Java as a MQTT subscriber and its role is to insert data coming from sensors into a graphical database Graphite. In such a way we have the next higher level of abstraction – the database level. This level of abstraction ensures independence form of real sources of information. Any piece of information coming from outside, which has some metric data, is stored and delivered to a higher level applications in a uniform way.

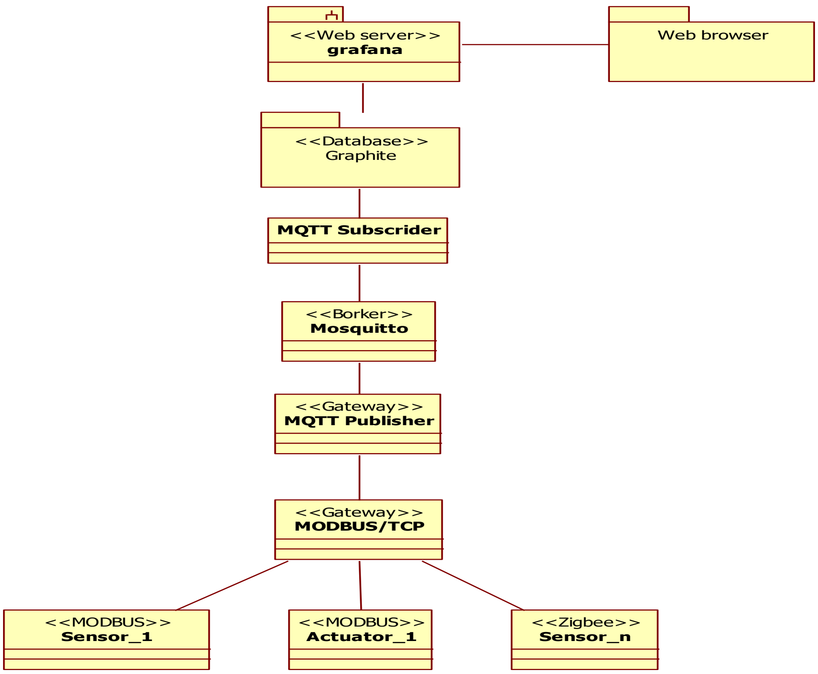


Figure 11. Multi-tier architecture for data collection and management of IoT devices

Figure 12. Graphical presentation of acquired data from some sensors managed

- In presented use-case the highest level of abstraction is a graphical presentation of the collected data Figure 12.

7 Conclusions

The paper describes a use-case driven approach and the basic principles of design of a HLA that enables interoperability of cross-domain application developments. The HLA is based on four cross-domain use cases and Cloud/Fog computing solutions which allows building digital innovation ecosystems oriented towards innovative SMEs and digital entrepreneurs that support development of IoT solutions with open re-usable building blocks for heterogeneous wireless sensor and actuator networks. Embedding the opportunities to use Fog applications makes possible implementation of IoT solutions for monitoring and analysing of real-time data from connected things and initiating actions (machine-to-machine communications, human-machine interaction) relevant to the use-case.

The adopted approach is appropriate for the implementation of regional strategies for smart/intelligent cities based on global functioning of local innovation clusters and regional systems of innovation [21]. Such cities offer the capability of online cross-domain knowledge processing and exchange and allows some basic activities in the innovation cycle (information provision, technology transfer, product development, partnership) to be performed collaboratively in a digital place. Development of a smart city and a city-based digital innovation ecosystem is the main goal of the Innovation Strategy for Smart Specialization of Sofia (www.sofia-da.eu). The authors are taking part in some of the teams of the Cluster Sofia Knowledge City (knowledgesofia.eu/en/) that are working for the implementation of this strategy.

Acknowledgements

The research is supported by the KoMEIN Project „Conceptual Modeling and Simulation of Internet of Things Ecosystems” funded by the Bulgarian National Science Foundation, Competition for financial support of fundamental research (2016) under the thematic priority: Mathematical Sciences and Informatics, contract № ДН02/1/13.12.2016.

References

[1] Christos G. Cassandras (2016), Smart Cities as Cyber-Physical Social Systems, Engineering, Volume 2, Issue 2, pp 156–158

[2] Nikolov, R., Novakova, G., Shoikova, E., (2016), A Model for Smart and Safe City, Scientific Proc. of the Intern. Conference High Technologies. Business. Society, Year XXIV, Vol. 5/191, pp. 99–103

[3] Nikolov, R., E.Shoikova, M. Krumova, E. Kovatcheva, V. Dimitrov, A. Shikalanov (2015), Learning in a Smart City Environment, Serdica Journal of Computing, Volume 9, Number 3-4

[4] Nikolov, R., Jekov, B., Mihaylova, P. (2015), Big Data in a Smart City Ecosystem: Models, Challenges and Trends, Proceedings of BdKCSE'2015 – Big Data, Knowledge and Control Systems Engineering, Sofia, 5-6 November, 2015, pp 101–118

[5] Komninos, N.(2011), "Intelligent cities: Variable geometries of spatial intelligence.". In Deakin, Mark; Al Waer, Husam. From Intelligent to Smart Cities. Journal of Intelligent Buildings International: From Intelligent Cities to Smart Cities 3 (3).

[6] Cavoukian, A. (2012), Privacy by design: origins, meaning, and prospects for assuring privacy and trust in the information era. In: Yee, G. (ed.) Privacy Protection Measures and Technologies in Business Organizations: Aspects and Standards, pp. 170–208. IGI Global, Hershey

[7] Siriwardena P. (2014) Security by Design. In: Advanced API Security. Apress, Berkeley, CA

[8] EU (2016), Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Dir 95/46/EC (General Data Protection Regulation), http://ec.europa.eu/justice/data-protection/reform/files/regulation_oj_en.pdf

[9] ISO/IEC 29182-1 (2013), Information technology — Sensor networks: Sensor Network Reference Architecture (SNRA), www.iso.org

[10] ISO/IEC 29182-2 (2013), Information technology - Sensor networks: Sensor Network Reference Architecture (SNRA) - Part 2: Vocabulary and terminology), www.iso.org

[11] AIOTI (2016), High Level Architecture (HLA), Release 3.0, AIOTI WG03 – loT Standardisation, June 2017, https://aioti.eu/aioti-wg03-reports-on-iot-standards/

[12] OpenFog (2016), OpenFog Whitepaper, February 2016, https://www.openfogconsortium.org/white-paperreference-architecture/white-paper-download-open-fog-reference-architecture/

[13] ETSI (2016), ETSI GS MEC 003 Mobile Edge Computing (MEC); Framework and Reference Architecture, ETSI GS MEC 003 V1.1.1 (2016-03), March 2016, http://www.etsi.org/deliver/etsi_gs/MEC/001_099/003/01.01.01_60/gs_MEC003v010101p.pdf

[14] Nikolov, R., Chikalanov, A., Shoikova, E., Paskalev, D., Rashevski, M. (2016), Smart Place as a Service: A Model for Providing Big Data Solutions for Smart and Energy Efficient Buildings and Places, Proc. of the Inter. Conference BdKCSE'2016 – Big Data, Knowledge and Control Systems Engineering, Sofia, 1-2 Dec

[15] Cisco (2015), Fog Computing and the Internet of Things: Extend the Cloud to Where the Things Are, Cisco White Paper, www.cisco.com/go/iot

[16] Lybenova, M., Pavlov, Y., Chikalanov, A. Spassov, K., Novakova, G., Nikolov, R. (2015), Smart Forest Ecological Management System, Journal of Balkan Ecology, Volume 18, No 4, pp 363-374

[17] Lyubenova M., Chikalanov A., Pavlov Y. (2016), Modeling of forest ecosystem services. – In: Proceedings of Abstracts of Seminar of Ecology, IBDES, 21-22, April, Sofia, Bulgaria.
[18] Lyubenova M., Chikalanov A., Lyubenova V., Vatov St. (2015), SPPAM 2.0 – Scientific description and Use. Journal of Environmental Science, Computer Science and Engineering & Technology, Sec.B, Vol. 4, Iss. 1, 37–51.
[19] Pavlov Y., Andreev R. (2013), Decision control, management, and support in adaptive and complex systems: Quantitative models, Hershey, PA: IGI Global.
[20] Kovatcheva E., Nikolov R, Madjarova M., Chikalanov A (2013) Internet of Things for Wellbeing: Pilot Case of a Smart Health Cardio Belt, in Proc. MEDICON 2013: XIII Mediterranean Conference on Medical and Biological Engineering and Computing, Seville, Sept. 25–28
[21] Komninos N.(2009), Intelligent cities: towards interactive and global innovation environments, Int. J. Innovation and Regional Development, Vol. 1, No. 4, 2009

Forest Ecological Management and Exhaustible Timber Production: Preferences Based Utility Approach

Yuri P. Pavlov

Institute of Information and Communication Technologies,

Bulgarian Academy of Sciences,

1113 Sofia, Bulgaria

yupavlov15@isdip.bas.bg, yupavlov14@hotmail.com

Abstract: In the paper is demonstrated a system engineering value driven approach within modelling of forest resource management and timber production. The principles of rationality and market efficiency need economically effective resource management of forest timber production. A multi attribute utility is evaluated and used in the dynamical lumbering multi-factor problem. The optimal control solution is based on the individual consumers' preferences represented as utility objective function.

Keywords: Preferences, Utility, Optimal Production, Timber Production, Ecology.

1 Introduction

In the paper we considered a lumbering multi-factor problem. Modelling of a forest ecosystem and lumbering is a difficult task, especially if you need its functions to be regulated and subordinated to the multifaceted sustainable use (Lyubenova and all, 2015). Everywhere the main purpose and the objectives of research are interpreted through human perceptions and concepts, which are generally even socially and economically conditioned. For such problems, the multiattribute utility theory is very useful and convenient to use (Pavlov & Andreev, 2013; Keeney & Raiffa, 1993). Sometimes it is the only way to achieve quantitative modelling and subsequent mathematical analysis based on this modelling. In this way, human behaviour can be evaluated and the decision maker involved as part of the mathematical model of the complex system. Thus, a mathematical "man-process" model is achieved, allowing computer management and even optimal control with the help of mathematical techniques and powerful computing capabilities.

2 Preferences, Utility and Timber Production Modelling

For the modelling purposes we accept three indicators (sub-objectives or factors) adequately describing the main objective of investigation (Lyubenova and all, 2015):

X_1-timber reserves ($m^3.ha^{-1}$) as representing criteria for the assessment of economic effects or material services; X_2 - species richness ($n.ha^{-1}$) as representing criteria for the assessment of ecological effect, or regulating and supporting services; X_3 - percentage of population employed in the forestry sector as representing criteria for the assessment of social effect or services.

The model is developed as multiattribute utility function with the three factors mentioned before (Lyubenova and all, 2015). The expert analysis and structuring carried out led to accepting the following sub-objectives and the appropriate criteria, which adequately describe the main objective and are real, physically measured quantities. We determine the domain of variation of representing criteria as follows:

- X_1-factor (material services as volume of timber per hectare- economic effect) [10-300 $m^3.ha^{-1}$];
- X_2- factor (regulating and supporting services - ecological effect) [1-200 number of species per hectare];
- X_3- factor (percentage of employed locals in the forestry sector - social effect) [1-30%].
- Graphically the structure of the decision maker's (DM) objective has the form shown in Figure 1. In the process of investigation independence by utility was found by the DM between the following factors:
- X_2 from X_1; X_2 from X_3;
- X_3 from X_1; X_3 from X_2;

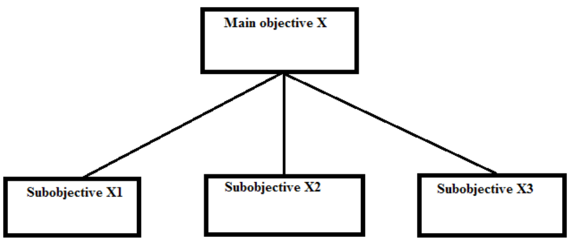


Figure 1. Structure of the main objective and sub-objectives

The DM’s preferences for x2 at different values of x1 and x3 do not change, suggesting utility independence of x2 from the changes of other two factors. Whatever the reserves of wood in the forest ecosystem are the employment of the population in the forestry sector may be different, but in any case the preferences are directed to the presence of a large species richness of the forest, i.e. great variety of species of trees, grasses, moss, lichen etc. that form the ecosystem and ensure its greater stability.

The preferences of DM for x3 at different values of x1 and x2 do not change, suggesting utility independence of x3 from the changes of other two factors. This means that whatever the reserves of wood are and at different species richness of the forest ecosystem, in any case, the preferences are aimed at the increasing of number of workers in the forestry sector. Using the theory for decomposition of multiattribute utility to simpler functions given in (Keeney & Raiffa, 1993) we determine the following utility structure:

$$\begin{aligned}
&u(X) = k_1 u(X_1; X^o_2; X^o_3) + \\
&+ f_2(X_1) \times [u(X^o_1; X_2; X^o_3)] + \\
&+ f_3(X_1) \times [u(X^o_1; X^o_2; X_3)] + \\
&+ f_{23}(X_1) \times [u(X^o_1; X_2; X^o_3)] \times [u(X^o_1; X^o_2; X_3)], \\
&\text{were } u(X^o_1; X^o_2; X^o_3) = 0 \\
&\text{and } u(X^*_1; X^*_2; X^*_3) = 1.
\end{aligned}$$

In the formula above $X^o = (X^o_1; X^o_2; X^o_3) = (10,1,1)$ and $X^* = (X^*_1; X^*_2; X^*_3) = (300,200,30)$.

The functions f_2, f_3 and f_{23} have the forms:

$$\begin{aligned}
&f_2(X_1) = u(X_1; X^*_2; X^o_3) - k_1 u(X_1; X^o_2; X^o_3), \\
&f3(X_1) = u(X_1; X^o_2; X^*_3) - k_1 u(X_1; X^o_2; X^o_3), \\
&f_{23}(X_1) = u(X_1; X^*_2; X^*_3) - f_2(X_1) - u(X_1; X^o_2; X^*_3).
\end{aligned}$$

Each of these six functions was evaluated based on the DM’s preferences (Pavlov, 2005). For example the function $u(X_1; X^o_2; X^o_3)$ has the form shown in Figure 2.

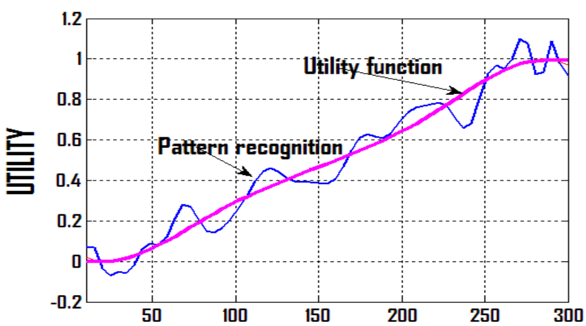
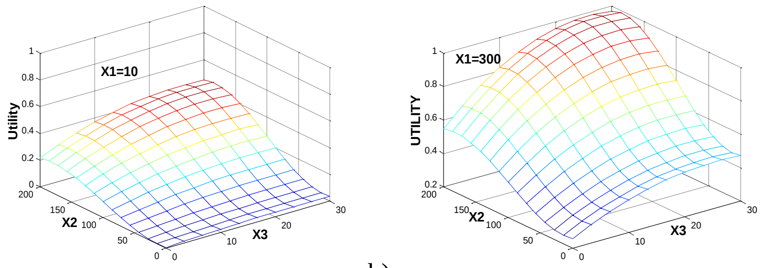


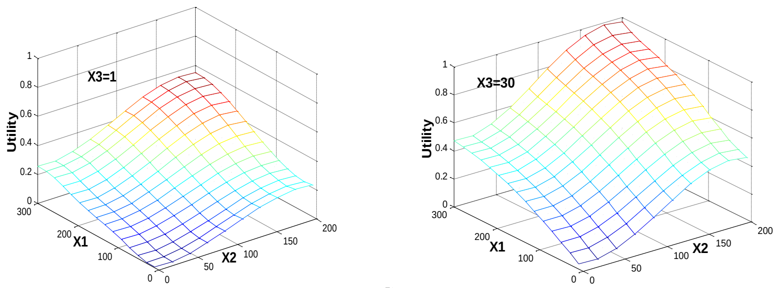
Figure 2. Utility function $u(X_1; X^o_2; X^o_3)$

The seesaw line is pattern recognition of the positive or of the negative DM's preferences (Pavlov, 2005). The solid line is the evaluated Utility function $u(X_1; X^o_2; X^o_3)$. In Figures 3, 4 and 5 are shown comparisons between the evaluated Utility function $u(10; X_2; X_3)$ and the evaluated utility function $u(300; X_2; X_3)$, between $u(X_1; X_2; 1)$ and the evaluated utility function $u(X_1; X_2; 30)$, between $u(X_1; 1; X_3)$ and the evaluated utility function $u(X_1; 200; X_3)$.



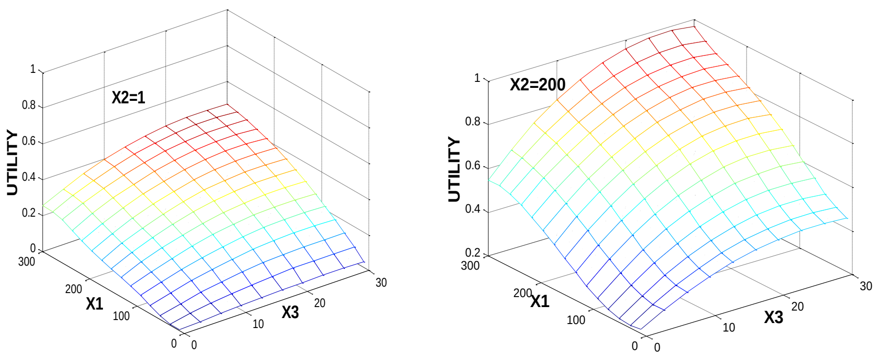
a) b)

Figure 3. Comparison between utility functions



c) d)

Figure 4. Comparison between utility functions



e) **f)**

Figure 5. Comparison between utility functions

This utility model could be perceived as a part of mathematical description of the social and ecological human expectations in the frame of the main objective of the complex lumbering problem. The described model was included as a part of the Smart Forest Ecological Management System ForEco which is based on Open Data and FIWARE. Potential users are policy makers, governmental and municipal organizations (forest and landscape planners) which are involved in forest management activities, including planning new forests in an environmentally sound way with minimal negative effects on the environment; economically effective resource management of forest timber; evaluate alternative strategies and providing comparable predictions; etc.

We will show how this utility model can be used for value based modelling and management, even for value based optimal control, taking into account the environmental, social and financial impact on the local population. For this purpose, we will use a model developed by Hotelling (1931). A logging model of forest areas of the following type is chosen:

$$\begin{aligned}
&\dot{y} = -\min(z, v) r(y), \\
&\dot{z} = \gamma w, \\
&y(0) > 0,\ z(0) > 0, \\
&0 \le v,\ 0 \le w,\ (v + w) \le L(t).
\end{aligned}$$

In the formula, with y(t) is marked the „extracted woods per hectare" - [$m^3.ha^{-1}$]. This new meaning of X_1 is just for demonstration of the logging model. With z is marked the capital invested in the equipment and with v is marked the labour as a workforce (one man - one shovel). Both dimensions are in standard units (for example, money, utils etc.). With **w** is marked that part of the labour that is used in the production of the equipment. As a criterion for optimal control, is chosen the integral:

$$\int_0^T \exp(-\delta t)(\pi \min(z, v) r(y) - \alpha v - \beta w) dt.$$

This integral has the meaning of net income for the user in a competitive market (Clarke, 1983). With T is indicated the period of work, π is the price of the timber, δ is a devaluation factor over time, α and β are respectively the value of the labour force and capital investments. Regardless of the relatively simple description, the determination of the optimal solution requires a special technique, namely non-smooth optimization (Clarke, 1983; Aubin, 2007). The solution includes the following steps. In the first step (v,w)=(z,L(t)-z). In the second step (v,w)=(z,0), and in the third step is (0,0). The function r(t) is the utility function u(X_1; 20; 5), shown on Figure 6:

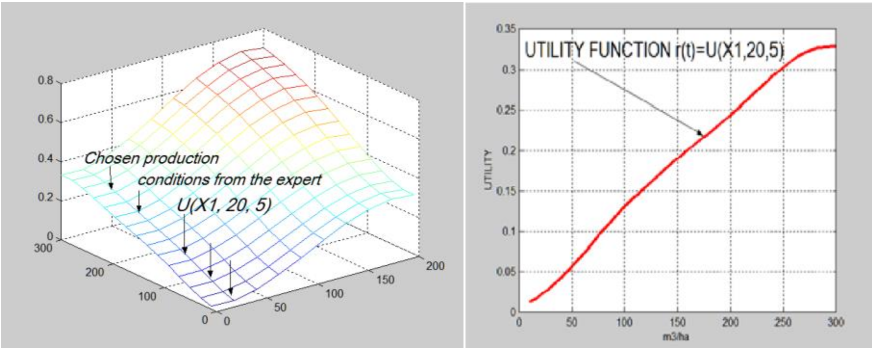


Figure 6. Utility function

The modelling is performed with parameters: y(0)=150 (m^3.ha^{-1}), T=10 (months), L=80 (thousand-BGNs), α=30 (BGNs), β=20 (BGNs) and devaluation factor δ=0.3. For the initial value of the investments is assumed z(0)=25 (thousand-BGNs) and for the coefficient of technology efficiency γ is accepted $\gamma = 1.7$. The logging was carried out under conditions of 5% social effect (X_3) and number of tree species richness, species in the forest X_2=20. The yield in cubic meters per hectare and the capital for equipment optimally varies over time as shown in Figure 7.

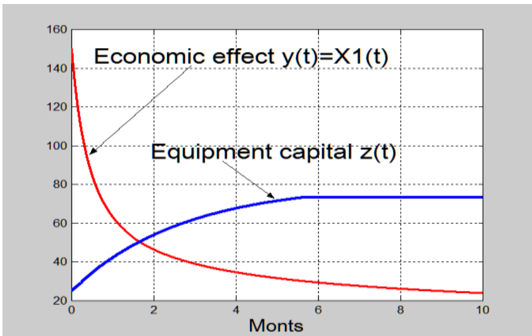


Figure 7. Optimal control solution

The changes over time of the parameter w (labour employed in the production of the equipment) is shown in Figure 8. The calculation demonstrates the ability for mathematically description of a complex production process taking into account social and economic effects. Value based models permit effective management of timber resource in accordance with DM"s preferences. The utility function u(X_1; 20; 5), is used for determination of the derivative r(t) in the dynamical model.

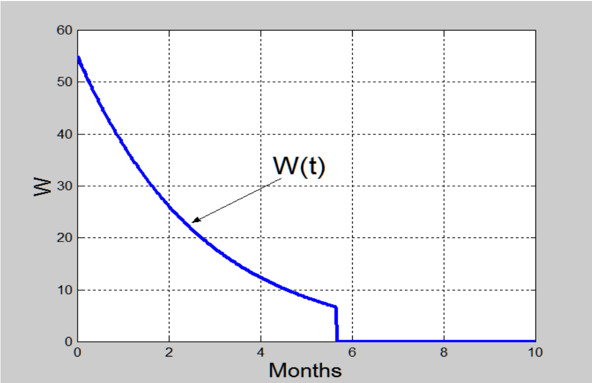


Figure 8. Labour used in the production of the equipment

That is, the DM's preferences are analytically reported in the model and indirectly in the optimal control criterion.

3 Conclusions

The optimal control is determined in line with the decision maker's preferences regarding the three factors, logging, social effect and salary and ecology on the forest. In conclusion, we would like to point out that the model is actually a mathematical model of the kind „human-dynamical production", taking into account the complex social and economic characteristics of the timber production process.

References

[1] Keeney, R. & Raiffa H. (1993). *Decision with Multiple Objectives: Preferences and Value Trade-offs,* Cambridge & New York, USA: Cambridge University Press.

[2] Hotelling, H. (1931). The economics of exhaustible resources. *J. Political Economy*, 39, 137-175.

[3] Clarke, F. (1983). *Optimization and Nonsmooth Analysis*, Wiley, New York.

[4] Pavlov, Y. & Andreev, R. (2013). *Decision control, management, and support in adaptive and complex systems: Quantitative models*, Hershey, PA: IGI Global.

[5] Pavlov, Y. (2005). Subjective preferences, values and decisions: Stochastic Approximation Approach. *Proceedings of Bulgarian Academy of Sciences*, Vol. 58 (4), 367-372.

[6] Lyubenova, M., Pavlov, Y., Chikalanov, A., Spassov, K.,· Novakova, G.,·Nikolov, R. (2015). Smart Forest Ecological Management System, *Journal of Balkan Ecology*, Vol. 18 (4), 363-375.

[7] Aubin, J-P. (2007). Mathematical Methods of Game and Economic Theory: Revised Edition, Dover Books on Mathematics.

Definite Integrals Computation with Very High Precision, Using Parallel Schemas

Velichko Dzhambov

Institute of Information and Communication Technologies, Bulgarian Academy of Sciences

Acad. G. Bonchev Str., bl. 2, 1113 Sofia, Bulgaria

vili_jambov@abv.bg

Abstract: An exceptionally effective implementation using parallel computations with two perspective quadrature schemes in a specific environment, namely .NET Framework, is discussed. Comparative tests are carried out demonstrating the possibility for effective computations with very high precision with specific tools of the environment being used. This paper describes a part of the research undertaken for the purpose of creating a tool library for arbitrary precision calculations [1]. The main goal is to demonstrate that very useful and mutually complementing computational tools, solving non-trivial problems with high precision computations may be implemented in a concrete environment, which is wide spread for personal computers, but underestimated by the software developers of scientific applications.

Keywords: High precision computation, parallel computations, numerical quadrature, computational mathematics, .NET Framework, X-MPIR.

1 Introduction

High precision (and the more so as arbitrary precision) computation is not of engineering purpose. Such computations, in particular, concerning definite integrals, are used in so called "experimental mathematics". One of the main reasons is the possibility of applying to the result an algorithm of type "identification of constant", including as an integral part "integer relation detection" algorithm (such as PSLQ), the latter requiring very high precision. Methods here presented are especially designed to use resources and tools, provided by .NET Framework. Effective implementation of two perspective quadrature schemes is discussed - *Clenshaw-Curtis* and *tanh-sinh* schemes. Our goal is to show that effective solution of non-trivial problems can be performed in this environment, which is wide spread for desktop and laptop computers, but underestimated as developing environment for scientific software. The general structure description of the program system involved and some extensions of the underlying main library, using .NET Framework C# + XMPIR, can be found in [1], but also shortly described below. More details can be found in [1], related to quadrature schemes used, as well as key fragments of program code, related to the specific implementations, using parallel computations.

2 About the Quadrature Schemes Used in the Program System

From practical point of view the problem of finding the numerical value of a given definite integral with a specified precision is almost never reduced to a single-time application of one particular quadrature formula. Probably the greatest advantage for a specific quadrature scheme is the possibility of using data from calculations previously done. In this sense the quadrature schemes of type *tanh-sinh,* a variant of the *Double Exponential Transformation*, or of type *Clenshaw-Curtis* have some advantage before the schemes using Gaussian type quadrature. Former may use already calculated abscissae and the values of the integrand in the next levels. At the *tahh-sinh* quadrature already calculated before weights may be used also but this is not of crucial importance. In Gaussian type quadratures rules like the Gauss-Kronrod are a step ahead, nevertheless if repeated calculation with greater and greater accuracy is needed this is not sufficient. In the general case, for different particular problems, the most appropriate type of quadrature scheme is different. Tanh-sinh is the best scheme for numerical integration of functions with singularities in the interval ends with requirements for high precision (1000 and more digits, [8]). It and its implementation without parallel computations have been already discussed in [2]. Here will be considered in some more details the Clenshaw-Curtis formula. The heuristic foundation of its exclusive effectiveness for some classes of integrands is the fact, that it may be interpreted like an interpolation quadrature formula not with algebraic polynomials, but with trigonometrical ones. In case when the integrand may be reduced to a smooth periodic function integrated for the interval equal to the period a result of the type "Paley-Wiener theorem" gives reasons exponential convergence to be expected with regard to the number of points in the quadrature formula. Besides if the integrand has n-th derivative with bounded variation (V) in the interval the convergence speed for both – Gaussian and Clenshaw-Curtis formulae for N points is the same, $O\left(V(2N)^{-n}\right)$. This fact refutes existing prejudices, related to the effectiveness of this formula, perceived only as algebraic order of accuracy ([4], [5]). An up-to-date concept is presented in [3].

In general both schemes used give exponential convergence, when applicable, concerning the number of points in the quadrature formula but the *tanh-sinh* scheme is of greater scope. It so to say regularizes the singularities in at the ends of the interval, transferring them to infinity.

On the other hand the *Clenshaw-Curtis* scheme, where applicable, is much more effective. It is not surprising – narrower class of admissible functions – smoothness and bounded variation of derivatives are required.

In the program system being presented a special graphical program tool NQTS, described in [2], as well as two console tools – THSHPar, CCPar are implemented. Parallel computations are used in THSHPar and CCPar. THSHPar uses the Double Exponential Transformation method. CCPar uses the Clenshaw-Curtis quadrature scheme.

3 Using the Multicore Architecture

Some preliminary details concerning base tools used for arbitrary precision computations: The possibility to use arbitrary precision computation library (MPIR [9], calling from C# in .Net Framework environment), generated from C source code is crucial (XMPIR, [10]). Detailed description can be found in [10]. As a result, dll files are generated (using mpir.c, mpir.h), enabling calls from C# via binding file (xmpir.cs). Unfortunately, this procedure don't include recently issued MPIR version. Details, below.

Specific modifications: Using quadrature schemes allowing parallel computations to be applied is one of the resources for speedup – the "theoretical" one. However the "practical" one, related to software adequacy, is important too. In this specific case some corrections had to be made in the generation of the dynamic library files of the MPIR library as well as in the wrapper file xmpir.cs.. In the original of 2010 Sergey Bochkanov uses versions of the static libraries generated by a quite old version of MPIR. The latest 2.7.0 is from June, 29, 2015. A possibility is provided libraries to be generated, which to be tuned and optimized for a particular variant of the processor. In this version however signatures of many functions are changed, e.g. in functions with mixed operands – integers with fixed and arbitrary precision, fixed integers with or without sign already correspond to C# types Int64 и UInt64. This imposed multiple changes in files 'mpir.c', used for dynamic library generation and 'xmpir.cs' used for linkage at call by C# code, so that they correspond to the head file 'mpir.h', received at the generation of the static library. The result is generation of dynamic library which is about 70% faster than the original for the particular processor.

Implementation for the tanh-sinh scheme: In the tanh-sinh scheme the nodes and weights do not depend on the integrand and the implementation of their parallel computation is almost straightforward. In such type of parallelization however the usage of independent objects and strict following for the manipulation of the current precision of the respective methods is a nicety. It should not handle with changes in the default precision – global variable for the

MPIR library. If at some places computations with different precision are necessary this should be done locally at the initialization of the respective variable – mpf_init2(*precision*).

An object of type *thsh* is in advance formed and includes the necessary lists, void, with abscissae and weights – *thsh.xc* и *thsh.wc*. Finally the effective number of abscissae and weights – *npmax*, necessary further is defined. A class *XW* is created whose basic method *xw* calculates the abscissae and weights in a given interval [idx1,idx2). It is realized quite economically. Abscissae and weights are respectively $x_k = \tanh\left(\frac{1}{2}\pi\sinh(kh)\right)$ and $w_k = \left(\frac{1}{2}h\pi\cosh(kh)\right)/\cosh^2\left(\frac{1}{2}\pi\sinh(kh)\right)$, but *sinh* and *exp* are called only once in each cycle with different arguments. *GetPi()* is repeatedly called but it is implemented in such way that uses a static field in the class *MPMath* so a value is calculated in fact only if a greater precision is required than the one, fixed in the static field.

Then the values of the function are computed in the necessary *npmax* points. Here again parallel computing is used. Method *f_full()* for object of class *MFF* is used whose main purpose is to use this method namely with input data from a list *thsh.xc* already formed, with abscissae of length *npmax*.

A separate object is again used for each problem – *FPS*, very simple in the case, whose method *processing* uses an object of class *Integrand*, independent for each problem, containing the method computing the integrand.

After this preliminary preparation the basic algorithm *tanh-sinh* is used with an essential modification consisting in that, that all values being calculated there are taken from lists which are ready and are not calculated on the spot. These latter computations are reduced to multiplication and addition only and they take a very small part of the time, needed for the through calculation.

Implementation of the Clenshaw-Curtis scheme: First of all the points in the interval [-1,1] are necessary with coordinates $\cos\left(\frac{\nu\pi}{2^n}\right)$, where *n* is the corresponding level and $\nu = 0,1,\ldots,2^n$. An exclusively quick implementation is the recursive generation [6], p. 417, using only two calls to the *sin* function.

To avoid loss of precision a little bit higher working precision must be used in calculations. If *digits* correspond to the number of decimal digits, then *digits* + log(*digits*) is enough. No parallel computations are used here. Recursive generation is very quick and takes a very small part of the total calculation time.

For computing the weights a scheme with N*log(N) operations is used [7], where $N = 2^n$ and n is the level. The core idea is in generating a vector, using the abscissae received above, for which a quick inverse Fourier transform is used, giving the weights. Here parallel computations are also not used.

The parallel computing of the integrand looks like in the same way as the one described in the *tanh-sinh* scheme. And here again for the calculations giving the successive levels no parallel computing is needed.

4 Comparative Tests

Tests were carried out on a laptop with the following characteristics: processor Intel (R) Core (TM) i7-3610QM CPU @ 2.30GHz (4 physical processors, 8 logical processors (threads) through hyper-threading technologies; up to 3.1 GHz at 4 active processors through Turbo Boost). 16 GB RAM, 64-bit operating system Windows 7 Enterprise (Microsoft Windows NT 6.1.7601 Service Pack 1). In tables below it is marked as TL (test laptop).

For a base of the investigation is accepted [8]. There 14 exemplary integrals are given computed with precision to 2000 decimal digits. The results received are compared between the proposed and developed program tools THSHPar, a scheme with parallel computing for tanh-sinh quadratures and CCPar, a scheme with parallel computations for Clenshaw-Curtis quadrature to some of the results received in [8]. As the program CCPar uses a Clenshaw-Curtis quadrature scheme data for it are shown only where it is effectively applicable. A necessary condition is that the integrand and its derivatives are continuous, with finite values for the whole interval.

Table 1. Comparative results for THSHPar and the set of examples in [8]

		Number of processors			
Example	Level required	4, [8]	16, [8]	1024, [8]	4, THSHPar+TL
Initialization for level 13		1085.34	271.87	5.02	382.31
1	10	101.63	25.55	0.53	6.11
2	10	294.32	74.04	1.54	129.03
3	10	317.01	79.69	1.83	44.75
4	10	328.73	82.13	1.63	130.31
5	9	51.62	12.90	0.30	4.69
6	10	5.62	1.42	0.05	0.43
7	10	11.46	2.87	0.10	0.65
8	9	50.98	12.85	0.27	4.70
9	10	333.24	83.60	1.84	17.54
10	10	245.45	61.39	1.44	11.84
11	11	5.17	1.30	0.04	1.67
12	12	161.99	40.71	0.80	112.47
13	13	216.50	54.13	0.97	214.40
14	13	1826.02	457.02	7.87	309.41
Total, no initialization		3949.74	989,6		988.00
Correction for hardware in [8]	1/[(3.1/2)* 1.3]	1960.17	492,12		988.00

In the last row a correction is added concerning the maximum possible acceleration related to the hardware used in [8], having in mind the maximum achieved clock frequency with Turbo Boost and the maximum percentage of improvement of about 30% provided by the hyper-threading technology in TL. Here only evident factors are considered. The speed of the main RAM memory and the processor cache may influence, but we have no data to compare to [8]. At such data not including the initialization time in both implementations – the one from [8] and the presented here, a rough account (492,12*2 = 982,24 ~ 988.00) shows that results of the present research by parallelization to 4 processors of TL are equivalent to 8 processors of the basic scheme in [8]. Without the correction the result is equal to 16 processors from [8].

For Example 14 at 13 levels (Table 1) wanted 2000 digits are not reached but up to 1971, both in [8] and THSHPar+TL. It is stated in [8] that these 2000 digits are reachable with 14 levels showing no data. This is really true. THSHPar+TL reach this precision with 14 levels in 620.90 calculations plus 768.55 seconds initialization for 14 levels.

Table 2. Comparative results for CCPar and a part of the examples in [8]

			Number of processors					
Example	Level required in [8]	Level required in CCPar	4, [8]	16, [8]	64, [8]	256, [8]	1024, [8]	4, CCPar +TL
(Initialization for level 13)/8, i.e. for level 10			135,67	33,98	8.61	2.22	0,63	0 Total time given below
1	10	12	101.63	25.55	6.45	1.65	0.53	1.84
2	10	12	294.32	74.04	18.83	4.99	1.54	10.95
3	10	10	317.01	79.69	20.42	5.24	1.83	1.23
4	10	12	328.73	82.13	20.84	5.52	1.63	10.99
Total, no initialization			1041,69	261,41	66.54	17.40	5.53	25.01
Total with init. each time, i.e. + (row #3)*4			1584,37	397.33	100.98	26.28	8.05	25.01
Total with hardware correction			786.29	197.19	50.11	13.04	4.00	25.01

Where the parallel Clenshaw-Curtis quadrature is applicable (Table 2) CCPar + TL with 4 processors is equivalent to about 128 processors of the system used in [8]. This is with the hardware correction as above. Otherwise this equals to about 256 processors. Here however another scheme is applied and such a comparison is somewhat speculative.

There is a difference which is to be noted in the denotation of the levels in the two types of schemes. In the *tanh-sinh* scheme level *k* means $20*2^k$ points, and in the *Clenshaw-Curtis* scheme level *k* means 2^k points. The computational method used here for weight calculation including fast inverse Fourier transform requires the number of points to be a power of two.

5 Discussion

Solving of a particular problem, as above mentioned, depends on many factors. At comparison some of them are unknown. Even the way of implementation of the elementary functions participating in the expression for the integrand is very important. As an example may be given the evolution of the program system considered here. It was mentioned in [2] about the program tool NQTS that the example given for integral $\int_0^{\pi/2} \frac{\arcsin\left(\sqrt{2}/2 \cdot \sin x\right)\sin x}{\sqrt{4-2\sin^2 x}}\,\mathrm{d}x$ whose analytical valuation hampers the wide spread specialized environments for mathematical calculations, after "long computations" gives the result with precision 1000 decimal digits. These "long computations" at that moment, about ttwo and a half years before THSHPar and CCPar were developed, took more than 4 hours and a half. THSHPar solves this problem in 49.51 sec. (11 levels required, i.e. $N = 20 * 2^{11} = 40960$). From them – 9.53 sec. for initialization of abscissae and weights. Computation of the integrand in the necessary *npmax*=8177 points – 39.90 sec. The difference is even greater for the achieved result with CCPar for the same problem – 3.87 sec. total, 11 levels required, i.e. $N = 2^{11} = 2048$ points). The great difference is due to improvements of all factors taking part as multipliers in the result for time. For THSHPar no interpreter is used. Increased precision is used in the scheme when this is necessary and the number of the necessary points for computing the integrand is decreased – the condition for interruption in the basic scheme of the algorithm forming parameter *npmax*. Parallel computing is used. There are improvements in the implementation of the function *atan*, *asin* is calculated through it, and in the trigonometric functions. The library XMPIR itself implements a newer version of MPIR. Besides, what is even more important, it is optimized for the concrete processor. All this, except the improvement of the scheme for *npmax* calculation, is valid for CCPar also. Something more, this scheme is more appropriate for the particular example. This does not belittle the qualities of the *tanh-sinh* scheme. It is applicable where the *Clenshaw-Curtis* scheme is not. Where there are singularities at the end of the interval, the *tanh-sinh* transform „regularizes them sending them to infinity". Something the *Clenshaw-Curtis* scheme cannot do in its classic form. Going back to the basic list of examples it can be noted however, that CCpar can solve example 13 and 14 too. Using the transform

$\int_{-1}^{1} f\left(\frac{2x}{1-x}+1\right)\frac{2}{(1-x)^2}\,\mathrm{d}x = \int_{0}^{\infty} f(x)\,\mathrm{d}x$, CCPar solves these examples for 60.62 and 652.43 sec respectively. 16 and 19 levels respectively are necessary. From formal point of view, the problem is of course in the fact that the interval for the examples is infinite and although the integrand and the derivatives behave well in it, the valuation of the error, including as a multiplier the interval length becomes undefined. Of course, transforming to finite interval can't resolve this problem, but simply move it elsewhere. For example 14 this is almost evident. At the transform the integrand's derivatives are not of bounded variation. Integrals of such type require a special approach for their effective solving.

6 Conclusion

An effective implementation of very high precision computations of definite integrals in a wide spread environment is described. Speedup achieved is due to various factors, including: 1) appropriate choice of quadrature schemes, allowing effective applying of parallel computations, 2) optimization of the basic program instruments for high precision computations, adapting them to the specific hardware, and 3) effective implementation of basic mathematical functions. Program tools are developed for computing definite integrals with high precision, THSHPar and CCPar. Tests are carried out for this implementation demonstrating the possibility for effective computations with very high precision with specific tools of the environment being used, including the multi-core processors massively built in in the modern desktop and laptop computers.

Acknowledgement

The research work reported in the paper is supported under Grant № DFNI-I-02-5 "InterCriteria Analysis: A New Approach to Decision Making".

References

[1] Dzhambov V., Retrieved from http://www.iict.bas.bg/konkursi/2017/VDjambov/disertacia.pdf. Accessed November 2017.

[2] Dzhambov V., High Precision Computing of Definite Integrals with .NET Framework C# and X-MPIR, Cybernetics and Information Technologies, 14, 1, 172-182, 2014.

[3] Reid H., Clenshaw-Curtis Quadrature, Retrieved from http://homerreid.dyndns.org/teaching/18.330/Notes/ClenshawCurtis.pdf, Accessed November 2017.

[4] Trefethen L., Six Myths of Polynomial Interpolation and Quadrature, Retrieved from https://people.maths.ox.ac.uk/trefethen/mythspaper.pdf, Accessed November 2017.

[5] Trefethen L., Is Gauss Quadrature Better than Clenshaw–Curtis?, SIAM Review, 50, 1, 67-87, 2008.

[6] Arndt J., Matters Computational, Springer, 2011.

[7] Waldvogel J., Fast Construction of the Fejér and Clenshaw–Curtis Quadrature Rules, BIT Numerical Mathematics, 46, 1, 195-202, 2006.

[8] Bailey D., Borwen J., Highly Parallel, Highly-Precision Numerical Integration, Retrieved from http://crd-legacy.lbl.gov/~dhbailey/dhbpapers/quadparallel.pdf, Accessed November 2017.

[9] MPIR site, Retrieved from http://mpir.org/, Accessed November 2017.

[10] XMPIR web location, Retrieved from http://www.alglib.net/x/xmpir/xmpir-0.2.zip, Accessed November 2017.

Challenges of Building Augmented Reality Applications Enhanced By IoT Data and Artificial Intelligence

Galia Novakova Nedeltcheva[1], Elena Shoikova[2], Elena Hristova[1]

[1] Sofia University Faculty of Mathematics and Informatics

5 James Boutchier Str., 1164 Sofia, Bulgaria

g.novak@fmi.uni-sofia.bg

[2] University of Library Studies and Information Technologies

119 Tsarigradsko Shose, 1784 Sofia, Bulgaria

e.d.shoikova@gmail.com

Abstract: Technology advancements nowadays like the Internet of Things (IoT), Augmented Reality (AR), Artificial Intelligence (AI), Cloud Technology, Mobile Internet, 3D printing, etc. have the potential to add significant business value and make daily life more convenient and enriching. Due to this every business becomes a digital business recently. The present paper focuses namely on the Digital Technologies as IoT, AI and AR which are potentially economically disruptive technologies. Some analysts are estimating AR applications for automotive, medical, military and other technology will reach $1 billion by 2018; while others estimate as much as almost $3 billion by 2019.

The present paper aims to present the current state of the art of AR, the most relevant opportunities and technical limitations as well as the challenges of building AR applications enhanced by the IoT Data and Artificial Intelligence features with the risks and rewards.

Keywords: Internet of Things (IoT), Augmented Reality (AR), Artificial Intelligence (AI), Applications, IoT Data, Risks, Rewards

1 Introduction

Cloud Computing is the underlying computing infrastructure and model that enables Big Data, IoT, social, community marketing, and many other trending topics. It is an enabler of new business models and builds the platform on which they run and this platform will enhance the growth of the business models laid above it. Cloud Computing is getting more powerful when it is enhanced with the power and capabilities of Artificial Intelligence (AI) algorithms.

In particular, the focus in the present paper is Internet of Things (IoT), Artificial Intelligence (AI), and Augmented Reality (AR) particularly which is a technology that superimposes a computer-generated image on a user's view of the real world, thus providing a composite view. This new technology blurs the line between what is real and what is computer-generated by enhancing what we see, hear, feel and smell.

The origin of the word augmented is augment, which means to add something. In the case of AR, graphics, sounds, and touch feedback are added into our natural world. Unlike virtual reality, which requires to inhabit an entirely virtual environment, AR uses the existing natural environment and simply overlays virtual information on top of it. As both virtual and real worlds harmoniously coexist, users of AR experience a new and improved world where virtual information is used as a tool to provide assistance in everyday activities.

Applications of AR can be as simple as a text-notification or as complicated as an instruction on how to perform a life-threatening surgical procedure. They can highlight certain features, enhance understandings, and provide accessible and timely data. Cell phones apps and business applications are a few of the many applications driving AR application development. The key point is that the information provided is highly relevant to what you want and you are doing.

So that, AR is a next generation of computing and next level in human-machine interaction. With its added sophistication, neatness, it adds a whole new dimension of human machine interaction, amplifying real world experience with machine added knowledge and intelligence.

Everything is changing drastically in our increasingly digitalized world: business models, working styles and customer communication. In particular, Internet is revolutionizing manufacturing. Digital technology is transforming how products are developed, produced and serviced. Big data transforms into smart data, and analytics help decision makers throughout the value chain with rich insights about the product and production life cycles. Customer requirements for product personalization are increasing as are expectations for shorter delivery times. Machine Learning (ML) as a subarea of AI provides efficient solutions for speech recognition which breaks down barriers in the act of physical communication between a human and a machine.

So, AR refers to the real-time use of information in the form of text, graphics, audio and other virtual enhancements integrated with real-world objects, according to technology research from Gartner [3]. It is the real-world element that differentiates AR from Virtual Reality (VR) as it adds value to the user’s interaction with the real world instead of just being a simulation. Having studied the potential of the technology, Gartner predicts that AR will become an important workplace tool. It is especially useful in the mobile environment because it enhances the user's senses via digital instruments to allow faster responses or decision-making [4].

On the spectrum between VR, which creates immersive, computer-generated environments, and the real world, AR is closer to the real world. AR adds graphics, sounds, haptic feedback and smell to the natural world as it exists. Both video games and cell phones are driving the development of AR. Everyone from tourists to soldiers, to someone looking for the closest subway stop can benefit from the ability to place computer-generated graphics in their field of vision.

The paper is structured as follows. After a short Introduction into the concepts of AR, VR, AI and IoT, Section 2 presents the state of the art in AR and IoT in terms of technology used, applications and limitations. The advantages and disadvantages of AR are explored and quantified, as well. Section 3 outlines the challenges to build AR applications enhanced by the AI. In Section 4 are drawn conclusions, and future research directions are also discussed in detail encompassing hardware, tracking and user-AR interaction.

2 Challenges to Build AR Applications Enhanced by IoT Data [1, 2]

The key components to AR devices are:

- *Sensors and Cameras* - Sensors are usually on the outside of the AR device and gather a user's real world interactions and communicate them to be processed and interpreted. Cameras are also located on the outside of the device and visually scan to collect data about the surrounding area. The devices take this information, which often determines where surrounding physical objects are located and then formulates a digital model to determine appropriate output.
- *Projection* - The projector can essentially turn any surface into an interactive environment. As mentioned above, the information taken by the cameras used to examine the surrounding world is processed and then projected onto a surface in front of the user; which could be a wrist, a wall, or another person. The use of projection in AR devices means that screen real estate will eventually become a lesser important component.
- *Processing* - AR devices are basically mini-supercomputers packed into tiny wearable devices. These devices require significant computer processing power and utilize many of the same components that our smartphones do. These components include a CPU, a GPU, flash memory, RAM, Bluetooth/Wifi microchip, global positioning system (GPS) microchip, and more.
- *Reflection* - Mirrors are used in AR devices to assist with the way one’s eye views the virtual image.

AR application development does have its challenges, even before IoT data is considered.

The most significant of these is the content creation which requires a certain degree of domain-level expertise to determine what to augment and where to get content from, in addition to the lead time required to do so. While using traditional tools to create a piece of content that is meaningful to interact (which can take about a week), the need for design experts or solution architects (i.e., multiple people with special and specific skillsets) makes this route prohibitive to scale.

For example, one of the upcoming AR devices such as Microsoft's Hololens, acts as a computer. This means that it is fully capable of performing complex computing operations, such as collecting data from sensors, understanding voice prompts, speech to text translation, recognizing gestures as well as computer vision, etc. With this in mind, strategically positioned solutions, exploiting benefits of this co-existence, can radically enhance end-user experience in real time.

Another example of AR in use is in the field of healthcare. From a routine checkup, to a complex surgical procedure, AR can provide immense benefits and efficiencies to both patient and healthcare professional.

AR has been popularized by consumer applications oriented around advertising, entertainment, and gaming due to the limited availability of industrial AR tools. However, enterprises are trialing AR applications today and are expected to be the primary users of AR applications through 2021 as the introduction of new solutions removes barriers to adoption by harnessing the power of these technologies. AR provides enterprises with the following benefits.

Data Reuse

AR applications allow more efficient reuse of existing enterprise application data which can include maintenance records, service tickets and usage logs—any combination of which can be tapped into to decorate an AR experience that blends the physical and digital worlds. Using these data in AR applications improves its value because they are delivered in a visually relevant way specific to the machine, thing or object to which the data apply. The advantage for the industrial enterprise is that a lot of these data are readily accessible via 3D/CAD software.

Data Contextualization

Second, AR applications can be used to better contextualize information across a range of industries and job functions. In manufacturing, this could mean giving linemen and plant personnel the ability to abstract backend systems information for 3D product navigation, step-by-step work instructions, and remote visual guidance; for design engineers, it might be the ability to accelerate the product development process without building expensive models and iterative testing procedures; for field service professionals, it is a means of getting machine data from an asset in need of attention to the service professionals that provide the fix. The common thread is that data contextualization derived from AR applications can help optimize operations measured by improvements in speed, efficiency and quality.

Work Error Reduction

AR applications can reduce errors made by workers in various occupational tasks. Whether the error is caused by too little information, too much information or poor quality information, AR applications place the right information and instructions at the location of the work task. Even simple AR applications, such as those that enable glanceable work instructions, have shown an improvement in first-time fix rates and reductions in audit-related errors (e.g., via photo work logs) ranging anywhere from 5% to 20%. In some cases, AR applications can save companies hundreds of thousands of dollars in a single day as happened when a service technician identified with his AR application that a pump flagged for replacement was mislabeled.

Workforce Multiplier

AR applications are a workforce multiplier in two ways. First, training resources are more effective and efficient. Training hours can be reduced because AR applications can include video instructions on repair/replacement tasks overlaid on the machines themselves. Second, workers can do more by themselves versus sending multiple workers or truck rolls to address the same problem or work task. These benefits are especially relevant in emerging markets where skilled and trained workers are scarce and language barriers limit the effectiveness of training time.

Safety

AR applications can also improve the safety of a work task. For instance, AR applications can include not only written instructions for removing or replacing a part on a machine, but can also include visual instructions. This assistance avoids unnecessary and possibly unsafe interaction with the machine or object.

All of these benefits are magnified when IoT data augments the AR application. For instance, operational vibration data of a generator may detect an installation issue that can be addressed at the same time when a faulty bearing is replaced. Another example is partial discharge monitoring data from utility substation transformers which can be used in an AR application that directs a utility technician to the right cabling components in need of repair or replacement. A third example is to use recommendations from analytics on IoT data to avoid certain machine components during an inspection by leveraging AR to identify internal pressure conditions or dangerous electrical charge build-up. A final example is how greater sensorization of a machine or nearby environment displayed in an AR application can improve work safety conditions for personnel in occupations such as fire and environmental protection.

The Power of IoT Platforms for Building AR Application

Driven by such a complex IoT market, the supplier community has responded by creating IoT platforms that aggregate many of the technologies and services for building IoT applications. The question is whether these platforms could be used to help building AR applications for connected products. The following are key requirements of an IoT platform for supporting AR applications today and in the future.

Device Connectivity Services

As noted earlier, before an AR application can enjoy the value of IoT data, the machine or device needs to be connected. This can include sensorization and data digitization services, protocol translation services, knowledge of device OS platforms, and relationships with gateway and communication providers. At a minimum, IoT platforms that offer a full suite of connectivity technologies/services greatly simplifies building a connected asset and speeds overall time to market for advanced AR applications.

IoT Application Development Assets

For AR applications, IoT platforms need to have, at a minimum, business rule development tools that describe actions based on values and patterns detected in the IoT data. These actions that are aligned to the business needs, can be displayed in the AR application. For example, the number of alarms that a motor has registered between inspection events. Another less often found application development asset in IoT platforms is composer tools for creating an application from IoT data. These tools can be as extensive as drag and drop environments for fast build and application deployment to a mix of traditional software development tools augmented with IoT-specific drag and drop menus. Composer tools are not necessary for AR applications, but can have an enhancing effect in driving more IoT services into an organization that can then accelerate AR application development.

Digital Twins (CAD Integrated IoT Cloud)

CAD programs, by design, build a digital twin of a device or object and AR applications and can leverage the digital twin to improve data assignment for different object features. When CAD software is enhanced with IoT data, a powerful combination is created for building AR applications. Now, developers do not have to go through the laborious process of integrating both business rule and machine attribute IoT data into the CAD-generated digital twin for use by the AR application. Pre-integration of IoT cloud data into CAD software removes these development tasks and speeds AR application development.

Analytics

Companies that have implemented IoT solutions are quickly moving from basic descriptive analytics to advanced analytics that can predict events such as machine failure, or simulate future events called prescriptive analytics. For AR applications, advanced analytics information supercharges its value. For example, field services personnel using an AR application to replace a seal or filter, may see that recent operating data suggest a more robust filter to extend the machine's lifetime under new operating conditions. This benefits the customer and the service organization, potentially eliminating an unplanned maintenance trip. IoT platforms that offer advanced analytics services are not common, but they are incredibly valuable for both IoT services and AR applications. Overall, the IoT cloud facilitates development of advanced analytics and makes it available to the AR application.

Business Integration

Because IoT platforms enable applications, many offer pre-built connectors from the IoT cloud to common enterprise applications. This speeds time to value as IoT data can now be applied to more applications and, by default, be used by more employees accessing these applications. For AR applications, with the IoT cloud as the common integration point, IoT-augmented enterprise application data can more easily be delivered and presented in the AR application.

The proceeding IoT platform capabilities greatly simplify building IoT-enhanced AR applications. However, other more important long-term benefits can be derived from using IoT platforms.

Scalability

IoT platforms provide a scalable way to both expand IoT to new AR applications and to enhance current AR applications with new IoT data. In the former case, new AR applications can leverage the same IoT cloud which has pre-established connectors for machine data and business rules. In the latter case, as analytics on machine data advances to new levels, predictive and prescriptive insights can be delivered to the AR application.

Agility

Businesses may not know which IoT data to incorporate into the AR application. IoT platforms provide flexibility and agility for selecting IoT data that provide value based on the business process and need. Today, businesses building AR applications shows simple machine operating statistics data, but may later find the need to add product and service recommendations derived from machine data analytics.

Security

Once an asset is connected, security risks are greatly raised. IoT platforms can reduce the security challenges through pre-approved device components. The approval process requires testing communications with the IoT cloud where security issues and concerns can be addressed. AR applications benefit because IoT data found in the application will be used to

interact with the object, which need to be true for operational and safety reasons.
While the AR market has been slowly overcoming barriers, the market is rapidly progressing in a way that extends AR experiences to connected things to capture the benefits outlined above. However, there are several challenges in building IoT applications to support AR application development efforts. (see Figure 1)

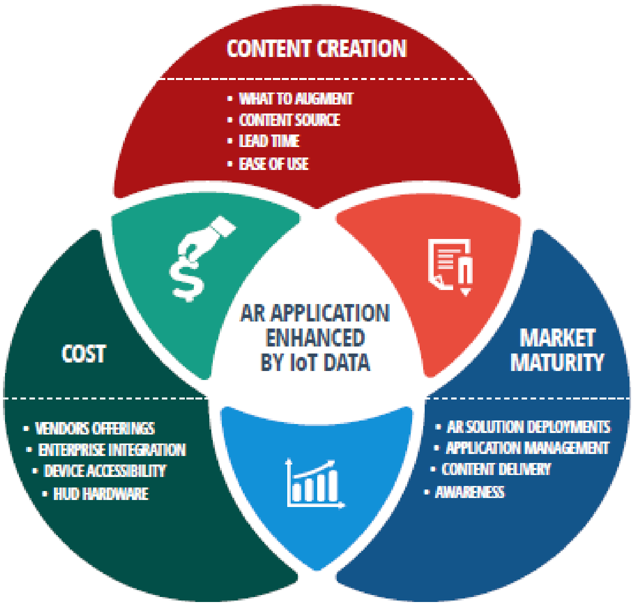


Figure 1. AR application enhanced by IoT Data [1]

The first challenge is connecting a machine or device to provide the data accessible to AR applications. Connecting a product can first involve sensorizing the machine or thing, which may require engineering expertise if off-the-shelf sensors are not available. This activity can also include use of software agents on the machine, messaging technologies and networking protocol translation services. Translation services are often required in industrial markets that use fieldbus or industrial Ethernet networking protocols. Physical transport of data from the machine is another requirement provided by multiple technologies, including fixed line, LAN, and WAN wireless connections and satellite. Once data are in the cloud, the data typically need to be normalized and categorized so applications can identify the data sets to extract from. All technologies and services that extract data from a machine or object and deliver them to a cloud are called application enablement activities which can also include writing business rules to act on machine data as well as device management services.

These activities require selecting the right technologies, but even before making technology decisions the second challenge of identifying and selecting the right suppliers arises. The

IoT boasts a vast supplier base, particularly for machine connectivity and application enablement services. In these parts of the IoT value chain alone, there are hundreds of suppliers to choose from which is why supplier diversity and offer complexity are consistently considered top challenges for the IoT market.

The second challenge is cost. Up to this point, HUD (head-up display) hardware has been prohibitively expensive to procure at enterprise-scale and the cost to develop enterprise applications usually in the tens of thousands of dollars has been a fraction of what it takes to create a tailored AR experience. But this is starting to change. Hardware is becoming cheaper as key components (optical modules, depth sensors, SOCs) become more widely available. Software development costs, the primary costs being time and labor, are also dropping because devices are becoming more accessible and vendors are expanding their offerings to include integrations with both new and existing enterprise infrastructure, in addition to traditional EMM/MDM tools.

The third challenge is market maturity. The AR market is still nascent and awareness is low. However, many enterprises that were testing AR solutions in 2015 and 2016 are now beginning to deploy them. This is where AR development platforms can help accelerate AR adoption by operationalizing not only the deployment process, but also AR content delivery and application management.

Competent IT teams can navigate these issues and develop customized AR applications. However, the difficulty is not in a single development effort, but in creating a scalable and sustainable application development capability that expands AR applications to a broader set of connected devices and continuously upgrades AR apps with new IoT data.

Organizations and their IT departments need to adapt their strategies to account for the risk and reward represented by the IoT and by commercial use of AR [2].

- *Extend social media monitoring to AR platforms*
  Since much of the AR risk is closely related to social media, leverage and extend current social media policies and monitoring to AR platforms, such as Layar and Wikitude SDK.
- *Start looking how AR can improve your business*
  Training and diagnostics are two key ways—imagine being able to have a wealth of additional information layered on to the doctor's screen, for example. Begin to think about how it can be used safely for marketing and other means to drive business for the organization.
- *Review your governance framework and update your policies*
  There are two aspects to AR. The aspect of bringing your own device (e.g., playing

Pokemon Go in the workplace which is an example of BYOAR/VR. Secondly, what is your organization's policy for use of AR as part of the business?

- *Consider the privacy aspect*

  How will the privacy policy be related to these new technologies?

- *Build security into everyday stage*

  Security is a crucial component of AR initiatives to ensure we can have confidence in the data.

Technology convergence is inevitable in times where business demands are ever increasing. On the one hand, where consumers are seeking extended coverage and ease-of-use, corporates are more concerned with ROI and improved productivity of their employees. AR and IoT powered systems are geared towards giving more intelligence to everyday job by optimizing productivity and user experience.

So that, consumer familiarity with IoT devices and AR-enhanced IoT devices should matter to enterprises for two reasons. First, as the use of AR apps grows among enterprises, it will be important for organizations to understand how much employee education will be needed. Secondly, employees may create a risk to an organization's information security or corporate reputation if they use IoT devices or AR apps in risky ways at work [2].

Table 1. Top 3 Internet of Things (IoT) devices owned by consumers [2]

AUSTRALIA	INDIA	SINGAPORE	UK	US
Smart TV (46%)	Smart TV (51%)	Smart TV (56%)	Smart TV (43%)	Smart TV (45%)
Connected car **(29%)**	Smart watch **(38%)**	Internet-connected camera **(33%)**	Internet-connected camera **(24%)**	Wireless fitness tracker **(26%)**
Internet-connected camera **(27%)**	Internet-connected camera **(35%)**	Employee access card with sensor **(30%)**	Connected car (18%), smart meter **(18%)**	Internet-connected camera **(22%)**

The IT Risk/Reward Barometer shows that a large majority of consumers in all five regions (see Table 1) see value in the range of potential applications of AR that was presented. Four of the five countries rate the following as the top three ways that AR-enhanced IoT can improve their quality of life [2].

User distraction is an obvious potential downside of AR. For example, a car windshield cluttered with too many items is probably going to be considered as a way of safe driving. For popular AR games, there is also the risk of criminals, such as the news reports about robbers who used Pokemon Go to lure players to remote spots and then rob them.

A more common threat is hackers, this is a concern shared by both IT professionals and consumers.

A risk that is specific to AR is a virtual graffiti attack – the use of AR-enhanced IoT devices to virtually deface buildings, landmarks, signage or other workplace surfaces with negative, unauthorized imagery, and then share with others.

There is an even bigger risk. AR virtual graffiti apps can also collect information from social media and sites such as Glassdoor, so companies need to be aware of this and monitor both those apps and social media. The risk is a corporate reputational one and few organizations appear to be ready to detect and manage it quickly. Organizations are better prepared for detecting content that shows up in an AR app through aggregation, i.e. publicly posted social media posts, pictures or videos that are geotagged to the organization or tagged to its signage or ads.

This gap between the risk and the preparedness of enterprises to detect and manage is not surprising as this technology is very new. For example, Pokémon Go caught many organizations off guard, and it took some time for organizations to figure out how to leverage the opportunity. These numbers are expected to increase significantly in the next year. In addition, most organizations still associate AR and VR with putting on 3D glasses. While that is certainly one aspect of it, the ability for people to use their smartphone as a viewfinder has caught some organizations off guard [2].

3 Challenges to Build Augmented Reality Applications Enhanced by Artificial Intelligence

These days, more and more companies concentrate on researching and examining AI for their purpose. Significant quantity of information about various aspects of AI application is being shared every single day all over the Internet. It is considered as only a trend by most of the people although, giants like Google, Tesla, Facebook and IBM are constantly expanding funds with an attempt to bring working products to market. AI is heard as a buzzword yet it promises better scientific development not only in the world of entertainment, but also in healthcare, transportation, robotics and other fields. Autonomous cars and trucks, image recognition, language recognition, solving IQ tests with ease are few examples of the tools built with AI which are currently put into practice.

The mixture of AR and AI could be accomplished successfully thanks to real-time recognition by the following:

- Deep Learning (DL) method

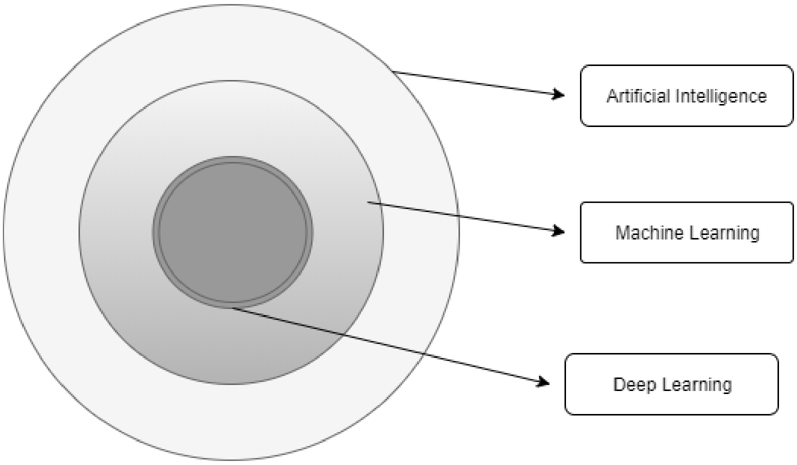


Figure 2. Relationship between Deep Learning (DL), Machine Learning (ML) and Artificial Intelligence (AI)

As the diagram above depicts, Deep Learning is the heart of AI. ML and DL differ in the notion of structuring algorithms, while ML applies algorithms to parse certain data, learn from it and make predictions or conclusions, DL construct sets of algorithms and forms an artificial "neural network" which mimic neurons in the human brain, so as to achieve an intelligent assumption on its own. In this paper the focus is set on the matter of how DL as a part of AI can enhance the value of AR applications. To benefit to their fullest from AR applications, user should be able to notice results responsively to the moment of changes. One tool enabling this is the real-time speech and image recognition accomplished with the means of DL [5].

- Possibility to use AI in the Cloud

Cloud Computing provides infrastructure services and platform services on-demand which is an advantage when there is a need for storing and managing constantly changing sets of training data used for the DL network.

- Decreased expenditure of local processing and storage and high availability

High availability and local processing and storage are extremely needed for complex DL models which tackle with large sets.

- Expanding network bandwidth, allowing richer data streams

Measured by bits per second, the network bandwidth should be big enough to enable richer data streams that can be transmitted in a fixed amount of time. Intelligent systems consume data streams of raw data receiving it from different sources so they can develop predictive analytics.

3.1 How Does AI Enhance the Value of AR Applications?

Reduction of errors resulted from humans' work

The main purpose of AI is to replace human beings in the task that they may produce errors. As it declines the rate of errors, the system is not prone to simple defects that can be caused by lack of concentration of a human. Then it comes to conducting compound tasks requiring a great amount of resources or even life-threatening medical procedures, AI is the technology which guarantees the quality of the automated operation. Combined with AR, it supplies a helpful environment which gives the opportunity to perform a successful implementation.

Identity detection

Security has a crucial role in every application which uses Internet. AR applications are threatened as well. AI is capable of protecting the applications by applying uninterrupted image recognition to check the identity specific objects or people. Moreover, identity detection can provide benefits in a communication AR environment. In business applications image detection can enrich user experience as they can offer detailed information about a project or colleagues' work.

Interaction on the next level

AI aims to imitate human and animal intelligence thus it can absolutely substitute for everyday tasks or even more advanced tasks as replacing the maid with a little home robot taking care of the hygiene at home, as well as perform some more responsible tasks in healthcare equipment which is strictly regulated by organizations.

The goal of AR is practically the same. It adds some virtual elements to reality so as to enhance user experience by artificial visualization. Combined in one, they provide the implementation to the next level. In AR the user comes as another object which is the center of the communication. They are the source of commands which are not always verbal, they can be composed from a movement of a body part or even a pupil motion. All the recognition in this scenario are in the hands of AI.

Privacy

Marked with a concern as a disadvantage of mixing these technologies, privacy in a world so dependent to technology will be in danger. Data explosion in the concept of business leaking but also of personal data leaking will be unpleasant and a major problem if strict regulation are not taken into consideration before products enter the market.

4 Conclusion

AR applications will drive the next wave of enterprise productivity through the delivery of more contextual and operationally relevant data overlaid onto a live image of a machine or object. Field services and maintenance workers will benefit the most, particularly in manufacturing, mining, logistics and supply chain, oil and gas, utilities and construction. Interestingly, these same industries are also aggressively pursuing use of IoT technologies and services to improve maintainability and reliability of the equipment and machines driving their businesses.

IoT platforms greatly accelerate the benefits of IoT applications and services that monitor equipment, predict machine outcomes and develop better products. IoT platforms simplify the selection of technologies and suppliers for machine data collection and integration into other IoT applications. Advanced IoT platforms offer deep integrations into other enterprise systems, including Computer-Aided Design (CAD), Product Life Cycle Management (PLM), Enterprise Resource Planning (ERP) as well as access to new predictive analytics toolsets. So, if IoT is the present of the interconnected world, AI is the future to come.

While AR applications are valuable without IoT data, when IoT data are added, AR applications become supercharged and IoT platforms become a powerful AR application enabler. But the benefits of IoT platforms for AR application go beyond enablement. IoT platforms provide a scalable, agile and secure way to cost-effectively enhance AR applications and extend both AR and IoT services across the business [1].

Some analysts are estimating that AR applications for automotive, medical, military and other technology will reach a $1 billion by 2018, while others estimate as much as almost $3 billion by 2019.

With the proliferation of IoT-enabled devices and the drive to provide enhanced user experiences, IoT, AR and AI combined have the power to become a source of unprecedented value and opportunity. Individuals and enterprises should focus on rapidly getting up to speed on these technologies while learning how to manage risk so they do not compromise their company's ability to innovate [2].

Acknowledgments

The research is supported by the KoMEIN Project (Conceptual Modeling and Simulation of Internet of Things Ecosystems) funded by the Bulgarian National Science Foundation, Competition for financial support of fundamental research (2016) under the thematic priority: Mathematical Sciences and Informatics, contract № DN02/1/13.12.2016.

References

[1] The power of IoT platforms for building AR applications. ©2017 ABI Research, New York, USA

[2] Risks and rewards of adding augmented reality to connected devices. ISACA 2016 IT Risk/ Reward Barometer

[3] Gartner Says Augmented Reality Will Become an Important Workplace Tool https://www.gartner.com/newsroom/id/2649315

[4] Bringing clarity to the supply chain with augmented reality https://www.singaporebusiness.com/2016/bringing-clarity-to-the-supply-chain-with-augmented-reality.html

[5] Brian J. Dooley (2017) Why AI with Augmented and Virtual Reality Will Be the Next Big Thing

Integrated System for Managing Microclimate, Lighting and Energy Consumption in Technological Premises WMCS 100

Dichko Bachvarov[1], Ani Boneva[1], Nesim Baruh[3], Simeon Angelov[2],
Yordanka Boneva[1], Zlatoliliya Ilcheva[1], Veronika Ivanova[4]

[1] Institute of Information and Communication Technologies – Bulgarian Academy of Sciences,
Acad. G. Bonchev Str., Bl. 2, Sofia 1113, Bulgaria,
dichko1952@abv.bg, a_boneva1964@abv.bg, dani_boneva@yahoo.com, zlat@isdip.bas.bg
[2]OMNITEL Ltd.,
Tsarigradsko Shousse blvd.125, bl.2, Sofia 1113, Bulgaria
simeon@omnitel.bg
[3] SD "ELL – Danev, Bozhilov &"
Stefan Karadzha Blvd 6,Office building, Floor 7, Sliven, 8801, Bulgaria
n_baruh@abv.bg
[4] Institute of Robotics - Bulgarian Academy of Sciences,
Acad G. Bonchev Str., bl.2, Sofia 1113, Bulgaria
iwanowa.w@abv.bg

Abstract: The wireless "Integrated system for monitoring environmental parameters, managing micro-climate, lighting and energy - WMCS 100" is designed to monitor and control various parameters of technological objects (greenhouses, warehouses, dryers, refrigerators, clean rooms and others) related the implementation of production processes. Software of WMCS 100 includes various modules performing specific functions and managed by a specially developed real-time operating system (M2MOS). WMCS 100 used several micro- processor modules functioning as controllers and a range of sensors for temperature, humidity, CO2, NH3, O2, atmospheric pressure, solar radiation, wind direction and wind speed, rain. For cases when have to control internal and external lighting to minimize the relevant electricity costs, WMCS 100 offers relevant LED modules and controllers for them, included into wireless local network. When to the main production facilities are added and additional activities (greenhouses, warehouses), in WMCS 100 can be included appropriate hardware and software for monitoring and controlling them according technological program. Different “WMCS 100” configurations are available, depending on customer needs

Keywords: WMCS100, MOS, ROS, interpreters, TCL, micro climate, controllers, Integrated systems, Wireless LAN.

1 Introduction

Optimizing energy consumption is important for medium and small businesses. Existing control systems are not adapted to low power consumption. For small businesses, it is important to maintain and manage the microclimate in the production premises due to the technological requirements, the requirements of the control organizations and the creation of suitable conditions for the workforce [1, 2, 3, 4]

Another task related to energy consumption is the management of support activities in the small business. These are objects such as small greenhouses, warehouses, lighting in the area of the company and others.

One possible solution to these tasks is the design of a modular distributed enterprise management and control system. As a result, an integrated management of the various activities and their centralized control is created through an operating station located on or away from the company (using wireless Internet).

The modular system presented in this article, WMCS-100, offers such a solution.

2 Architecture of WMCS 100

WMCS 100 is designed as a distributed network that includes as network nodes both wired and wireless microcontrollers. Each of them can act as a control device serving multiple own wired and wireless sensors and actuators, as well as initiators of other wired or wireless LANs. WMCS 100 supports three types of wireless networks:

- **Radio-LAN** (RLAN) uses radio-channel on 433 MHz[3];
- **Wireless-LAN** (WLAN) based on IEEE 802.15.4 and uses radio-channel on 2.44 MHz[1];
- **Hybrid-LAN** (HLAN) uses radio-channel on 433MHz and incorporates intermediate segments of wired RS 485[3].

The following devices may be used in the configuration of WMCS 100:

- <u>as controllers [8]</u>:

- ➢ **WMCS-C** - It is functioning as GPRS 3G client, “global MASTER”, MASTER or SLAVE device in WLAN, RLAN, HLAN and peripheral controller;
- ➢ **WMCS-D** - It is functioning as GPRS 3G client, “global MASTER”, MASTER or SLAVE device in WLAN, RLAN, HLAN and peripheral controller. WMCS-D is using in the systems for controlling of LEDs lighting;
- ➢ **WMCS-THC** - Peripheral controller (TC and HM) and SLAVE in RLAN and HLAN;
- ➢ **WMCS-R** - Peripheral controller of LEDs and SLIVE in WLAN;

- **WMCS-S** - Peripheral controller of specialized sensors (CO2, NH3, O2 and SLAVE in RLAN, HLAN);
- **WMCS-M** - Peripheral controller of meteorological sensors in greenhouse complex or other objects. SLAVE in RLAN or HLAN;
- **WMCS-S+** - Peripheral controller for continuous and discrete process control. SLAVE in WLAN. GPRS 2G client (via an external modem);
- **WMCS-R+** - Controller functioning as SLAVE in WLAN. It is used as an intermediate node for network expansion.

- as wires sensors:

- Sensors, working on “1-Wire bus ": electronic thermometer DS1822, humidity sensor SHT10 on 1-Wire bus, Clock-Calendar on 1- Wire bus [6, 7, 9];
- Gas sensors, forming potential outputs, and processing by WMCS-S controller: CO2, O2, NH3;
- Sensors measuring the parameters of the external environment processing by WMCS-M controller: for atmospheric pressure, solar radiation, wind direction and wind speed, rain.

The controllers are supporting wireless communications based on custom stack uMac [1] at carried frequency of 2.44 GHz and are included in common radio LAN working at 433 MHz additionally. Some of them can duplicate the radio link with wires connections by RS 485, using custom RS485+ stack [3]. Three of the controllers: WMCS-C, WMCS-D and WMCS-S+ have embedded infrastructure allowed them to working as GPRS clients [1].

WMCS 100 may have a different configurations and topology according to the specifics of managed objects. The object structure and the target of the controlling determine the choice of types of controllers and sensors used in the concrete realization. Each configuration of WMCS 100 includes one controller (global MASTER) that communicates with the operating station (MOS) using USB wired interface, remote operation station (ROS) and the other network controllers (SLAVES). Always “global MASTER” controller is of types WMCS-C or WMCS-D, since these have appropriate communication infrastructure:

Table 1. WMCS-C and WMCS-D embedded communication tools [3, 5]

Processor:	JN5168-001 32 bits, 32 MHz, RISC ATxMegaA8 8 bits,32 MHz, RISC
Interfaces:	RS232C/GPRS (TCP) USB 2.0 Wireless LAN(WLAN)

	Radio LAN(RLAN) RS485+ Hybrid LAN(HLAN) SPI: SPI0 - with integrated 128 KB FLASH SPI1 – internal unit SPI2 – with 64MB onboard FLASH
Timers:	3 16 bits programmable timers 24 bits programmable timer
Memory:	128 KB fast memory RAM 128 KB self-contained FLASH 64MB self-contained FLASH
Accessories:	Integrated in communication infrastructure modules (2.4 GHz) GPRS\TCP modem SIM900 Radio LAN SIM20

Exemplary configurations of WMCS 100 associated with particular applications are shown in Figures 1a, 1b and 1c. The figures show a hybrid WMCS 100 system divided into three control segments, each of which is used for a particular application. The WMCS 100 includes an operating station (MOS) to monitor and control all tasks. WMCS 100 topology performs three types of network connections. Sensors and actuators are used by controllers to influence technological processes.

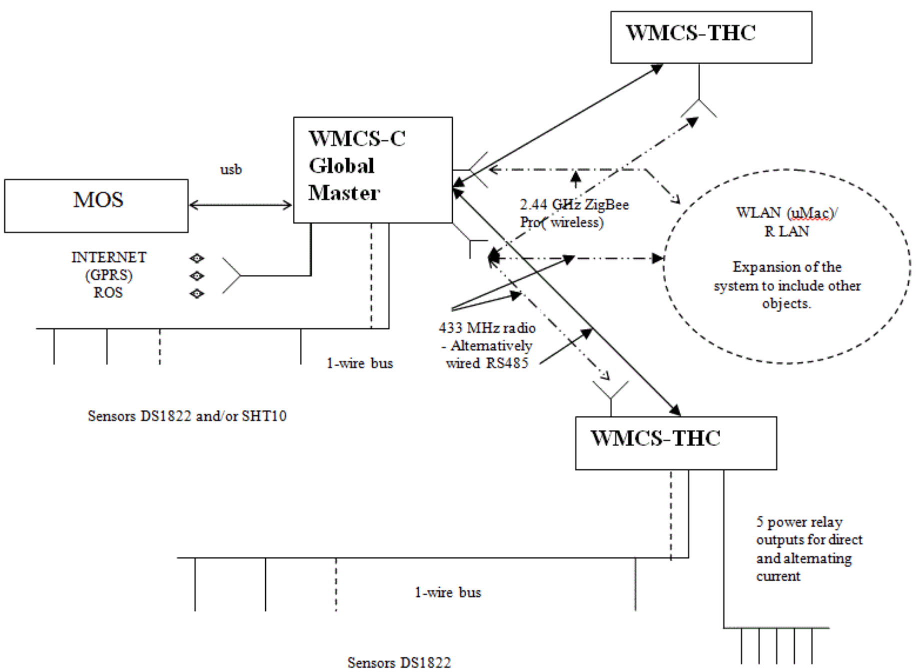


Figure 1a. Segment of WMCS 100 configured for temperature controlling [6, 7]

In Figure 1a is shown a segment of WMCS 100 designed for controlling the micro- climate in production premises including an Operator Station (MOS), a WMCS-C controller operating as a MASTER, two remote WMCS-THC controllers operating as SLAVES, sensors for measuring of the temperature and humidity in remote objects, actuators for controlling heating elements in these objects. SLAVE controllers are included in RLAN, supported by MASTER, duplicated with a wired network based on the RS 485(HLAN). The WMCS-C also supports WLAN and RLAN for communication with other objects included in external wireless LANs. It receives commands from MOS or an external computer (ROS), processes them and forms command sequences for each SLAVE or for the external networks. All controllers send their reports to MASTER, which generates a generic report and returns it to MOS or ROS.

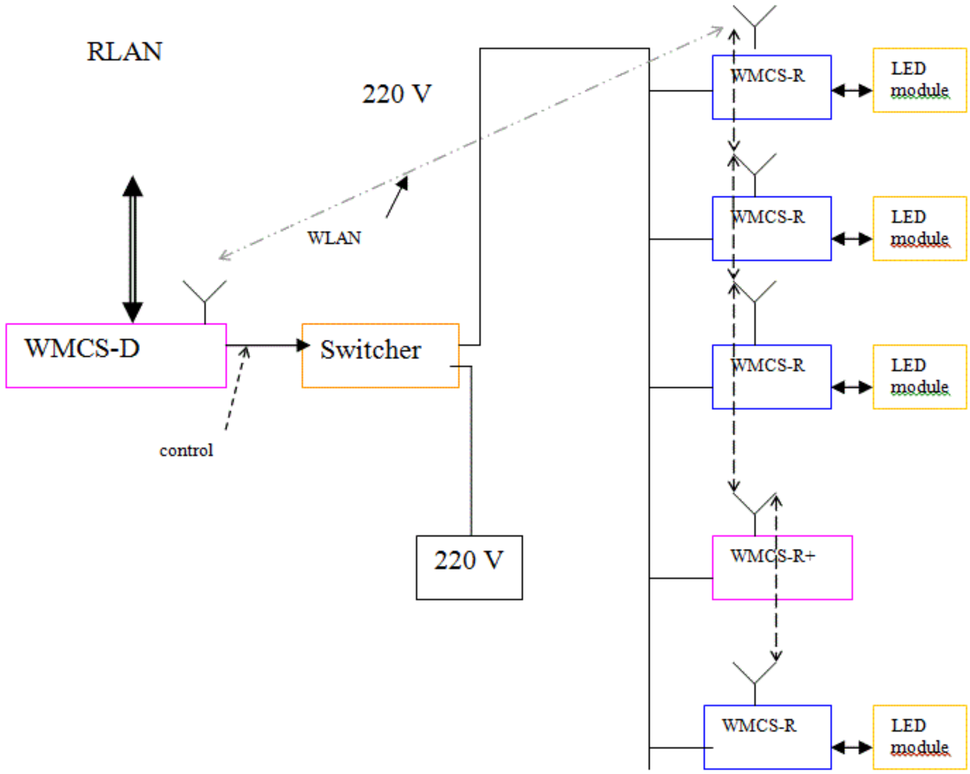


Figure 1b. Segment of WMCS 100 configured for controlling of factory lightings

In Figure 1b is shown a segment of WMCS 100 configured for controlling of factory lightings. It includes WMCS-D controller, working as SLAVE in RLAN with global MASTER

WMCS-C and as MASTER in the wireless WLAN of LED module controllers WMCS-R and WMCS-R+.

WMCS-R controllers, being nodes into WLAN, have addresses associated with LED modules, controlled and located into the different factory premises. In the case when the distance between the WMCS-R are out of range is possible to be used WMCS-R+ controllers, functioning as intermediate routers into WLAN.

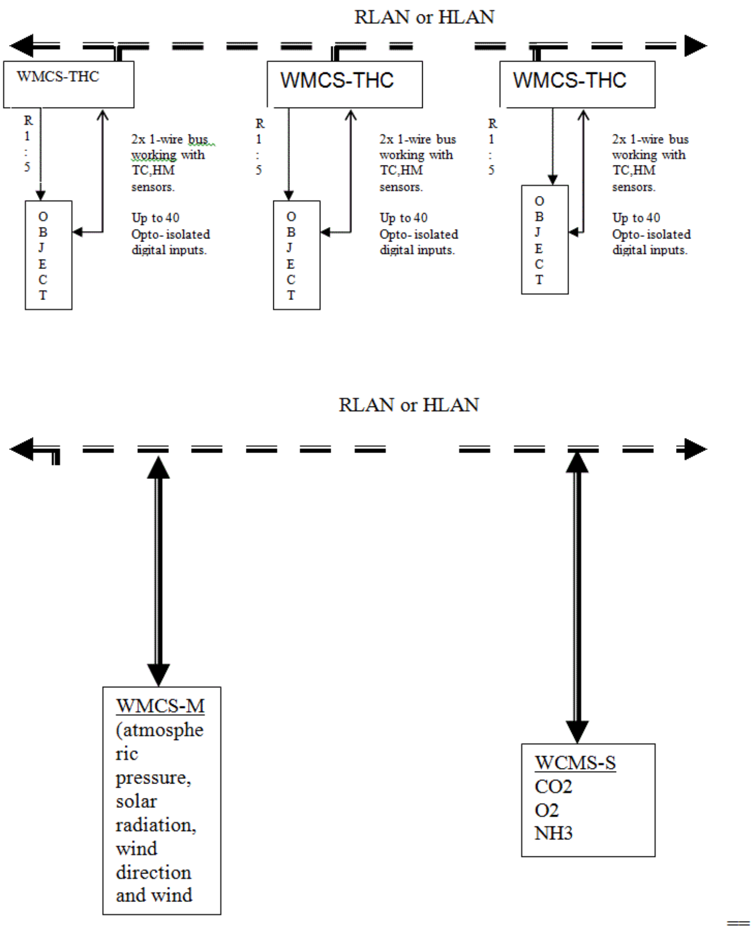


Figure 1c. Segment of WMCS 100 configured for controlling of greenhouse factory premises

In Figure 1c is shown a segment of WMCS 100 configured for controlling of a greenhouse object. It includes three controllers of type WMCS-THC for measuring of temperature and humidity parameters into the objects, detecting the states of digital inputs and controlling of the electromechanical actuators.

Two controllers, one of type WMCS-M and one of type WMCS-S monitor environmental parameters outside and gases concentration inside greenhouse objects.

All controllers are working as SLAVE nodes in the RLAN or HLAN.

3 Network and Communication Infrastructure

WMCS 100 uses various communication tools and network means to connect the controllers included in a given configuration. Some of them are wireless. These include the use of 2.44 GHz and 433 MHz radio channels. Others are wired – RS 485. There is also a hybrid interface RS485+, which allows simultaneous use of wired and wireless areas within a communication session.

There may exist one or more networks in a given configuration. WMCS 100 is creating always a wireless RLAN that contains all network controllers as network nodes. It can be implemented and as a hybrid - HLAN, according to PC485 +.

In some segments of the WMCS, WLAN built on a stack uMac can be included. In this case one of the SLAVE controllers in RLAN, acts as MASTER for the network devices of WLAN.

4 Software Solutions in WMCS 100

4.1 Operating Stations

WMCS 100 includes two types of Operating station:

- **MOS** – operating station connected to the "global MASTER" controller by USB wired interface;
- **ROS** – operating station using Internet (TCP) and GPRS for remote connecting to the "global MASTER" controller.

The operating stations are implemented in PC-Laptop under Windows and are built as applications designed in the TCL / TK environment [11, 12]. They have similar architecture and are using software tools embedded in the TCL and known as "creating and managing groups of interpreters"[11]. The interpreter is a software tool corresponded to the thread in the "multithread Operating systems".

Each interpreter is independent of others: it has its own space for commands, procedures, and global variables. The Interpreters form subordination structures: MASTER - SLAVE, similar to those between controllers in WMCS 100. Each interpreter, in its capacity of MASTER, can create interpreters that function as SLAVE (in relation to it).

A MASTER interpreter may create connections between its SLAVES and itself using a mechanism called an "alias". An alias is a command in a SLAVE interpreter which, when invoked, causes a command to be invoked in its MASTER interpreter or in another SLAVE interpreter. The only other connections between interpreters are through environment variables (the env variables), which are normally shared among all interpreters in the application.

Through the abstraction of the subordinate interpreters, it is possible to create a model into the Operating station corresponding to the concrete structure of WMSC 100. The program code in Operating station for controlling of given controller is executing from corresponded interpreter.

This is possible because each interpreter has the ability to execute all commands of the primary language, with exceptions of the limitations imposed on him by its creator.

The highest level in interpreter hierarchy has this one corresponds to "global MASTER" controller. The program code executed from it controls the communication with connected to the Operating station "WMSC 100" network, forms, sends and receives message sequences to and from the controllers, selects and activates the low level interpreters, corresponded to other controllers.

4.2 Principles of Addressing of WMCS 100 Controllers

Since the VMC 100 supports three types of networks, the controllers included in them are addressed in different ways:

- **RLAN** unites and connects all controllers included in the WMCS. The latter have unique addresses arranged in a growing order. During the initialization time, for each controller operating as a MASTER is allowing the access to the associated SLAVE controllers connected to it. When a message is sending by any MASTER it is receiving from all its SLAVEs. The sent message shall contain information about the final recipient and one for the next intermediate. All controllers whose addresses are different from the intermediate do not accept the message. The one whose address matches the current intermediate accepts it, analyzes whether its own address is not the end of the message, and if not, modifies the intermediate address to provide access to the next intermediate controller.

Upon reaching the message to the final address, the latter controller interprets the message and forms a report that goes back to the "Global MASTER" [3].

- **HLAN** follows the RS 485+ protocol described in the draft BPMIO[3].
- **WLAN** uses protocol/stack uMAC [1].

The preferred way for addressees in the SMP allows access to remote nodes at the expense of network traffic.

4.3 Command Languages and Drivers

Each of the controllers has built-in driver programs relevant to its functionality. They are activating when the controller is receiving messages. After receiving they form relevant reports.

Command messages are generated by the active interpreters into the operating station and include an address and command strings. After sending them to the global MASTER, the latter has the task of forwarding them to the corresponding controller address.

The reports are sent back to the active interpreter, which on their basis generates visualization for the operator of the Operating Station. In this way, the user program and the visualization of the results are determined by the program code executed in the corresponding interpreter and the one in the interpreter from the highest hierarchical level.

5 Conclusion

The paper presents an approach to design a distributed power management system in a small company. A set of wired and wireless devices are used to construct different system configurations to match the features of the managed object. The technology used allows the construction of a centralized system with an operating station within the site or a remote mobile one.

The software and hardware used by the system provide a reliable connection between the different networks and their controllers, regardless of the distance between network nodes.

References

[1]. D. Batchvarov, A. Boneva, Z. Ilcheva, S. Angelov , V. Ivanova, Tools for control of mechatronic objects using the wireless network stack uMAC, *Proceedings for International Conference AUTOMATICS AND INFORMATICS'2017* 4-6 October 2017, 2. Session ROBOTICS AND MECHATRONICS, ISSN 1313-1850, CD: ISSN 1313-1869, JOHN ATANASOFF SOCIETY OF AUTOMATICS AND INFORMATICS, ,Sofia, Bulgaria, 2017, pp. 77 – 80.

[2]. Bachvarov D., Boneva A., Angelov S. "SYSTEM FOR MONITORING AND DOCUMENTATION OF TEMPERATURE AND HUMIDITY IN THE MANUFACTURING FACILITIES OF A COMPANY",*Proceedings: International Conference "ROBOTICS, AUTOMATION AND MECHATRONICS ' 16" (RAM 2016),* ISSN 1314 - 4634 , Sofija, Prof. Marin Drinov", pp.5-14.

[3] Project BPMIO. www.bpmio.free.bg (last visited 31.10.2017).

[4]. Lu, X. , Wang, S. , Li, W. , Jiang, P. and Zhang, C. (2015) 'Development of a WSN based real time energy monitoring platform for industrial applications', *in Proceedings of the 2015 IEEE 19th International Conference on Computer Supported Cooperative Work in Design*, CSCWD 2015, ISBN 978-147992002-0, (pp: 337-342). IEEE http://dx.doi.org/10.1109/CSCWD.2015.7230982

[5]. NXP, *Data Sheet*: JN516x - IEEE802.15.4 Wireless Microcontroller, JN-DS-JN516x v1.3 Production, NXP Laboratories UK 2013, pp. 1-94, http://www.nxp.com/

[6]. Maxim Integrated, DS1822 Econo 1-Wire Digital Thermometer, Maxim Integrated Products, 2015, pp. 1-21, http://datasheets.maximintegrated.com/en/ds/DS1822.pdf

[7]. SENSIRION, *Datasheet* SHT1x (SHT10, SHT11, SHT15) Humidity and Temperature Sensor IC, Version 5 – December 2011, pp. 1-12, www.sensirion.com

[8]. ATMEL Corporation, ATXMEGA32A4 *Datasheet*, Atmel Corporation. 2008, pp. 1-64 http://pdf1.alldatasheet.com/datasheet-pdf/view/236665/ATMEL/ATXMEGA32A4.html

[9]. Dallas Semiconductor Corp., Automatic Identification Data Book, Dallas, Texas, pp. 117-136, 1995

[10]. Hipp R., Mktclapp A Toll For Mixing C/C++ with Tcl/Tk, Charlotte, NC, 1999.

[11]. http://www.tcl.tk/software/tcltk/ (last visited 31.10.2017).

[12]. https://www.tcl.tk/ (last visited 31.10.2017).

A Survey in Deep Convolutional Neural Networks for Diagnosis of Diseases

[1]Hidir Selcuk NOGAY, [2]Tahir Cetin AKINCI, [3]Emel ONAL

[1] Erciyes University, Mustafa Cikrikcioglu Voctional College
Kayseri, Turkey
[2, 3] Istabul Technical Universtiy, Department of Electrical Engineering
Istanbul, Turkey
nogay@erciyes.edu.tr, akncitc@itu.edu.tr, eonal@itu.edu.tr

Abstract: Before designing a deep convolutional neural network (DCNN), it is necessary to preliminarily investigate the decisions that need to be made about design processes. When designing a DCNN for medical diagnosis, specific information about the disease itself is needed. In this study, the stages of designing a DCNN model that can be used to diagnose diseases have been investigated and information affecting decision-making processes at each stage has been identified.

Keywords: Convolutional Neural Networks, Deep Learning, Artificial Intelligence.

1 Introduction

Developments in computer technology and artificial intelligence methods have become one of the important tools to reach greater success in the medical field. In recent years, machine learning-based deep learning approaches have been used to start the treatment process in a short time with the definite diagnosis of diseases.

Machine learning is the general name of computer algorithms that model a given problem according to the data obtained from the probing environment. Today, machine learning algorithms are used in many areas. Search engines, sharing sites, e-mail services use this learning algorithm. In addition, machine learning is used in the diagnosis and treatment of diseases using patient records, gene and DNA sequencing [1].

Deep machine learning is a kind of artificial intelligence method with more data sets and more layers. The Deep Convolutional Neural Network (DCNN) is one of the most successful machine learning methods used in the deep learning approach. In order to design the DCNN model appropriately for its purpose and to be successful. For the DCNN architecture to be designed and successful for its purpose, the activation function, gradient descent algorithm, the number of convolutional layers, the number of dropout layers and the number of neurons in the output layer need to be determined. While these decisions are being made, detailed investigations are required to be carried out as required. In this study, specific information was provided in the DCNN design process used to predict diseases. It has been researched and determined which activation function and which data set is most suitable to use. In addition, it is also mentioned about the layers of convolution, dropout and output layers.

3 Deep Machine Learning

It is seen that the deep machine learning algorithm has superior performance when used with image processing techniques, classification and estimation approaches depending on the data set [1-3].

As shown in Figure 1, there is more than one layer in deep learning. Each layer is subject to feature presentation through a learning method. The purpose of deep machine learning; to create higher-level representations of the data by using multi-layered nonlinear models [4]. On each layer, a hierarchy is created from feature down to top-level features with feature presentations. Deep machine learning can only be applied to certain neural network architects. There are 13 types of deep machine learning architecture in the literature and applications. The most successful and most advanced architecture in image processing is the "Deep Convolutional Neural Network, (DCNN) [5]. Deep convolutive neural networks can be in different architectures depending on the number of layers [6].

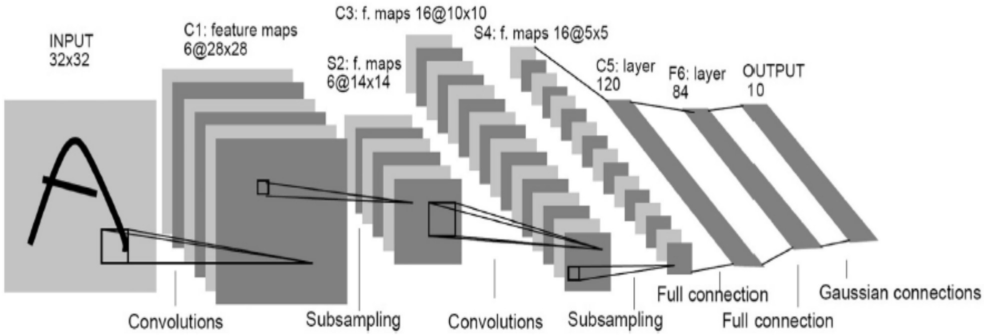


Figure 1. A sample DCNN architecture [5]

4 Convolutional Neural Networks

Convolutional neural networks are initially a type of feedforward neural network proposed [5]. Feed forward neural networks can be examined in two groups as full connected networks and sparse networks. Convolutional neural networks are a kind of structurally sparsely connected networks.

The number of parameters can be reduced by the weights shared between the receiver fields of each neuron in a sparse layer. In this case it will be the same as the weights at the same position between the receiver fields. Since the same weights are applied through the input nodes in the same configuration, the sparse layer can be thought of as a convolute filter with a certain shape. In the convoluted neural network of Figure 2, the same color of the connections indicates that their weights are equally distributed. Since this process is a convolutional process, this network is called the convolutional neural network. In this case, the middle sparse layer in Figure 2 is the convolutional layer [7,8].

Supervised learning method is used in convolutional neural networks. The neurons are connected to the next hidden layer, depending on their relative location in the neural network. When an image is presented for input into the neural network, each neuron in the first hidden layer receives input vectors (small images) from the small regions of the image. In addition, neurons are replicated so that the same features can be detected in different parts of the input image [8,9].

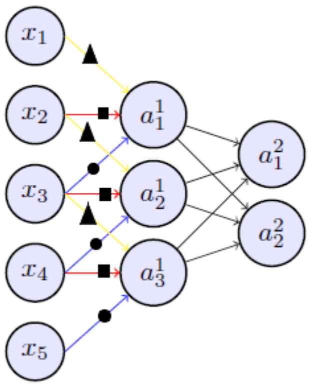


Figure 2. Example of convolutional neural network [8]

3. 1 Convolutional Layer

This type of layer takes as input the image to be classified and applies a convolution filter set to this input. The output of the layer corresponds to the feature maps of the convolution filter. The output of the convolutional layer depends on the convolution filter and the bias. Parameters learned during training correspond to convolution filter and bias weights [7,10].

The convolutional layer reduces the number of parameters that the model has to learn in large scale, and it also prevents the model's overfitting. It also reduces calculation times. For example, in Figure 2 only 9 weights are calculated instead of 21 weights. Also, each layer can be designed to learn multiple filters [8].

3.2 Pooling Layer (Subsampling)

In general, pooling is a form of dimensional contraction used in convolutional neural networks. The aim is to remove unnecessary information and to protect only the most critical information [11]. In convolutional neural networks, a connection is made to the pooling layer after the convolutional layer. There are two options for the pooling layer. The first is average pooling and the other is max-pooling. To create the output in the average pooling layer, the averages of the input elements within a given window are taken. In the maximum pooling layer, the largest input of the input elements in each input window is taken as output. With the pooling layer, the size of the feature map input is reduced. Thus, local spatial invariance is achieved. The pooling layer divides the input maps into sub-regions that do not overlap each other, and each performs a specific pooling function [10].

3.3 Dropout/Drop Connect

Dropout prevents the test data from being excessive. Dropout work; is the random skipping of half of the neurons in every training of the neural network. The dropout process prevents the network from overfitting [12].

3.4 Output Layer

This layer type catches the relationship between final layer feature maps and class labels. The output of the output layer is a vector of elements, each representing a numerical score of a class. The output layers can be seen as convolution operations where the filters have the same dimensions as the input maps [10].

3.5 Gradient Descent

The most common method to minimize the total error is the gradient descent algorithm. According to this algorithm, the minimum point of the function between the first value and the last value is tried to be found in the data set. The gradient descent algorithm can be applied in three ways.

The first is the batch gradient descent (BGD) algorithm. In the BGD algorithm, all m number of samples are used in each iteration. Equation (1) represents the hypothesis function, and equation (2) represents the cost function for neural network training [13].

$$h_\theta(x) = \sum_{j=0}^{n} {}_0\theta_j \, x_j \tag{1}$$

$$J_{train}(\theta) = \frac{1}{2}\sum_{i=0}^{m}\left(h_\theta(x^{(i)} - y^{(i)}\right)^2 \quad (2)$$

The second is the stochastic gradient descent (SGD) algorithm. SGD algorithm is preferred in DCNN models especially for diagnosis of diseases. In the SGD algorithm, only one sample is used for each iteration. In SGD the training data is randomly mixed [13].

The third is the mini-batch gradient descent (MBGD) algorithm. In MBGD algorithm, fewer samples are used for each iteration [13].

3.6 Activation Functions

The most important point in the design of the neural network architecture is the selection of the activation function [14]. Activation functions determine how inputs must be transformed across the network. Thus the network gains the ability to learn complex functions. A completely linear activation function can be selected for each layer in a neural network. However, in this case the outputs are simply a linear transformation of the inputs. In this case, the network cannot have the ability to learn complex functions. Therefore, nonlinear activation functions are preferred in deep machine learning networks [7].

Sigmoid activation functions are frequently used daily in artificial neural networks. Sigmoidal activation functions include sigmoid and hyperbolic tangent (tanh) functions. Sigmoidal functions are limited to minimum and maximum values. The limits of the sigmoid activation function cause saturated neurons in the upper layers of the neural network. These saturated neurons cause a gradient descent to the lower layers of the neural network and limit the depth of the neural network that can be trained. Therefore, they are usually not preferred in deep neural networks. The sigmoid activation function (3) and the hyperbolic tangent (tanh) activation function (4) are shown below [7,9].

$$sig(x) = \frac{1}{1+e^{-x}} \quad (3)$$

$$tanh(x) = \frac{e^{2x}-1}{e^{2x}+1} \quad (4)$$

Another commonly used activation function is the Rectified Linear Unit (or ReLU) activation function. In the ReLU function, if the input value is less than zero, the value is zero [15]. The positive linear (RLU) activation function is therefore less affected by the decreasing gradient descent. Equation 5 shows ReLu [7-9].

$$ReLU(x) = Max(0,x) \quad (5)$$

Recently, a special kind of activation function has been developed to complete the drop-out. A dropout layer randomly skips half of the activation. Maxout is an activation function specially created to take advantage of the dropout feature. Maxout does not have the rarity property and does not have a constant zero like ReLU. One feature of dropout, however, is that it causes a dilution in the distribution of activations. This is why MaxOut does not have to be infrequent [9]. The Maxout activation function is designed so that it does not have a zero gradient value. The Maxout activation function allows the network to obtain a good approximation of the model average using Maxout and Dropout. The Maxout activation function works by multiplying the input vector by the matrix giving a vector. The maximum value is selected from this vector. Only the link weight is affected from the maximum value in back propagation. Equations 6 and 7 show the maxout activation function [9].

$$Maxout_i(x) = \max_{j \in 1,k]} z_{ij} \tag{6}$$

$$z_{i,j}(x) = x^T W_{\cdots i,j} + b_{i,j} \tag{7}$$

$$W \in \Re^{dxmxk} \text{ and } b \in \Re^{bxk} \tag{8}$$

The maxout activation function can be used in deep neural networks as it is less affected by the decreasing gradient descent. Figure 3 shows all the activation functions [7,8].

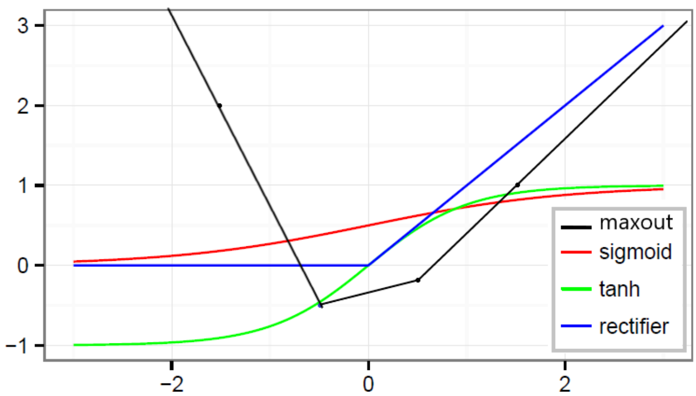


Figure 3. Activation functions [7]

5 Conclusion

As a result, the number of convolution layers in a DCNN design for the diagnosis of diseases varies depending on the number of data in the data set in which the model is trained. Likewise, the number of pooling layers can be selected as two layers, depending on the number of data. Following the training of the DCNN model, a method of skipping half of the whole data in the selection of each test data at the time of testing contributes to a more accurate prediction of the model. Dropout also prevents the memorization of the data. Therefore, the use of dropout in the test phase is suitable for every purpose. The last filtering of the DCNN model is done at the output layer. The output layer is the final convolution layer. In the DCNN model, one of three gradient descent algorithms is used to minimize the error. In general, since the diagnosis of diseases is concerned, the stochastic gradient descent (SGD) algorithm is preferred from these three gradient descent algorithms. Finally, the most important point in designing a DCNN architecture for diagnosing diseases is the selection of the activation function. The most important criterion in selecting the activation function is that a function that is not affected by the gradient descent is preferred. For this, the most preferred activation function is the Rectified Linear Unit (ReLU) activation function.

References

[1] Lemley, J., Bazrafkan, S., and Corcoran, P. (2017) Deep Learning for Consumer Devices and Services, *IEEE Consumer Electronics Magazine*, DOI:10.1109/MCE.2016.2640698, Date of publication: 15 March 2017, 2162-2248.

[2] Zeiler, M., Fergus., R. (2013) Regularization of Neural Networks using DropConnect, access date: 12th Nov. 2017, http://cs.nyu.edu/~wanli/dropc/

[3] Litjens, G., et al., (2017) A survey on deep learning in medical image analysis, *Medical Image Analysis*, 42, 60–88.

[4] Bengio, Y., Courville, A., and Vincent, P. (2013) Representation learning: A review and new perspectives, *Analysis IEEE Transactions on Pattern and Machine Intelligence*, 35,8, 1798-1828.

[5] Lecun Y., Bottou L., Bengio Y., Haffner P. (1998) *Gradient-based learning applied to document recognition*, Proceedings of the IEEE, 86, 2278–2324.

[6] Cecotti, H., (2017) Convolutional neural networks for event-related potential detection: impact of the architecture, *39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2031-2034.

[7] Doorn, J.V. (2014) Analysis of Deep Convolutional Neural Network Architectures, 21th Twente Student Conference on IT June 23st.

[8] Walker, I. (2015) Deep Convolutional Neural Networks for Brain Computer Interface using Motor Imagery (Master Thesis) Imperial College London, Computing Science Department of Computing Imperial College of Science, Technology and Medicine, U.K, London.

[9] Goodfellow, I. J., Warde-farley, D., Courville. A. (2013) Maxout Networks, *Proceedings of the 30th International Conference on Machine Learning*, Atlanta, Georgia, USA, 2013. JMLR: W&CP, 28.

[10] Hassan H. H., (2017) Classifications of Multispectral Colorectal Cancer Tissues Using Convolution Neural Network, *Journal of Pathology Informatics*, doi: 10.4103/jpi.jpi_47_16.

[11] Boureau, Y.L., Ponce, J., LeCun, Y. (2010). A theoretical analysis of feature pooling in visual recognition, *In Proceedings of the 27th International Conference on Machine Learning* (ICML-10), 111-118.

[12] Hinton G. E., et al. (2012), Improving neural networks by preventing co-adaptation of feature detectors. arXiv preprint arXiv:1207.0580. access date: 12th Nov. 2017,

[13] https://arxiv.org/pdf/1207.0580.pdf

[14] Andrew, N. (2017) Stanford University Lecture 10. access date: 12th Nov. 2017, https:///www.coursera.org/learn/machinelearning/lecture/8SpIM/about

[15] University of Montreal LISA Lab, *Deep learning tutorials*, access date: 12th Nov. 2017, http://deeplearning.net/tutorial/

[16] Glorot, X., Bordes, A. Bengio Y. (2011) Deep sparse rectfier networks. In Proceedings of the 14th *International Conference on Artificial Intelligence and Statistics*, JMLR W&CP,15, 315-323.

Dual MLP for Time Series Forecasting with Hidden Layer Sharing

Ivan I. Blagoev, Tatiana V. Atanasova, Iliyan A. Zankinski

Institute of Information and Communication Technologies – Bulgarian Academy of Sciences

acad. Georgi Bonchev Str., block 2, 1113 Sofia, Bulgaria

i.blagoev@iit.bas.bg, atanasova@iit.bas.bg, iliyan@hsi.iccs.bas.bg

Abstract: Modern management and control systems in the fields of business, telecommunications and energy require short term planning on daily basis due to great dynamics, changing regulations and penetration of new technologies and new markets. Because of this, short-term forecasting is a daily obligation and need. Despite numerous studies, forecasting tasks in the mentioned fields are still challenging due to the high complexity of these systems. Recent advances in the development of mathematical data retrieval methods and artificial intelligence gives a potential for improving the accuracy of forecasting. With modern tendencies the tasks of detecting regularities (trends, processes) in time series become increasingly important.

Research approaches to exploring time series can be divided into two categories: statistical methods and computational intelligence. Statistical methods investigate dependencies between baseline and relevant factors after learning on past data while the other group of methods mimics the human way of thinking and logical conclusion to gain knowledge from past experience (as artificial neural networks) and anticipate future values.

This paper focuses on the capabilities of regular multilayer perceptrons (MLPs) to work in dual mode by sharing of a common hidden layer. This hybrid structure of a neural network is used for financial time series forecasting.

Keywords: neural networks, time series forecasting, multilayer perceptron.

1 Introduction

Big data in general are unstructured. Unstructured data are data that does not follow a specified format [1]. The analysis of unstructured data is able to reveal important interrelations that were previously difficult or impossible to determine. In order to be efficiently handled, the unstructured data needs to be structured. Artificial Neural Networks (ANNs) are a promising approach for unstructured data handling. Nowadays, ANNs are becoming one of the most popular tools for processing big data.

For different application areas and datasets, it is not possible to apply one model or algorithm. New machine learning algorithms are being developed intensively as their number grows. For a specific application, it is necessary iteratively to try different alternatives and approaches to discover and create the best model.

Financial time series is typical example of unstructured data. The use of neural networks in finance is based on one fundamental assumption - the replacement of prediction by recognition. A neural network does not predict the future, it "tries to learn" in the current state of the market the previously encountered situation and to reproduce the market reaction as accurately as possible.

For forecasting financial time series, it is possible to use multilayer perceptrons (MLP) [2]. The use of networks with feedback in this case is inexpedient due to the difficulty of training such networks and the irrelevance of the main characteristic of networks with feedbacks - short-term memory [3]. ANNs provide a powerful black box method for modeling time dependent phenomena because they require no prior assumptions or knowledge of the underlying dynamics of the time series data as with the Kalman filter [4]. The Kalman filter does not have enough features to predict prices and capture rapid movement.

One would have to build a more complex model [5], but in that case one needs to estimate the state matrix, i.e. one needs to understand the dynamics between future values and past historical data.

It is necessary to determine which inputs to feed the network and what data to receive at the output. Input data are often time-series (1-D time series), spatially correlated (2-D images) or can be transformed to obtain such properties.

Training the model with the selected machine learning algorithm is done by one part of the data and the other part is used for validation. Validation of the model is a key factor for successful solution of a problem. The model improvements are achieved by changing of variables, formulas or the overall algorithm.

Multivariate time series often do not show obvious relationships between variables as they describe different phenomena. The choice of appropriate input variables and their grouping is in the field of Input Variable Selection (IVS).

2 Model Proposition

Classical time series definition includes a set of observations recorded at the time t [6]:

$$y(t) = S(t) + T(t) + E(t)$$

Where *y(t)* is the data in the period *t*, *S(t)* is the seasonal component at period *t, T(t)* is the trend-cycle component at period *t* and *E(t)* is the remainder (or irregular, or error) component at period *t*.

The features of time series are as follows:

1. Often the basic model is not known in advance;
2. Time series can be highly non-linear or even chaotic;
3. Time series data can be severely nosy;
4. The sample size of the data can be insufficient.

With so many difficulties in the time series forecasting every innovation in the field would be great appreciated [7].

Multilayer perceptron is the most used type of ANN. By merging the hidden layer of two MLPs a modified neural network is produced, which can be called Dual MLP. The hidden layer sharing can be efficiently used for knowledge transfer between both MPLs. This modification allows each MPL to speed up its training because of the extra information provided by the other MPL. Such network topology can be very useful in problems like time series forecasting where different financial instruments can be provided at the inputs of the Dual MLP. The lack of extensive research for such ANN topologies makes this field interesting and it gives the focus of the current paper.

The model proposed in this paper combines two MPLs by sharing of the hidden layer (Figure 1). A number of studies emphasize that neural network performance is highly dependent on input data or data representation in hidden layers [8, 9].

Dual MPL is trained with financial time series from the foreign exchange market (FOREX).

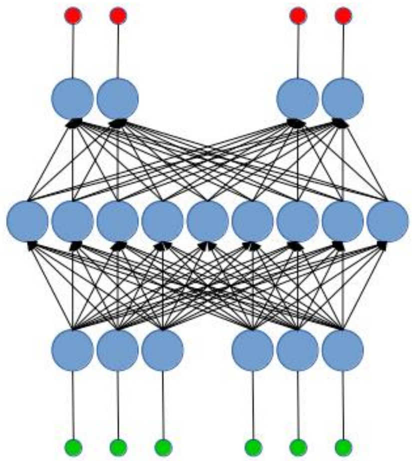


Figure 1. Dual MLP with 3-3-9-2-2 topology

In our case two currency pairs are used (EUR/USD and USD/JPY). The main idea is to exploit possible correlation between European, Japanese and US economies (Figure 2 and Figure 3).

The specificity of the research object imposes certain features on the use of neural networks for data analysis. It should be noted that one of the important components of data analysis with the help of neural networks is preprocessing of data aimed at reducing the dimension of network inputs, increasing the joint entropy of input variables and normalizing input and output data.

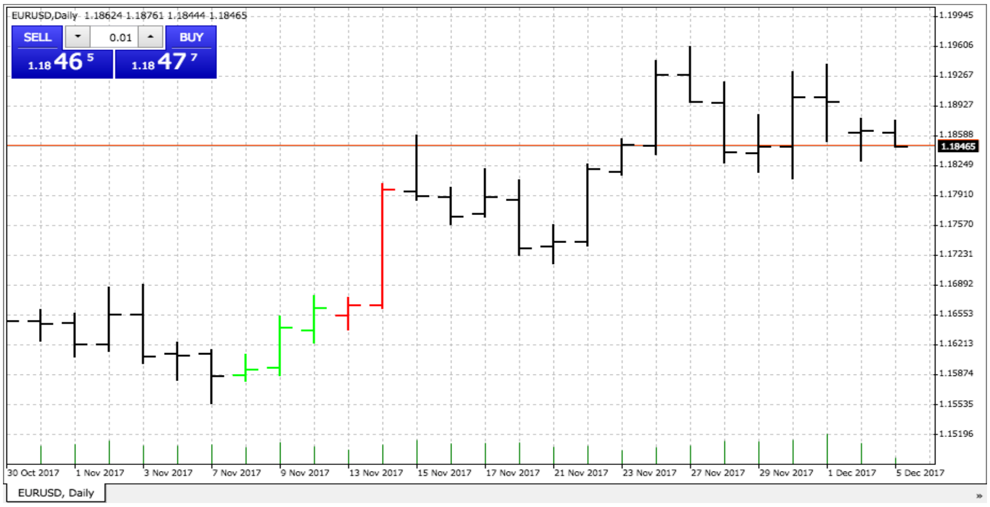


Figure 2. Daily rate of EUR/USD

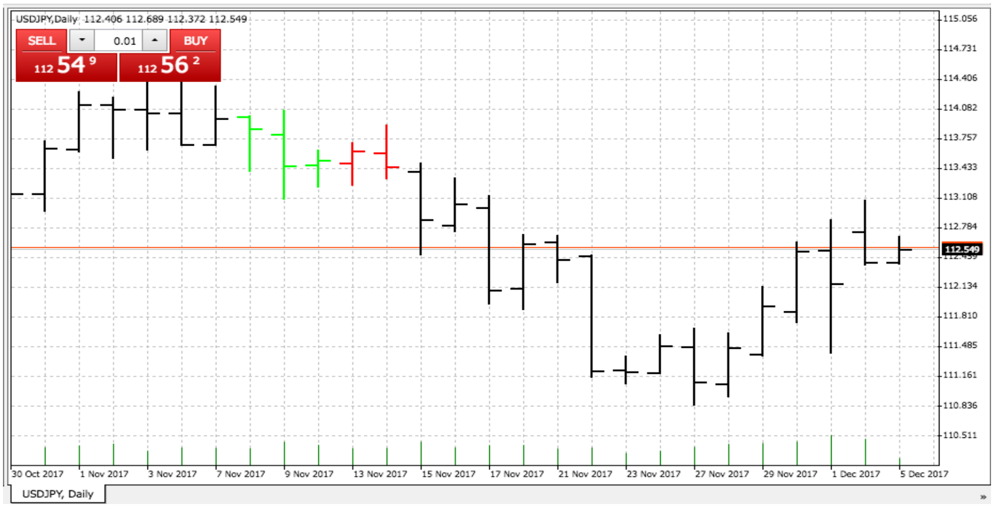


Figure 3. Daily rate of USD/JPY

Left part of Dual MLP is feed with scaled values of EUR/USD data (first three inputs) when right part of the Dual MLP is feed with scaled values of USD/JPY data (second three inputs). Scaling is needed because MLPs are well working with input signals in the range 0.0 to 1.0 when the activation function is sigmoid and -1.0 to +1.0 when the activation function is hyperbolic tangent. The other reason of scaling is related with the currency pairs ration difference. For example the EUR/USD ratio is about 1.18 in Figure 2, when the USD/JPY ratio is about 112.58 in Figure 3. Without scaling input signals would influence internal ANN neurons with different magnitude which will make the system less accurate. The expected forecast is taken from the MLPs output and it is re-scaled to the original time series data range. During Dual MLP training time series are conditionally divided in two sets lag (past value marked with green in the figures) and lead (future values marked with red in the figures). The same color marking is used in ANN topology figure.

3 Conclusions

Information systems in telecommunications, energy, business and many other areas are gathering more and more data which need processing and support. Artificial neural networks can be an effective tool for analyzing time series and big data in those areas. The proposed model is a promising extension of the neural networks topologies.

Acknowledgements

This work was partially supported by private funding of Velbazhd Software LLC.

References

[1] Silipo R. and Phil Winters, „Big Data, Smart Energy, and Predictive Analytics - Time Series Prediction of Smart Energy Data”, KNIME.com AG, 2013.

[2] Oancea B., Șt. Cr. Ciucu, „Time Series Forecasting Using Neural Networks”, Challenges of the Knowledge Society, *IT in Social Sciences*, pp.1402-1408, 2013

[3] Balabanov, T., Zankinski, I., Barova, M., Strategy for Individuals Distribution by Incident Nodes Participation in Star Topology of Distributed Evolutionary Algorithms, *Cybernetics and Information Technologies*, Vol.16, 1, 2016.

[4] Alexandrov, A. Ad-hoc Kalman Filter Based Fusion Algorithm for Real Time Wireless Sensor Data Integration. Advances in Intelligent Systems and Computing, vol. 400: Flexible Query Answering Systems, *Springer*, 2015.

[5] Tashev T. The Modeling of Data Link Layer Receiver with Generalized Net - a Lifetime Value Restriction, Proc. 3-th Int. Workshop on Generalized Nets, 2002, Sofia, pp.44-45.

[6] Box G., Jenkins G., and G. Reinsel. Time series analysis: forecasting and control, volume 734. *Wiley*, 2011.

[7] Atanasova T., Barova M., Balabanov T., Using of neural model for analyzing of time series in large volumes of data. NVU "Vasil Levski", Bulgaria, pp.193-198, 2016.

[8] Tomov, P., Monov, V. Artificial neural networks and differential evolution used for time series forecasting in distributed environment. Proc. of the Int. Conf. "Automatics and informatics 2016", 4-5 October 2016, Sofia, Bulgaria, pp.129-132, 2016.

[9] Armando Bernal, Sam Fok, Rohit Pidaparthi, Financial Market Time Series Prediction with Recurrent Neural Networks, 2012.

Petri Nets in Modelling of Supervisor Based Agent Cooperation

František Čapkovič[*], Lyubka Doukovska[**], Vassia Atanassova[***]

[*]Institute of Informatics, Slovak Academy of Sciences

Dúbravská cesta 9, 845 07 Bratislava, Slovakia

Frantisek.Capkovic@savba.sk

[**]Institute of Information and Communication Technologies, Bulgarian Academy of Sciences

Acad. G. Bonchev Str., Block 2, 1113 Sofia, Bulgaria

doukovska@iit.bas.bg

[***]Institute of Institute of Biophysics and Biomedical Engineering, Bulgarian Academy of Sciences

Acad. G. Bonchev Str., Block 105, 1113 Sofia, Bulgaria

vassia.atanassova@biomed.bas.bg

Abstract: An approach to modelling and control synthesis of the agent cooperation in multi agent systems (MAS) based on place/transition Petri nets (P/T PN) is presented. It is based on utilizing the P/T PN place invariants and Parikh's vector. The autonomous agents are forced by a supervisor to mutually cooperate in order to achieve a prescribed goal. The approach is illustrated in details on an example.

Keywords: Agent, modelling, control synthesis, multi agent system, place/transition Petri nets.

1 Introduction

An agent is a real or virtual entity that exhibits an autonomous behavior. Multi agent system (MAS) is a group of cooperating agents. The agent cooperation can be understood to be a kind of discrete event systems (DES). The causality in DES behavior was described in [1]. P/T PN are, for their user friendliness, frequently used for DES modelling and for the control synthesis for them. The basic definition and mathematical model of P/T PN was described in details in [2]. Hence, the analytical model of P/T PN has the form of the following linear discrete integer system

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{B}.\mathbf{u}_k, \quad k = 0, 1, \cdots, N \tag{1}$$

$$\mathbf{B} = \mathbf{G}^T - \mathbf{F} \tag{2}$$

$$\mathbf{F}.\mathbf{u}_k \leq \mathbf{x}_k \tag{3}$$

where

k is the discrete step of the dynamics development;

$\mathbf{x}_k = (\sigma_{p_1}^k, \cdots, \sigma_{p_n}^k)^T$ is the *n*-dimensional state vector; $\sigma_{p_i}^k \in \{0, 1, \cdots, \infty\}, i = 1, \cdots, n$ express the states of atomic activities of the PN places by 0 (passivity) or by $0 < \sigma_{p_i} \leq \infty$ (activities);

$\mathbf{u}_k = (\gamma_{t_1}^k, \cdots, \gamma_{t_m}^k)^T$ is the *m*-dimensional control vector;

$\gamma_{t_j}^k \in \{0,1\}, j = 1, \cdots, m$ represent occurring of the elementary discrete events (e.g. starting or ending the activities, failures, etc.) by 1 (presence of the corresponding discrete event) or by 0 (absence of the event);

B, F, G are matrices of integers being the structural parameters.

2 Place Invariants and Parikh's Vector in the Supervisor Synthesis

The P/T PN place invariants are defined as the state vectors reachable from the initial state fulfilling the relation

$$\mathbf{w}^T.\mathbf{x}_k = \mathbf{w}^T.\mathbf{x}_0 \tag{4}$$

As it was presented in [3], the place invariants are important at the control synthesis of (1). When restrictions imposed on the state vectors have the vector-matrix form $\mathbf{L}.\mathbf{x}_k \leq \mathbf{b}$, the supervisor structure and initial state are as follows

$$\mathbf{B}_s = -\mathbf{L}.\mathbf{B} \tag{5}$$

$$\mathbf{B}_s = \mathbf{G}_s^T - \mathbf{F}_s \tag{6}$$

$${}^s\mathbf{x}_0 = \mathbf{b} - \mathbf{L}.\mathbf{x}_0 \tag{7}$$

Other important vector characterizing the P/T PN behaviour is the Parikh's. To define the Parikh's vector let us develop the model (1) behaviour as follows

$$\mathbf{x}_k = \mathbf{x}_{k-1} + \mathbf{B}.\mathbf{u}_{k-1} = \mathbf{x}_{k-2} + \mathbf{B}.(\mathbf{u}_{k-1} + \mathbf{u}_{k-2}) = \ldots = \mathbf{x}_0 + \mathbf{B}.(\mathbf{u}_{k-1} + \mathbf{u}_{k-2} + \ldots + \mathbf{u}_0) \tag{8}$$

Here, $\mathbf{v} = (\mathbf{u}_{k-1} + \mathbf{u}_{k-2} + \ldots + \mathbf{u}_0)$ is named to be the Parikh's vector of the system (1). It is very important vector because its entries give us information how many times the particular transitions are fired during passing the system (1) form the initial state $\mathbf{x}_0$ to the terminal state $\mathbf{x}_k$.

When the restrictions are imposed also on control vector and on Parikh's vector, i.e.

$$\mathbf{L}_p.\mathbf{x} + \mathbf{L}_t.\mathbf{u} + \mathbf{L}_v.\mathbf{v} \leq \mathbf{b} \tag{9}$$

the structure of the supervisor and its initial state are the following

$$\mathbf{F}_s = \max(\mathbf{0}, \mathbf{L}_p.\mathbf{B} + \mathbf{L}_v, \mathbf{L}_t)$$
$$\mathbf{G}_s^T = \max(\mathbf{0}, \mathbf{L}_t - \max(\mathbf{0}, \mathbf{L}_p.\mathbf{B} + \mathbf{L}_v)) - \min(\mathbf{0}, \mathbf{L}_p.\mathbf{B} + \mathbf{L}_v) \quad (10)$$
$${}^s\mathbf{x}_0 = \mathbf{b} - \mathbf{L}_p\mathbf{x}_0 - \mathbf{L}_v\mathbf{v}_0$$

The approach is illustrated by means of two examples appertaining to synthesizing the supervisor.

3 Illustrative Example 1

Consider two agents A_1, A_2 with the same structure given in Figure 3. Both agents are modelled by PN subnets, i.e. by PN_1, PN_2, respectively. Here, PN_1 is created by $p_1, \dots, p_4$; $t_1, \dots, t_3$ while PN_2 is created by $p_5, \dots, p_8$, $t_4, \dots, t_6$. They have the same structure and the same initial state.

$$\mathbf{F}_i = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 2 & 0 \end{pmatrix}; \quad \mathbf{G}_i^T = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 2 \end{pmatrix}; \quad \mathbf{B}_i = \begin{pmatrix} -2 & 1 & 1 \\ 1 & -1 & 0 \\ 1 & 0 & -1 \\ 0 & -2 & 2 \end{pmatrix} \quad (11)$$
$${}^i\mathbf{x}_0 = (0, 1, 1, 0)^T; \quad i = 1, 2$$

Consider that it is necessary to fulfil the following constraints (12) at the agent cooperation with the initial conditions (13)

$$2.\gamma_{t_1} + \sigma_{p_3} \le 4; \qquad 2.\gamma_{t_4} + \sigma_{p_2} \le 4 \quad (12)$$

$$\mathbf{u}_0 = (1, 0, 1, 1, 0, 1)^T; \qquad \mathbf{v}_0 = (0, 0, 0, 0, 0, 0)^T \quad (13)$$

On Figure 1 the PN model of the supervised agents is displayed (left), while the corresponding reachability graph (RG) expressing the relations among the PN feasible states is displayed in the Figure 2. The edges among the reachable states represent the transitions. Firing of them realizes the state-to-state passing at the system *dynamic* development. The supervisor is synthesized on the base of prescribed conditions for the global system (cooperating agents) behaviour. The conditions for the agent cooperation are prescribed as follows

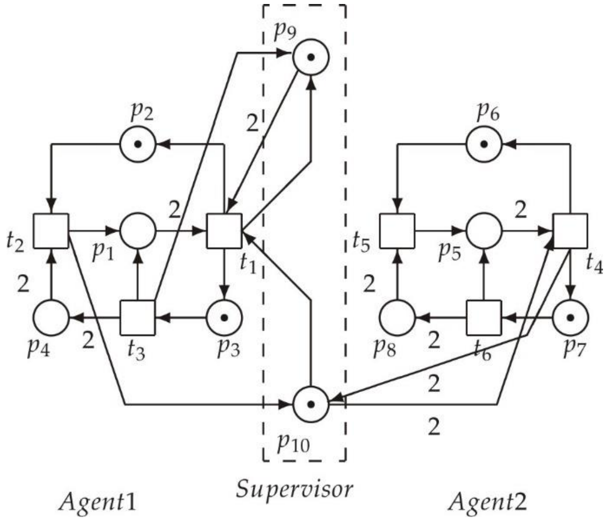


Figure 1. The PN model of the supervised agents

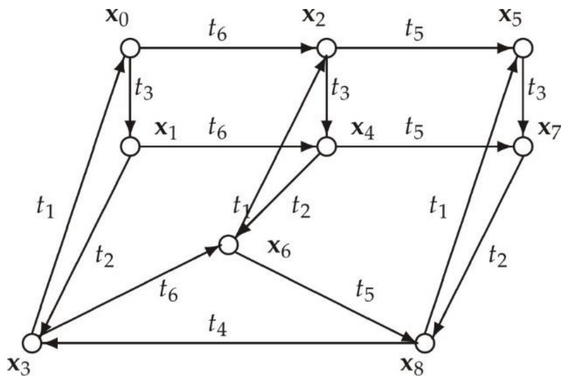


Figure 2. The RG of PN model of the supervised agents

$$\mathbf{b} = (4, 4)^T; \quad \mathbf{L}_v = \mathbf{0}_{2\times 6}$$

$$\mathbf{L}_t = \begin{pmatrix} 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 \end{pmatrix}; \quad \mathbf{L}_p = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Hence, the structure and the initial state of the supervisor are the following

$$\mathbf{F}_s = \begin{pmatrix} 2 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 2 & 0 & 0 \end{pmatrix}; \quad \mathbf{G}_s^T = \begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 2 & 0 & 0 \end{pmatrix}; \quad {}^s\mathbf{x}_0 = (1, 1)^T$$

The nodes of the RG are represented by the feasible state vectors being the columns of the following matrix

$$
\mathbf{X}_{reach} = \begin{pmatrix} 0 & 1 & 0 & 2 & 1 & 0 & 2 & 1 & 2 \\ 1 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 2 & 0 & 0 & 2 & 0 \\ 0 & 0 & 1 & 0 & 1 & 2 & 1 & 2 & 2 \\ 1 & 1 & 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 2 & 0 & 2 & 0 & 0 \\ 1 & 2 & 1 & 2 & 2 & 1 & 2 & 2 & 2 \\ 1 & 1 & 1 & 2 & 1 & 1 & 2 & 1 & 2 \end{pmatrix}
$$

4 Illustrative Example 2

Consider the same two agents, but with different initial states, namely

$$
{}^{1}\mathbf{x}_0 = (2, 0, 1, 0)^T; \quad {}^{2}\mathbf{x}_0 = (0, 1, 1, 0)^T
$$

Let the constraints are different too, namely as follow

$$
2.\gamma_{t_1} + \sigma_{p_5} \leq 4; \qquad 2.\gamma_{t_4} + \sigma_{p_1} \leq 4
$$

The conditions are expressed by the matrices $\mathbf{L}_v$, $\mathbf{L}_t$ and vectors $\mathbf{b}$, $\mathbf{u}_0$, $\mathbf{v}_0$ like in the previous Example 1. Only the matrix

$$
\mathbf{L}_p = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}
$$

has the new form. The supervisor structure and its initial state are given as follows

$$
\mathbf{F}_s = \begin{pmatrix} 2 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 2 & 0 & 0 \end{pmatrix}; \quad \mathbf{G}_s^T = \begin{pmatrix} 2 & 0 & 0 & 2 & 0 & 0 \\ 2 & 0 & 0 & 2 & 0 & 0 \end{pmatrix}; \quad {}^{s}\mathbf{x}_0 = (2, 0)^T
$$

The PN model of the supervised agents are displayed in Figure 3 and the RT of the model is given in Figure 4.

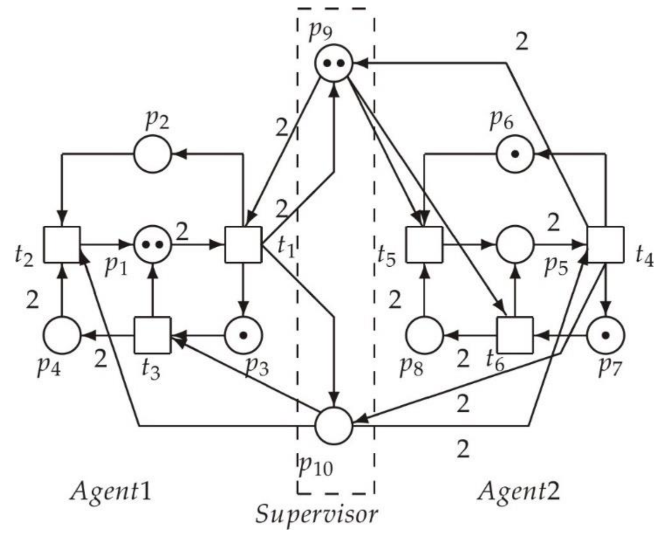


Figure 3. The PN model of the supervised agents

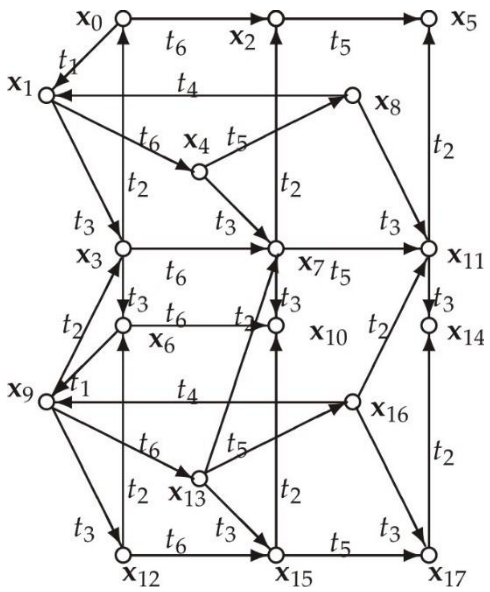


Figure 4. The RG of PN model of the supervised agents

Here, the nodes of the RG are represented by the feasible state vectors of the supervised cooperation being the columns of the following matrix

$$\mathbf{X}_{reach} = \begin{pmatrix} 2 & 0 & 2 & 1 & 0 & 2 & 2 & 1 & 0 & 0 & 2 & 1 & 1 & 0 & 2 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 2 & 1 & 1 & 2 & 2 & 1 & 2 & 2 & 2 \\ 1 & 2 & 1 & 1 & 2 & 1 & 0 & 1 & 2 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 & 4 & 2 & 0 & 4 & 4 & 2 & 6 & 4 & 4 & 6 & 4 & 6 \\ 0 & 0 & 1 & 0 & 1 & 2 & 0 & 1 & 2 & 0 & 1 & 2 & 0 & 1 & 2 & 1 & 2 & 2 \\ 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 2 & 0 & 0 & 2 & 0 & 0 & 2 & 0 & 0 & 2 & 0 & 2 & 0 & 0 \\ 2 & 2 & 1 & 2 & 1 & 0 & 2 & 1 & 0 & 2 & 1 & 0 & 2 & 1 & 0 & 1 & 0 & 0 \\ 0 & 2 & 0 & 1 & 2 & 0 & 0 & 1 & 2 & 2 & 0 & 1 & 1 & 2 & 0 & 1 & 2 & 1 \end{pmatrix}$$

When e.g. it is necessary to reach $\mathbf{x}_{17}$ from $\mathbf{x}_{0,}$ it can be possible by means of the *trajectory*

$$\mathbf{x}_0 \xrightarrow{t_1} \mathbf{x}_1 \xrightarrow{t_3} \mathbf{x}_3 \xrightarrow{t_3} \mathbf{x}_6 \xrightarrow{t_1} \mathbf{x}_9 \xrightarrow{t_6} \mathbf{x}_{13} \xrightarrow{t_5} \mathbf{x}_{16} \xrightarrow{t_3} \mathbf{x}_{17} .$$

5 Conclusions

The procedure of the discrete event systems (DES) supervisor synthesis was utilized here for realizing the cooperation of autonomous agents in order to fulfil the prescription for the collective behavior of the group of agents. The approach is supported by two basic sources of knowledge, namely, the general known results from the DES control theory and the previous results of the author’s works in the area of PNs and DES intelligent control.

Acknowledgement

The authors thank for the partial support of the Slovak Grant Agency for Science VEGA under grant 2/0029/17, and the BAS-SAS grant on the bilateral joint project “Modern Paradigms in the Field of Intelligent Systems” solved during the years 2015-2017. The research work reported in the paper is partly supported under Grant № DFNI-I-02-5 “Inter-Criteria Analysis: A New Approach to Decision Making” and partly supported under Grant № DN 02/10 “New Instruments for Knowledge Discovery from Data and Their Modelling”.

References

[1] Čapkovič F. (2007) Modelling, Analysing and Control of Interactions among Agents in MAS. *Computing and Informatics*, vol. 26, no. 5, 507–541.

[2] Čapkovič F., Doukovska L., Atanassova V. (2015) Automated Control Synthesis for Discrete Event Systems. *Proc. of the conference Big Data, Knowledge and Control Systems Engineering*, November 5-6, 2015, Sofia, Bulgaria, 139-148.

[3] Čapkovič F. (2017) Petri Nets in Discrete-Event and Hybrid Systems Modelling, Analysing, Performance Evaluation and Control. In: R. Szewczyk, C. Zieliński, M. Kaliczyńska, ed., *Automation 2017,* Advances in Intelligent Systems and Computing 550. Kluwer Academic Publishers, London, Chap. 7, 193-212.

[4] Čapkovič F., Doukovska L., Atanassova V. (2016) Cooperation of substantial agents in MAS. *International Journal of Data Sciences*, vol. 1, no. 4, 2016, pp. 353–369.

Key Drivers and Main Challenges for Renewable Energy in Telecommunications

Maria I. Atanassova[1], Tatiana V. Atanasova[2]

[1]Sofia University "St. Kliment Ohridski"

Bul. "Tsar Osvoboditel" 15, 1504 Sofia, Bulgaria

[2]Institute of Information and Communication Technologies – Bulgarian Academy of Sciences

Acad. Georgi Bonchev Str., block 2, 1113 Sofia, Bulgaria

atanasova@iit.bas.bg

Abstract: In a constantly and exponentially digitalizing world, demands for denser and more reliable telecommunication networks have been increasing tremendously. This puts a major challenge for the telecommunication and ICT companies – how to meet the growing demands for faster and better networks, balance costs and pursue sustainable development. Traditional telecommunication technologies and architectures logically need to become "greener". Telecommunication companies generate CO_2 emissions via their vehicle fleet, heating or air-conditioning, but the lion's share comes from their fixed and mobile telecommunication networks. The scale of the problem can be illustrated easily by pure facts and figures: worldwide there are over 4.1 million telecommunication sites (incl. ground based towers and rooftops), providing mobile connection for more than half the world's population – around 4 billion people are now within reach of a 4G network, while 350 million people are covered by a 3G network (2016), a total of 410 million machine-to-machine (M2M) connections, 2.5 billion users of social media on mobile device. The increasing number of users and the booming data traffic lead to increase in power demand by the networks. A single GSM base station could consume between 5-10 kilowatts in the urban areas and a typical cellular base station in rural areas consume 10 - 20 kilowatts of power. We should add to those numbers the billions of kilometers of cables of fixed line access architectures that add up to the energy demands, hence emissions. In general, telecommunications equipment accounts for over 4% of all power used globally and this amount will inevitably increase.

The rise of cloud computing, Internet of Things (IoT) and big data demand will put additional pressure on finding the proper way to power all these new capacities. Key drivers and main challenges for implementation of renewable energy in telecommunication industry are described in the paper. Several examples of implemented policies and measures to minimize CO_2 emissions are shown.

Keywords: Telecommunications, CO_2 emissions, renewable energy.

1 Introduction

The global climate change and the urgent need efficient policies and measures to gain control over increasing CO_2 emissions in the atmosphere, has been affecting all aspect of the modern industries, businesses and day-to-day lives of people. The common urge to improve energy efficiency and increase the share of renewable energy sources in the overall energy mix has become a major source for competitive advantage across all industries.

These trends are also valid for the companies in telecommunications and ICT. They are facing a major challenge to meet the increasing demands for better and faster networks, keep costs under control and ensure sustainable development – environmental and social. Various studies conclude that modern telecommunications industry is accountable for 2.5 - 5 percent of worldwide CO_2 emissions, with a dominating portion due to the telecommunication networks, themselves. This, combined with the exponentially growing traffic, is the basis to expect that these numbers will almost double by 2020.

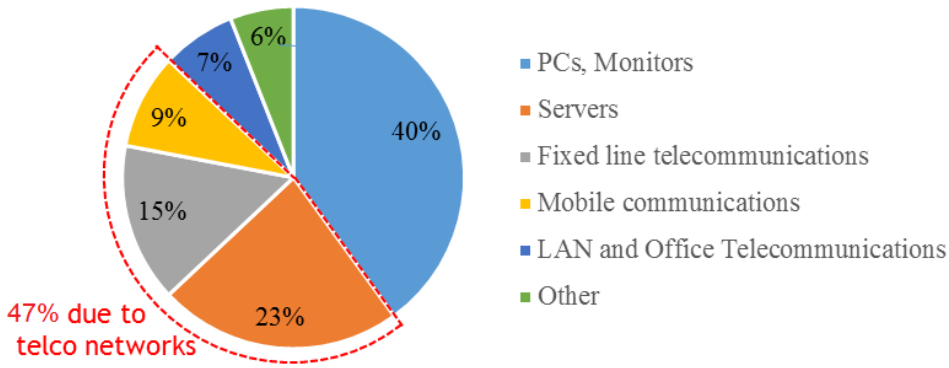


Figure 1. Energy consumption in ICT sector, adapted from [1]

Half of the power consumption of the ICT sector comes from fixed line and mobile communication networks (Figure 1). Various studies have estimated that the energy consumption of a typical telecommunications network operator varies between 1.5 TWh to 9 TWh per annum [3].

Worldwide there are over 4.1 million telecommunication sites (incl. ground based towers and rooftops), providing mobile connection for more than half the world's population – around 4 billion people are now within reach of a 4G network, while 350 million people are covered by a 3G network (2016), a total of 410 million machine-to-machine (M2M) connections, 2.5 billion users of social media on mobile device [4]. A single GSM base station could consume between 5-10 kilowatts in the urban areas and a typical cellular base station in rural areas consume 10 - 20 kilowatts of power [5]. We should add to those numbers the billions of kilometers of cables of fixed line access architectures that add up to the energy demands, hence emissions. In general, telecommunication equipment accounts for over 4% of all power used globally and this amount will inevitably increase [1].

Telecom operators can control directly below one-third of the power consumption in the fixed networks, which is their own operational expenditure (OPEX). The remaining over 70 percent occurs in the user segment. Moreover, fixed networks are characterized with high power losses. Different estimates conclude that these losses can reach up to one-fifth of the consumed energy, due to cable transmissions, switches, routers, data centers, etc.

The situation is different with the mobile networks, where only small portion (near 10 percent) of the overall power consumption corresponds to the cellular user, whereas almost 90 percent is incurred by the operator, and mainly for the operation of base stations.

Based on the above, the main drivers for more and more telecoms to start exploring renewable and innovative energy sources worldwide, can be divided into two groups: internal and external. In this paper we will discuss in bigger details some of these factors and their effect on modern telecommunication industry.

2 Internal Factors

The group of internal factors consists of those drivers that are generated from the organization: the need to optimize operations from within, decrease OPEX, manage growth needs and increase network footprint in a cost effective manner.

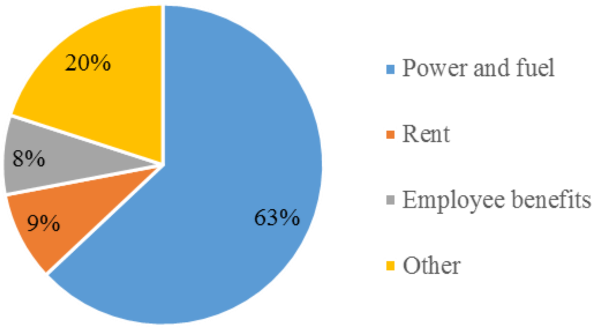


Figure 2. Structure of a Tower company's annual operating costs [2]

Telecoms generate CO_2 emissions through their vehicle fleet, heating, air-conditioning, but especially by the power requirements of the telecom infrastructure – mobile and fixed. Analysis indicates that energy costs account for one third of network operating costs for a large number of telecom operators. The share of energy costs in total OPEX can reach 60 percent for specialized mobile infrastructure companies (tower companies) [2] (Figure 2). Logically, 85% of the total energy costs of a telecom operator are generated by network and data center infrastructures.

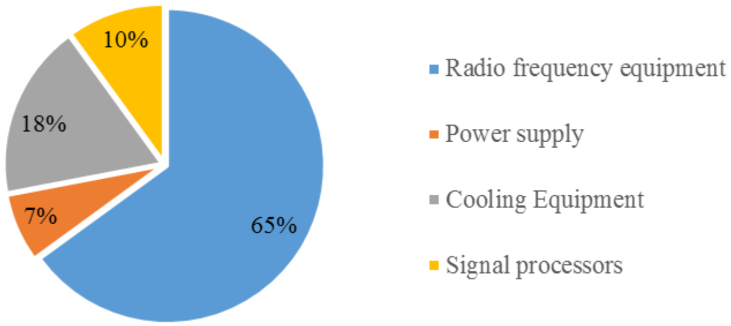


Figure 3. Breakdown of energy consumption of a typical base station with cooling [2]

In mobile networks more than 65% of the power is consumed by the radio equipment and amplifiers, 7% is consumed by the DC power system and 18% by the cooling equipment, an air conditioning unit (Figure 3).

There are already operational examples of implemented measures mainly targeting optimization of power demand and CO_2 emissions:

- massive investment programs to replace the legacy technologies (for example, DOCSIS - Data Over Cable Service Interface Specification is a telecommunications standard used to provide Internet access via a cable modem) with new generation fiber technologies, that require much less energy and provide much better quality services to the end users. Moreover it has been empirically proven that optical fiber is the best fitted solution for energy saving, suffers fewer energy losses, and minimizes the required amplification and signal processing units within the network.
- Smart network planning, aimed at optimal amount of base stations, which are dominating the energy consumption in cellular networks
- Introduction of renewable and hybrid energy sources to replace diesel in the remote areas and conventional energy sources.
- Special attention and innovative solutions for base station equipment cooling technologies. Practice shows that most of the energy is wasted for cooling purposes. To address these problem, a variety of solutions are now available – among the less financially demanding is allowing fresh air flow within the base station, on the other side – investment in new base station equipment, capable of operating in higher temperatures.

Many operators face the challenge that renewable sources, especially solar technologies, still cannot supply reliable 24x7 electricity supply, batteries need to be installed and maintained. Another problem is that the deployment of renewable energy sources is highly dependable and need to be individually adjusted to the meteorological conditions in each area. The main factors, determining the deployment of mobile network sites are the regional cost of conventional energy (in case there is access to the grid) and the diesel prices, the solar and wind conditions and the load requirements of the site. This imposes the need for individual and case-by-case planning for the base stations. In certain cases require a balanced combination of diesel generators, conventional electricity, renewable energy supply and batteries.

Despite that, renewable sources have already proven to be sustainable in the long term and it is the solution to constantly increasing conventional energy prices.

One clear example of a CO_2 neutral network is the incumbent telecom operations in Austria. The company has implemented a set of cost reduction measures and has managed to power its entire network with renewable energy, thus making it 100 percent CO_2 neutral since 2014. The program includes all relevant organizational processes – from network planning, operations and maintenance of mobile radio stations, to data centers and to the customer premises. Thus the company has addressed directly and indirectly 15.9 tons of emissions, or 59% of the total company emissions.

Another major trend, originating from the ultimate goal to optimize operations and decrease operational costs is network consolidation with the support of a dedicated infrastructure management company (tower company). The benefits in energy consumption and CO_2 emissions are clear – instead of supporting and running several networks on one territory, the competing telecoms enter into a partnership and merge their networks into one, or rent the existing infrastructure from a specialized company owning the assets. The optimization varies depending on the type of infrastructure sharing (passive/active equipment sharing). This trend of combining networks, sharing infrastructure in seek of financial and operation synergies, has a positive effect on energy consumption and hence CO_2 emissions.

3 External Factors

The group of external factors include the market trends, regulatory requirements, the more and more demanding standards for social and especially environmental responsibility and, of course – competition.

The main external factor that drives energy consumption is the exponential digitalization - booming traffic demand, the increase in the number of connected machines - machine-to-machine (M2M) connections, etc. The increasing transported data volumes are resulting in rising energy demand of the networks.

Telecoms in most cases are required by regulatory regimes to provide stable and reliable connection throughout a certain territory, which often includes remote areas, and areas without any connection to the electricity grid. As a result, more than one quarter of all tower sites worldwide, or over 1 million telecom towers operate in remote, off-grid or in bad-grid areas, and the numbers are projected to rise. More than 90 percent of these towers operate using in total 130 million barrels of diesel fuel per year, which releases over 40 million tons of CO_2 in the atmosphere [2]. The most vivid example for this case is India. The country has more than 250,000 mobile towers which consume on average 3-5 kilowatts per tower. This results in a consumption of 2 billion liters of diesel per annum. As a result of targeted governmental policy – set of measures and obligatory requirements, India's telecommunication industry has become an example of implementation of renewable energy in mobile telecom infrastructure world-wide. Around 50 percent of India's rural towers and 20 percent of urban towers are powered by renewable hybrid generation sources and grid power in India and the program is ongoing. It is expected to save 5 million tons CO_2 emissions, and USD 1.4 billion in annual energy spending by telecoms.

4 Summary of Key Drivers and Main Challenges

The main challenges for implementation of renewable energy in telecommunication industry are:

- 24-hour supply of electricity;
- Local prices of diesel, local weather conditions (sun and wind);
- Maintenance and batteries;
- Cell tower convectional energy supply availability and demand.

The following key drivers can be summarized as:

- Drive down operating expenses;
- Reduce emissions;
- Improve operational efficiency and hence, quality of services to end users;
- Reduce operational risk, gain competitive advantages.

5 Conclusions

The reduction of the energetic consumption, hence CO_2 emissions by the modern telecommunication industry has become a critical factor in its development – it allows for a major saving of economic resources to be redirected towards realization of sustainable development actions. It has become the key to gaining competitive advantages, enabling new services. Fixed network operators rely on fiber optics to replace legacy technologies, and energy efficiency in their data centers. Mobile network operators are increasingly looking for the optimal balance between renewable energy sources and conventional energy and diesel generators, and innovative cooling techniques. There is also a clear trend towards implementation of various IoT solutions for energy monitoring, improvement of energy consumption behavior and reduction of losses.

The trend for consolidation of existing networks under the management of a professional infrastructure company is likely to speed-up the above trends in many regions worldwide, since the energy cost burden in their total operating costs is significantly higher, compared to the traditional telecoms.

Apart from being a substantial financial opportunity for the telecoms, improving energy efficiency is also a major contribution in the fight against global warming; it has clear social and environmental advantages in the introduction of technologies based on renewable sources.

References

[1] Koutitas G., "A Review of Energy Efficiency in Telecommunication Networks," in IEEE Telfor Journal, Vol. 2, No. 1, 2010.

[2] GSMA – Green Power for Mobile, The Global Telecom Tower Esco Market Overview Of The Global Market For Energy To Telecom Towers In Off-Grid And Bad-Grid Areas, www.gsmaintelligence.com, 2014.

[3] Lubritto C., “Telecommunication Power System: energy saving, renewable sources and environmental monitoring”, Chapter 8, 2010.

[4] GSMA – 2017, “Mobile Industry Impact Report: Sustainable Development Goals”, www.gsmaintelligence.com, 2017.

[5] Ahmad T., Sh. Kalyanaraman, F. Amjad, L. Subramanian, “Solar vs Diesel: Where to draw the line for cell towers?”, 2015.

A Study about Reliability Assessment of Power System

[1]Emel Onal, [2]Tahir Cetin Akinci, [3]Serhat Seker, [4]Hidir Selcuk Nogay

[1, 2, 3] Istabul Technical Universtiy, Department of Electrical Engineering
Istanbul, Turkey
[4] Erciyes University, Mustafa Cikrikcioglu Voctional College
Kayseri, Turkey
eonal@itu.edu.tr, akncitc@itu.edu.tr, eonal@itu.edu.tr, nogay@erciyes.edu.tr

Abstract: Reliability Evaluation involves calculating reliability indices which are a measure of reliable of a system. This evaluation is important for electric utilities since it can show important information about the functioning of the power system. In this report it is aimed to evaluate the reliability of the power systems. The electrical power system reliability can be defined as the degree to which the performance of the elements in the bulk system results in electricity being delivered to customers within accepted standards and in the amount desired. There are software such as DigSilent, Neplan and any other possible solution. In this study, DigSilent software is used for reliability analysis. First a simple network is analyzed then a 96 Bus power system is used to demonstrate results. As a result, load flow analysis gives more realistic results according to connectivity analysis. Finally, for further study suggestion an introduction about multi objective optimization is presented and one of the heuristic methods called the non-dominated sorting genetic algorithm 2(NSGA II) and its codes in Matlab for a object function is presented.

Keywords: Connectivity analysis, Digsilent software, load flow analysis, multi objective optimization, reliability.

1 Introduction

Reliability in general can be defined as the ability of a device or system to perform its required function under stated conditions for a stated period of time or number of cycles or events. Due to the ability of a power system to operate with contingencies, it can be said that the reliability of a power system is the ability of the system to perform its main function. The main function of a power system is to supply electrical energy to its customers. Reliability can further be divided into two parts, adequacy and security. Adequacy is having enough facilities in the system to meet the demands of the customer loads. The facilities referred to are generation, transmission and distribution facilities. Security is the ability of the system to respond to disturbances. The distribution system is the part of the power system that takes the power from the transmission system to the customer. Distribution systems are usually configured as radial type systems, if not they are operated as such. A radial distribution system consists of a set of series components including lines, isolators, circuit breakers, bus-bars etc. Reliability assessment involves determining, generally using statistical methods, the total electric interruptions for loads within a power system during an operating period. Reliability assessment tools are commonly used to quantify the impact of power system equipment outages in economic terms. The results of a reliability assessment study may be used to justify investment in network upgrades such as new remote control switches, new lines /transformers, or to assess the performance of under voltage load shedding schemes. The basic user procedure for completing a reliability assessment consists of the following steps as shown in Figure 1. Steps on the left are compulsory, while steps on the right are optional and can be used to increase the detail of the calculation.

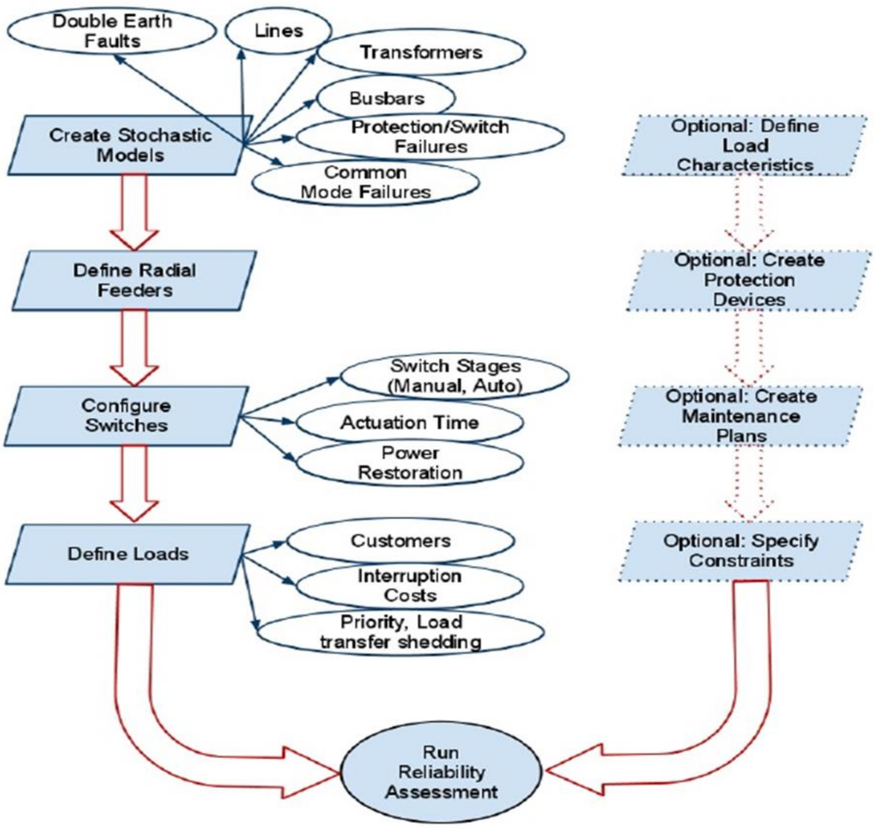


Figure 1. Reliability assessment user procedure

2 The Indices of Reliability Assessment

The Reliability Assessment procedure considers the network topology, protection systems, constraints and stochastic failure and repair models to generate reliability indices. The generation of reliability indices, using the Reliability Assessment tool also known as 'reliability analysis', consists of the following: The failure models describe how system components can fail, how often they might fail and how long it takes to repair them when they fail. The load models can consist of a few possible load demands, or can be based on a user defined load forecast and growth scenarios.

The combination of one or more simultaneous faults and a specific load condition is called a 'system state'. Internally, DigSilent system state generation engine uses the failure models and load models to build a list of relevant system states [1]. Subsequently, the Failure Effect Analysis (FEA) module analyzes the faulted system states by simulating the system reactions to faults. The objective of the FEA function is to determine if system faults will lead to load interruptions and if so, which loads will be interrupted and for how long. The results of the FEA are combined with the data that is provided by the system state generation module to create the reliability statistics including indices such as SAIFI, SAIDI and CAIFI. The system state data describes the expected frequency of occurrence of the system state and its expected duration however, the duration of these system states should not be confused with the interruption duration.

The network reliability assessment produces two types of indices: Load point indices and System indices. These indices are separated into frequency/expectancy indices and energy indices. Furthermore, there are indices to describe the interruption costs. Load point indices are calculated for each load, and are used in the calculation of many system indices. This section describes the simplified equations for the reliability indices. However, note that the DigSilent reliability assessment calculations use more complex calculation methods. Nevertheless, the simplified equations shown here can be used for hand calculations or to gain insight into the reliability assessment results [2]. (C: total custom, a: annual, h: hour)

These indices are defined as follows:

SAIFI, System Average Interruption Frequency Index, in units of [1/C/a], indicates how often the average customer experiences a sustained interruption during the period specified in the calculation

SAIFI_P, Average Interruption Frequency (Contracted Power), in units of [1/a], indicates how often there are contracted power interruptions during the period of the calculation.

CAIFI, Customer Average Interruption Frequency Index, in units of [1/a], is the mean frequency of sustained interruptions for those customers experiencing sustained interruptions. Each customer is counted once regardless of the number of times interrupted for this calculation.

ASIFI, Average System Interruption Frequency Index, in units of [1/a], the calculation of this index is based on load rather than customers affected. ASIFI can be used to measure distribution performance in areas that supply relatively few customers having relatively large concentrations of load, predominantly industrial/commercial customers.

SAIDI, System Average Interruption Duration Index, in units of [h/C/a], indicates the total duration of interruption for the average customer during the period in the calculation. It is commonly measured in customer minutes or customer hours of interruption.

SAIDI_P, Average Interruption Duration (Contracted Power), in units of [h/a], indicates the total duration of contracted power interruptions during the period of the calculation.

CAIDI, Customer Average Interruption Duration Index, in units of [h], is the mean time to restore service.

ASIDI, Average System Interruption Duration Index, in units of [h/a], is the equivalent of SAIDI but based on load, rather than customers affected.

ASAI, Average Service Availability Index, this represents the fraction of time that a customer is connected during the defined calculation period.

ASUI, Average Service Unavailability Index, is the probability of having all loads supplied.

MAIFI, Momentary Average Interruption Frequency Index, in units of [1/Ca], evaluates the average frequency of momentary interruptions. The calculation is described in the IEEE Standard 1366 'IEEE Guide for Electric Power Distribution Reliability Indices'.

3 The Reliability Assessment Verification of Test Results for a Simple System

Firstly, a simple test system which has been built in DigSilent software in used for reliability analysis and the reliability indexes are extracted for different conditions of failure in system. The simple electrical distribution system is shown in Figure 2.

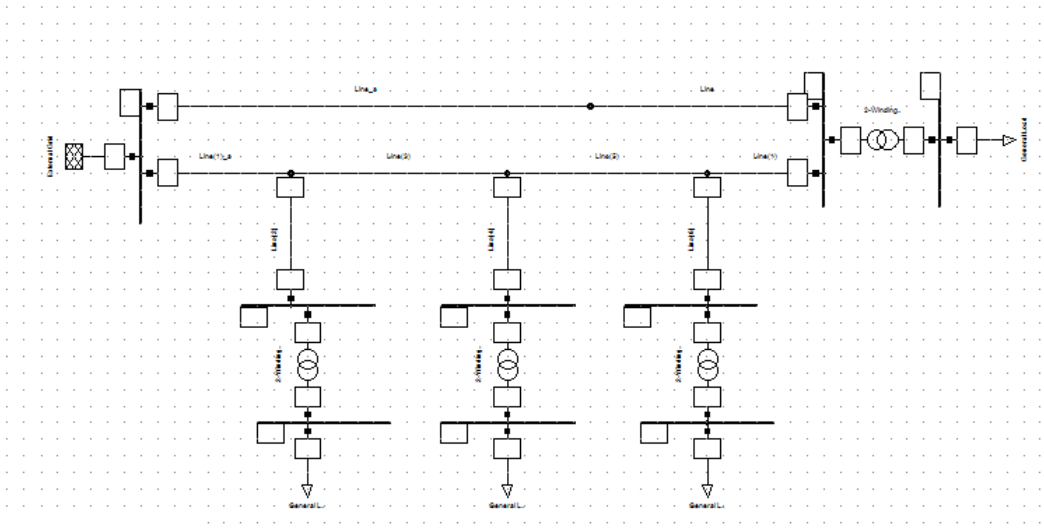


Figure 2. The simple test system for electric distribution system

The different assumptions in reliability analysis:

- Case 1:

The assumptions for case 1 are the following: availability of alternate feed, no fuses on the laterals, disconnects along the feeders and repaired transformers.

Table 1. The reliability system index for case 1

Index	SAIFI	SAIDI	CAIDI	ASAI	ASUI	ENS	AENS
Value	0.6568	5.9632	8.0212	0.9996	0.0007	19.3256	0.0145

- Case 2:

For case 2 the assumptions are: no alternate feed, fuses on the laterals, circuit breakers instead of disconnects long the feeders and repaired transformers.

Table 2. The reliability system index for case 2

Index	SAIFI	SAIDI	CAIDI	ASAI	ASUI	ENS	AENS
Value	0.2312	3.9862	18.2130	0.9995	0.0002	15.2361	0.0254

- Case 3:

Case 3 has the following assumptions: availability of alternate feed, fuses on the laterals, no disconnects along the feeders and repaired transformers.

Table 3. The reliability system index for case 3

Index	SAIFI	SAIDI	CAIDI	ASAI	ASUI	ENS	AENS
Value	0.3125	4.7652	15.5227	0.9992	0.0005	16.4568	0.0126

- Case 4:

The following are the assumptions for case 4: availability of alternate feed, fuses on the laterals, disconnects along the feeders and repaired transformers.

Table 4. The reliability system index for case 4

Index	SAIFI	SAIDI	CAIDI	ASAI	ASUI	ENS	AENS
Value	0.3256	3.6458	12.3687	0.9997	0.0006	11.2365	0.0111

- Case 5:

Assumptions for case 5: no alternate feed, fuses on the laterals, disconnects and the transformers are replaced instead of being repaired.

Table 5. The reliability system index for case 5

In-dex	SAIFI	SAIDI	CAIDI	ASAI	ASUI	ENS	AENS
Value	0.3269	1.2356	4.2333	0.9996	0.0005	7.2355	0.0065

4 The Reliability Assessment of Test Results for a Medium Voltage Distribution System with 96 bus

Now, reliability analysis of is done in medium voltage 33kv/11kv distribution power system (Figure 3). First step in reliability analysis in power system is to define a stochastic (reliability) model for every element in power system and use the necessary data are entered for reliability analysis. Afterwards, nest step is defining feeders in power system and then, regulating the switches and lads for reliability analysis. According to that, it’s possible to model various types of demands, costs, faults, planned outage in power system and reliability analysis.

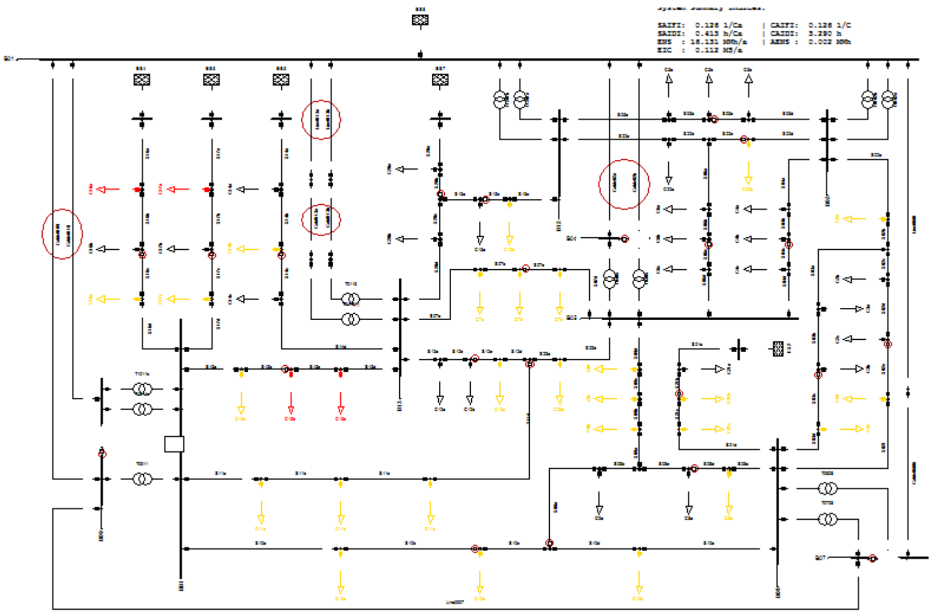


Figure 3. The electrical distribution system with 96 bus

There are mainly two different methods in the reliability analysis of power system in DigSilent. These are the connectivity analysis and the load flow analysis.

In connectivity analysis, the main defect is neglecting the power system limitation and voltage limitations in it. Being that, it is assumed the load is feed only when it is connected to network. As a result, no loading and bus voltage limitations for elements is considered in this method. So that the load flow calculation is for reliability analysis is unnecessary in this method and the burden of the calculations are decreased dramatically but this can cause faults in the results or even make them far from the real state[3].

The difference of the load flow analysis from the connectivity analysis is at following. In this method, all of the operational limits of power system such as load shedding, thermal limits in load transfer approaches, bus voltage limits and so on, are considered. So that, the load flow analysis is unavoidable in every reliability analysis. [4]

As can be seen from the experimental results, the flow analysis is superior to the connectivity analysis.

Table 6 and Table 7 show the reliability indices for the medium-voltage electrical distribution network with 96 bus. The results obtained using the Digsilent program are quite different each other.

Table 6. The results of connectivity analysis for the distribution system with 96 bus

```
|                |                                                     | DIgSILENT     | Project:
|                |                                                     | PowerFactory  |----------------------
|                |                                                     |   15.1.7      | Date:  5/8/2017
| Reliability Assessment
|   Method                                        Connectivity analysis
|   Network                                       Distribution (Optimal Power Restoration)
|   Calculation time period                       2016
|   Consider Maintenance                          No
|   Fault Clearance Breakers                      Use all circuit breakers
|   Switching procedures                          Concurrently
|   Consider Sectionalizing (Stages 1-3)          Yes
|   Time to open remote controlled switches       1.00 min.
|
| Automatic Contingency Definition
|   Selection                                     Whole System
|   Busbars / terminals                           Yes            Common mode                             No
|   Lines / cables                                Yes            Independent second failures             No
|   Transformers                                  Yes            Double earth faults                     No
|                                                                Protection/switching failures           No
| Study Case: Reliability                                                             | Annex:          / 1
| System Summary
| System Average Interruption Frequency Index         :  SAIFI  =    0.424291 1/Ca
| Customer Average Interruption Frequency Index       :  CAIFI  =    0.424291 1/Ca
| System Average Interruption Duration Index          :  SAIDI  =    0.785     h/Ca
| Customer Average Interruption Duration Index        :  CAIDI  =    1.849     h
| Average Service Availability Index                  :  ASAI   =    0.9999104346
| Average Service Unavailability Index                :  ASUI   =    0.0000895654
| Energy Not Supplied                                 :  ENS    =   30.646     MWh/a
| Average Energy Not Supplied                         :  AENS   =    0.003     MWh/Ca
| Average Customer Curtailment Index                  :  ACCI   =    0.029     MWh/Ca
| Expected Interruption Cost                          :  EIC    =    0.217     M$/a
| Interrupted Energy Assessment Rate                  :  IEAR   =    7.090     $/kWh
| System energy shed                                  :  SES    =    0.000     MWh/a
| Average System Interruption Frequency Index         :  ASIFI  =    0.424282 1/a
| Average System Interruption Duration Index          :  ASIDI  =    0.784576 h/a
| Momentary Average Interruption Frequency Index      :  MAIFI  =    0.000000 1/Ca
```

Table 7. The results of load flow analysis for the distribution system with 96 bus

```
|                |                                                     | DIgSILENT     | Project:
|                |                                                     | PowerFactory  |----------------------
|                |                                                     |   15.1.7      | Date:  5/8/2017
| Reliability Assessment
|   Method                                        Load flow analysis
|   Network                                       Distribution (Optimal Power Restoration)
|   Calculation time period                       2016
|   Consider Maintenance                          No
|   Fault Clearance Breakers                      Use all circuit breakers
|   Switching procedures                          Concurrently
|   Consider Sectionalizing (Stages 1-3)          Yes
|   Time to open remote controlled switches       1.00 min.
|
| Automatic Contingency Definition
|   Selection                                     Whole System
|   Busbars / terminals                           Yes            Common mode                             No
|   Lines / cables                                Yes            Independent second failures             No
|   Transformers                                  Yes            Double earth faults                     No
|                                                                Protection/switching failures           No
| Study Case: Reliability                                                             | Annex:          / 1
| System Summary
| System Average Interruption Frequency Index         :  SAIFI  =    0.487638 1/Ca
| Customer Average Interruption Frequency Index       :  CAIFI  =    0.487638 1/Ca
| System Average Interruption Duration Index          :  SAIDI  =   27.583     h/Ca
| Customer Average Interruption Duration Index        :  CAIDI  =   56.565     h
| Average Service Availability Index                  :  ASAI   =    0.9968512415
| Average Service Unavailability Index                :  ASUI   =    0.0031487585
| Energy Not Supplied                                 :  ENS    = 1077.364     MWh/a
| Average Energy Not Supplied                         :  AENS   =    0.103     MWh/Ca
| Average Customer Curtailment Index                  :  ACCI   =    1.009     MWh/Ca
| Expected Interruption Cost                          :  EIC    =    8.713     M$/a
| Interrupted Energy Assessment Rate                  :  IEAR   =    8.088     $/kWh
| System energy shed                                  :  SES    =    0.000     MWh/a
| Average System Interruption Frequency Index         :  ASIFI  =    0.487627 1/a
| Average System Interruption Duration Index          :  ASIDI  =   27.582283 h/a
| Momentary Average Interruption Frequency Index      :  MAIFI  =    0.000000 1/Ca
```

5 The Multi Objective Optimization

The presence of multiple objectives in a problem, in principle, gives rise to a set of optimal solutions (largely known as Pareto-optimal solutions), instead of a single optimal solution. In the absence of any further information, one of these Pareto-optimal solutions cannot be said to be better than the other. This demands to find as many Pareto-optimal solutions as possible from the user. Classical optimization methods (including the multi criterion decision-making methods) suggest converting the multiobjective optimization problem to a single-objective optimization problem by emphasizing one particular Pareto-optimal solution at a time. Over the past decade, a number of multiobjective evolutionary algorithms (MOEAs) have been suggested [5-7]. The one of the most effective of these algorithms is Non-dominated Sorting Genetic Algorithm II (NSGA II). It is a very effective algorithm but has been generally criticized for its computational complexity, lack of elitism and for choosing the optimal parameter value for sharing parameter. One of the evolutionary algorithms is pareto front optimal solutions in one single simulation run. In this study, the codes are written in Matlab software and the results are tested in a object functions. These are the values are in contradiction with each other and optimizing one of them spoils the other one. So instead of one optimized point there is a group of the points (continuous) that no one has any prominence to other one. As a result, for a simple power system, one of the functions (f_1) can be the cost of the changing or repairing the elements in the networks and the other one (f_2) can be the outage period or number of the loads that can be disconnected. The implementation is adapted in two abjectives. For example, one of both can be SAIFI index the other one can be loss power. Here, there is no unique solution. The pareto front optimal solution presents an extensive solution set. The relationship between the number of interruptions and active power loss in a system is one of the most important criteria of optimization. Here, the weak points can be eliminated. In conclusion, this optimization offers many options. Figure 4 offers these possibilities.

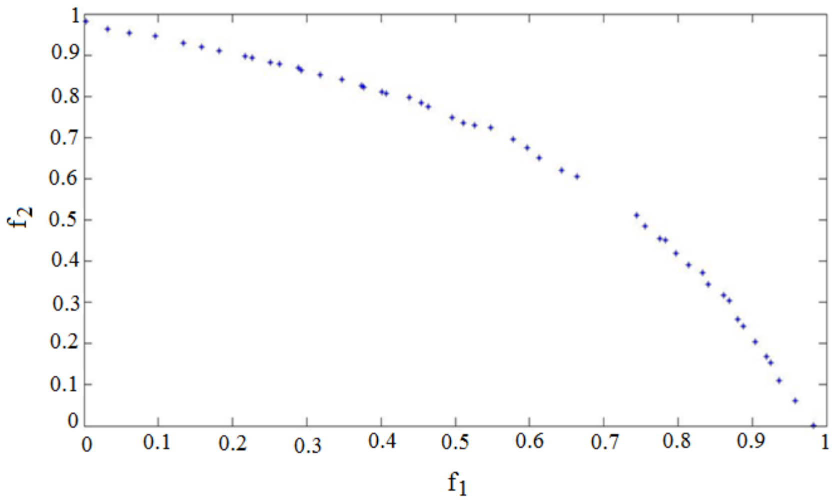


Figure 4. The Pareto front optimal solution for test results

6 Conclusions

In this study, reliability analysis of power systems is done and the necessary data are extracted through DigSilent software. These data can be used for future programming and economic expansion of the power grids. Besides, the multi objective optimization is introduced and one of its methods called NSGA II is programmed and the results are presented. It is possible to combine this optimization for further studies and the results that extracted from the reliability to gain optimal future condition assessment of the network and establishing new equipment in power systems. In this study, reliability analysis indices of radial electrical distribution system have been calculated by using the connectivity analysis load flow analysis. When these results based on system limits are compared, it is shown that load flow analysis gives more realistic results.

References

[1] *DigSilent Power Factory V.15 Tutorial*. (2014), DIgSILENT GmbH.

[2] Suthapanun C., Jirapong P., Bunchoo P.,Thararak P. (2015) Reliability assessment tool for radial and loop distribution systems using DIgSILENT PowerFactory, *12th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*, Thailand, 212-225.

[3] Ngaopitakkul N., Pothisarn C., Bunjongjit S., Suechoey B., Thammart C., Nawikavatan A. (2013) A Reliability Impact and Assessment of Distributed Generation Integration to Distribution System. *Energy and Power Engineering*, 5, 1043-1047.

[4] Mokoka O.K.P., Awodele K.O. (2013) Reliability Evaluation of distribution networks using Neplan & Digsilent power factory, *IEEE Africon conference*.

[5] Deb K., Pratap A., Agarwal S., Meyarivan T.(2002) A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 6, 2, 182-197.

[6] Yijie S., Gongzhang S. (2008) Improved NSGA-II Multi-objective Genetic Algorithm Based on Hybridization-encouraged Mechanism. *Chinese Journal of Aeronautics*, 21, 6, 540-549.

[7] Tomoiagă B., Chindriş M., Sumper A., Sudria –Andreu A., Villafafila-Robles R. (2013) Pareto Optimal Reconfiguration of Power Distribution Systems Using a Genetic Algorithm Based on NSGA-II, *Energies* 6, 1439-1455.

Trends and Challenges of Big Data: A Review

Katia Todorova, Valentina Terzieva, Petia Kademova-Katzarova
Institute of Information and Communication Technologies – BAS
Acad. G. Bonchev Str. Bl. 2, Sofia, Bulgaria
katia@isdip.bas.bg, valia@isdip.bas.bg, petia@isdip.bas.bg

Abstract: The amount of data pouring into organizations as well as their variety and in addition, the speed at which they arrive is increasing tremendously. This tendency is observed in almost every field. Huge amounts of unstructured data have to be transformed into structured, useful and valuable information in order organizations to improve their decision-making process and to fulfil consumers' preferences. Big Data provides the opportunity to deal with these challenges and to derive useful insights. The main contribution of this paper is to present an overview of Big Data trends as well as the challenges they pose.

Keywords: Big Data, Big Data trends, Big Data Challenges

1 Introduction

Nowadays there is an increasing tendency the contemporary world to be run on the flow of data. Almost every sector of the economy, society and politics are data-driven. The free flow of data, as well as their adequately sharing, helps experts to track and spread important issues across the globe. Often the data is not generated intentionally, but as a by-product of manufacturing, commerce, business, administration, health-care and educational systems, social networks, etc. All of this data (human-sourced, process-mediated, and machine-generated) has to be reliable and continuously and timely available in order their potential to be fully exploited [9].

It should be taken into account that data sharing threatens security and individuals' privacy. It's absolutely crucial that everyone has to properly protect the information that manages and controls. On the other side restricted access to data leads to limited access to information, ideas and innovation that can prevent evolving, development and growth in the today's fast-changing global digital era. Considering the aforementioned aspects, it becomes clear that a strong balance between data access and use, from one side, and responsibility and security measures from the other side is needed.

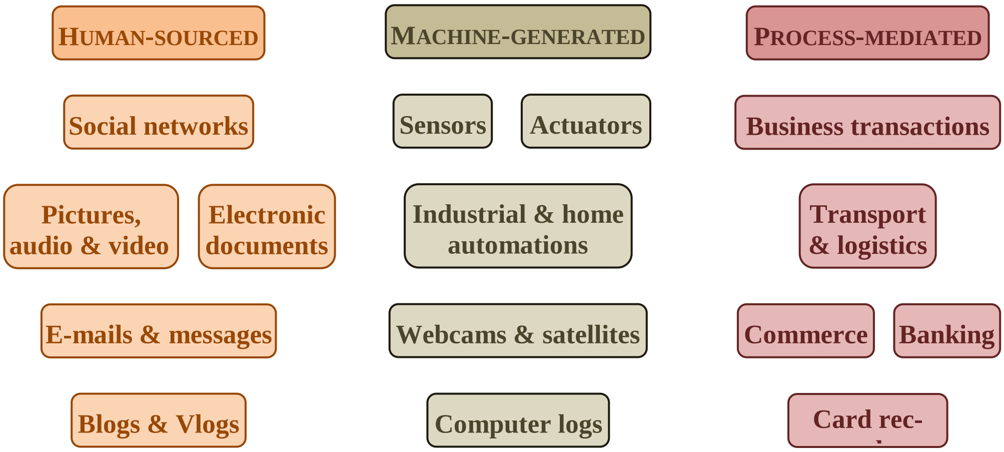


Figure 1. Big Data Source-Diversity

With all this huge amount of data pouring into organizations every day, IT specialists now have to respond to ever larger data requests, to do ad-hoc analyses and timely reports. In order to make informed decisions, decision-makers need to spend more time looking for answers to questions if they can get it at all. For consumers, it is important the information be clear, understandable and usable. So, data visualization, as well as many other issues concerning the data valuability, is of a great importance of today increasingly driven by data society. The Big Data concept and its applications have the potential to deal with these challenges. Big Data is evolutionary – diverse and complex data generated with high speed by machines, networks, and human interaction – and offer many opportunities for stakeholders to get more detailed insides and correlations from the data (Figure 1). Traditional data gathering, storing, analysing are inappropriate for Big Data and raise some important issues that will be considered in this paper.

2 Current Trends

What are the current trends in Big Data development? Big Data is quite big, but its volume continues to grow immensely at cosmic speed: about every two years it doubles [6]. This raises questions about storage space (volume, organization, access times), comprehensibility, understanding, knowledge extracting, efficiency, innovation launching. The following main trends can be outlined:

- Building of Data Lakes – because of the swelling storage consumption [4]

- Increasing data complexity – more and more varied and detailed data reflecting the world's diversity are captured, stored and analyzed [2]
- Ensuring timely data accessing and processing – enable real time and quasi real-time implementations [7]
- Getting insights from the collected data – find sense of them and convert them to actionable items [6, 7]
- Establishing data-driven actions – analysis, management, decision making: these activities are completed by data, rather than by intuition or by personal experience [4]
- Establishing a Big Data environment – re-organize processes and people to increase efficiency [5]
- Accelerating the speed of innovations knowledge – intensive industry fast launching of new products and services obtaining a competitive advantage [5]
- Using data visualization techniques – support understanding [7]

Many of the trends concerning Big Data lead to some serious challenges.

3 Challenges

The successful incorporation of Big Data as a technology that supplements traditional data sources brings many challenges – methodological, organizational, technical, and financial as well as qualification of specialists. So, we look more closely at some of the most important issues arising when using Big Data.

The potential of Big Data for the economic growth as well as for the welfare of the society is still underestimated. But more and more their utility is recognized to solve tasks in each life area so that the number of their applications increases immensely. Of course, this opens up further possibilities for their use, and thus the questions to be answered and the solutions to be found are growing.

It is well known that there is a mass of data available, but are we capable and competent to use and control them productively and effectively? The utilisation of the available data is the core challenge in this matter, but not the only one. Hereinafter the most important challenges are mentioned below.

Heterogeneity. Accumulated data is of different origin and from disparate sources – it has to be stored and organized, i.e. integrated into the Big Data to be prepared for future processing and usage and further submitted to different data sinks.

Compatibility. The disparate sources/sinks of data aforementioned imply the existence of different data input/output formats and data structures and therefore the problem of the data compatibility and consistency arise. Furthermore, the inflexible data structures often imposed from the data users lead to ineffective data processing so the initial problems continue to swell. The aforementioned difficulties require standards to be established and the question arises: is it time to face the problem with a common language for Big Data? [2]

Access and connectivity. Unstructured data cannot be easily accessed by traditional applications. Even across a company a lot of data isn't still connected. There is a lack of appropriate platforms to manage and aggregate the data across an organization. In order to address this problem large research and development enterprises have recently formed joint ventures to find solutions for system integration and data management [1, 2]. This approach will enable to integrate information with operational technologies (using IoT in real-time).

Embedding complex data. This problem is connected with the previously mentioned. It requires a more elaborated data processing methodology than originally assumed (lists, tables, graphs etc.) [2]. Because of the increasingly encompassed, circumscribed and presented reality, the data stored and processed is of more complexity and variations: images, videos, data from smart devices, internet, satellites, manufacturing, healthcare, scientific research etc. The escalation complexity also requires a more sophisticated management and analysis algorithms.

Scalability. As the amount of data in Big Data grows in an avalanche, there is a problem that cannot be postponed: which data is exactly needed to accomplish a task. The issue is not only about the amount of memory required to store, but also the speed at which data can be accessed and appropriately processed for a particular purpose. One of the possible solutions is the separation of local data from data of a more general or of global importance and validity by determining where and how the different types will be stored and processed. It is likely to require splitting into several levels. Generally, scalability on-demand, a mechanism for sifting and supplying the data needed has to be ensured, in order to provide just the required data [4].

Quality. Big Data is created at a top speed and its variety is very large, so its quality is not perfect. According to Data Quality Assessment Framework (developed as an assessment methodology for International Monetary Fund – IMF), seven important properties are defined to characterize the data [8]:

- credibility – relates to the reliability of the data sources

- accuracy – correct representation of the required values
- accessibility – on a physical, logical and conceptual level
- relevance – appropriate to the users' goals
- timeliness – not only actual but also in time, when needed (extremely important in real-time implementations)
- coherence – logically connected and mutually consistent
- interpretability – understandable to be adequately used

Validation. What we have said about the abovementioned data properties results in the requirement to execute data validation. Validation is an important process to recognize and improve data quality. Similar pieces of data from different sources don't always coincide and the task of the validation is to make them match [6]. Validation has to perform two very important operations and make sure that the proper data is supplied:

- removing input errors (inaccurate data retrieval), duplicated data, irrelevant data (inappropriate to issue) and incorrectly linked data
- cleaning „dirt "– detecting, deactivating and eliminating fake data (purposely manipulated) and garbage (unintended and pointless data)

Security. Big Data is liable to hacker attacks because they contain the raw, processed and analyzed data of many organizations. The affected subjects do not lose only intellectual and financial resources, but also effectiveness, reputation, customers etc. On the other hand, security measures are a great obstacle for businesses to exploit the full potential of their own data. To minimize the risks of compromised data, limited data access according to job position, data access authentication, and proper data encryption have to be ensured [4, 6].

Privacy violations. The huge amount of data collection for statistical, research and other public purposes (healthcare, education, effective management of economics and other processes) leads to problems with the privacy violation. It results in an obvious contradiction between privacy and the ubiquitous use of Big Data. The current approach to preserving the balance between privacy and society is no longer appropriate. Mechanisms for privacy protection have to be used when gathering, generating, storing and processing data and the use of data has to be in a responsible manner [3].

Time issues. Collecting, storing, processing and analyzing a huge amount of data (activities requiring significant time resources) on the one hand, and their timely finding and delivering to customers on the other is one of the biggest challenges facing the Big Data [7]. It is hardly expected that Big Data will be used to manage production processes requiring real-time responses. In many cases, however, we can talk about quasi-real-time work, like, for example, in stock exchange operations or in health-related situations where the extraction of significant quantities of appropriately processed information is of great importance. An appropriate approach to the provision of information suitable for each case is necessary to be found and applied. The question of embedding time variables also arises.

Education and training. The successful implementation of Big Data requires employment of appropriately educated data scientists, a sophisticated team of developers and analysts who also have a sufficient amount of domain knowledge. Experienced data specialists, capable to work with new technologies and interpret the data are rare [4]. The demand for experts with Big Data skills often imposes on the organizations to educate, train and cultivate their staff alone.

Cost management. Concerning the usage of Big Data the following costs should be calculated [4]:

- acquiring new hardware, paying cloud services
- hiring additional personnel
- training, maintenance, and expansion

Organizational resistance. The implementation of the Big Data causes changes in the human aspect. Often there is a deficiency of middle-level management and lack of understanding and accepting the changes by the staff. The solution is, as aforementioned, to educate and train the staff [6].

4 Conclusions

In our fast-changing and always connected world, more and more organizations realize the importance to be well informed in order to face the frequently altering customers' preferences. An increasing number of them start to utilize Big Data to make better tactical, strategic and operational decisions based on systematized data and detailed real-time reports rather than on intuition or personal experience. Using Big Data enables analysis of tremendous amounts of data and extraction of hidden insights that otherwise, from "ordinary" data, remains veiled or are not intended for that purpose. All these positive prerequisites inevitably lead to some challenges and if there is a willingness to deal with them, the potential for the business opportunities is unlimited.

References

[1] Abhay Kumar Bhadani, Dhanya Jothimani, Big Data: Challenges, Opportunities and Realities, Retrieved from https://arxiv.org/ftp/arxiv/papers/1705/1705.04928.pdf

[2] Aisha Javed, Challenges of Big Data, Retrieved from https://dzone.com/articles/challenges-of-bigdata

[3] Bart van der Sloot, How to assess privacy violations in the age of Big Data? Retrieved from http://www.tandfonline.com/doi/abs/10.1080/13600834.2015.1009714?journalCode=cict20

[4] Big Data Challenges and Opportunity, Retrieved from https://www.qubole.com/resources/big-data-challenges/

[5] Cornelia Hammer, Diane Kostroch, Gabriel Quiros Big Data: Potential, Challenges, and Statistical Implications, Retrieved from https://books.google.bg/books?isbn=1484318994

[6] Cynthia Harvey, Big Data Challenges, Retrieved from https://www.datamation.com/big-data/big-data-challenges.html

[7] Five big data challenges, Retrieved from https://www.sas.com/content/dam/SAS/en_us/doc/other1/five-big-data-challenges-106263.pdf

[8] IMF's Data Quality Assessment Framework, Retrieved from https://unstats.un.org/unsd/accsub/2010docs-CDQIO/Ses1-DQAF-IMF.pdf

[9] Steven Vale, Classification of Types of Big Data, Retrieved from https://statswiki.unece.org/display/bigdata/Classification+of+Types+of+Big+Data

Personal Assistance with Ontology Based Knowledge

Jordan Todorov[1], Boris Daskalov[2], Nina Stancheva[1],
Vladimir Valkanov[1], Stanimir Stoyanov[1]

[1]Dept. of Computer Systems, Plovdiv University "Paisij Hilendarski", Bulgaria

[2]Institute of Information and Communication Technologies, BAS, Bulgaria

jtodorov@uni-plovdiv.bg, bdaskalov07@gmail.com, st.nina@abv.bg,
vvalkanov@uni-plovdiv.bg, stani@uni-plovdiv.bg

Abstract: The paper presents the concept for generic personal assistant that is going to use ontologies as its knowledge base. We introduce a basic approach for the interface between the ontology and the JADE agent that is going to be used as a middle man between the knowledge base and the JADEX mobile assistant.

Keywords: JADE, JADEX, Ontologies

1 Introduction

Electronic learning is being gradually included as an important element in the educational strategy of even more educational institutions. This process directly relates to the most important aspects of progress of a certain society and of mankind. Over the last years there is increasing interest to electronic learning as well as the creation of electronic systems which support and aid its development. A deeper look at the types of media "proves" that there is hardly a difference in studying from an electronic environment than from the traditional face-to-face process. The conditions that a virtual environment can provide to its users are very close to the real conditions but without the added restrictions of spatial limitations and time. There are many different technological solutions which claim and define their products as environments for electronic learning. In that regard there is no strict definition to help us clearly differentiate between what is and what is not an environment for electronic learning. On a global scale two types of systems for electronic learning have become prevalent: Learning Management Systems (LMSs) and Learning Content Management Systems. The first type is focused on the students' progress in all learning activities. The second type is for creating and managing learning content where the authors can create, store and manage the content, and can use it repeatedly. Such types of systems that have already proved their usefulness (Moodle [1], ILIAS [2], Sakai [3], ECollege [4], et al) are becoming more popular in high education institutions.

In the Faculty of Mathematics and Informatics at the Plovdiv University "Paisii Hilendarski" there is such a system – Distributed eLearning Centre (DeLC)[5], which was developed at the university. Its value is proven by the thousands of students who used it and continue to use it every year. Despite its proven efficiency in the process of electronic learning, DeLC has a serious flaw – it cannot register processes taking place in the physical world. This flaw is the reason that a new reengineering project was started for transformation of DeLC into VES [6]. VES contains a set of intelligent components which are controlled by intelligent agents. A well designed space allows us to further implement new components which allow its expansion. One of the main goals of the space is to support the learning process for students through personal assistants. To achieve that goal, the space is inhabited by intelligent assistants. The main role to achieve this goal goes to three types of assistants – guards who are responsible for safety and performance of certain scenarios, operative assistants which provide a suitable interface for the existing electronic services, and personal assistants who provide the necessary entry points to the space and in special occasions act as middle men for activating certain scenarios or services. In order to expand the usefulness of the space, we are introducing a general ontology assistant that is going to allow the transformation of all the current knowledge of the space into ontologies and allow every agent to access and use them.

2 Related Works

Nowadays the use of ontologies in personal assistance is even more widely used since they allow more complex understanding of the information that the agents are using and providing. There are many works currently being developed in this field but some of most notable ones are:

- Siri[7]: Apple was the first large technological company, which in 2010 integrated the intelligent assistant Siri in its operating system. In its newer versions Siri supports different techniques for self-education and uses ontologies be able to understand the queries asked by the users and to be able to provide better information to them.
- MELISA[8]: An agent whose purpose is to allow smart search of medical literature to provide aid to medical professionals. It does that by analyzing the query input by the user, then searches for the needed information in different medical ontology bases and the returned information is once again filtered and then provided to the end user.

- KRAFT[9]: The KRAFT project permits the extraction of knowledge from relational databases, knowledge systems et al, and transfers them to ontologies that are subsequently used.
- OIDSA[10]: An ontology based intelligent agent used for monitoring and control of CMMI[11]. It allows natural language processing and includes fuzzy agent that computes the planned progress against the actual progress based on the CMMI ontology and then the final decision support agent measures the results based on all the information provided.

All of the projects mentioned above use different approaches and perform different tasks with ontologies. Using some of their findings and integrating them with our own ideas, our goal is to create a general ontology agent for the needs of VES.

3 Architecture

The system is composed of Jadex[12] agent, JADE[13] agent and interface for processing the knowledge data to ontologies as shown in Figure 1.

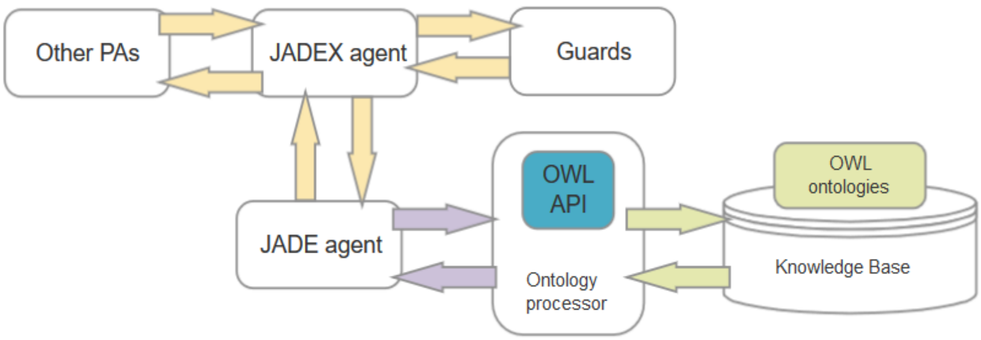


Figure 1. Architecture

The Jadex agent contains BDI[14] architecture and its purpose is to communicate with the other agents located in VES and provide them with the needed information on request or store the new information into ontologies. It acts as a middle man between the JADE agent and the rest of the virtual education space.

The Jade agent is a reactive agent that does the actual job of retrieving the ontologies and sending the newly acquired information to be transformed and stored into the ontologies base. Jade agents work more easily with OWL API (Web Ontology Language Application Programming Interface) which is why we are using this separate agent for this purpose instead of doing that in the Jadex agent. A basic approach for achieving the connection between JADE agents and ontologies is introduced. This feature is realized by implementing an ontology processor entity that would work as an interface between the ontology and an agent. The idea of this manager is to provide services that could be used by the agent when certain knowledge is needed. These services include loading of one or more ontologies, extracting knowledge with defined constraints, adding knowledge to the ontology or verifying knowledge using the ontology and even extracting and transferring information to ontologies from a relational database.

The ontology processor is realized as Java class, and Protégé-OWL is used for developing the ontologies. Due to this, OWL API is used as a library to access the ontologies. OWL API is a Java API, so it is suitable for the purpose.

A similar approach is applied for developing an environment for automatic generation of test questions [15, 16]. The ontology in the UML domain is accessed by the test operatives, using an ontology helper entity and OWL API. The ontology helper entity is similar to the ontology processor included in the architecture, which presented in this paper. The test operatives are JADE agents, which use the helper entity to get the needed knowledge in a suitable format. The knowledge is extracted from the ontology by the helper class on certain criteria given by the JADE agent. The criteria are formed by the agent according to its tasks and specific purpose. In the environment for automatic question generation, the test operatives need specific knowledge from the ontology to generate test questions on the UML topic. One of the features of the environment is the generation of different question types. Each type requires different type of knowledge and axioms, so based on this the agent forms its criteria. The questions are also thematic, so other criteria include the topic of the test that is being generated by the operative.

4 Life Cycle

The life cycle of the system goes through two phases:

1. **Initialization phase** – In this phase the system is checking all of the available knowledge base of the space against the existing ontologies, and if there are missing ontologies it goes to the process of creating them.
2. **Working phase** – In this phase the system has completed its initial tasks and waits for a change in the environment. Such change can be a query coming from another agent, in that case the Jadex agent processes the message and then sends a request

for specific ontologies to the Jade agent. Another example is when the system acquires new information that needs to be stored as a new ontology. In this case, the Jade agent uses the ontology processor to generate it and then stores it.

5 Conclusions

In this paper we have introduced a new ontology based assistant whose purpose is to create deeper use of ontologies in VES by creating ontologies for the entire knowledge base of the space. The system is going to do the same with all of the newly acquired knowledge and also allow easy access to it to every agent in the space.

References

[1] Moodle, https://moodle.com/, Accessed October 2017.

[2] ILIAS, http://www.ilias.de/docu/, Accessed October 2017.

[3] SAKAI, https://sakaiproject.org/, Accessed October 2017.

[4] ECollege, http://www.ecollege.ie/, Accessed October 2017.

[5] Stoyanov, S., Popchev, I., Doychev, E., Mitev, D., Valkanov, V., Stoyanova-Doycheva, A., Valkanova, V. , Minov, I., DeLC Edu-cational Portal, Cybernetics and Information Technologies (CIT), Vol.10, No 3., Bulgarian Academy of Sciences, 2010, pp. 49-69.

[6] Stoyanov S., A Virtual Space Supporting eLearning, Proceedings of the Forty Fifth Spring Conference of the Union of Bulgarian Mathematicians Pleven, April 6–10, 2016, 72-82.

[7] Siri, http://www.apple.com/ios/siri/, Accessed October 2017.

[8] Abasolo J.,Gómez M., An ontology-based agent for information retrieval in medicine., Retrieved from http://www.ics.forth.gr/isl/SemWeb/proceedings/session3-1/paper.pdf, Accessed October 2017.

[9] Pepijn V., Valentina T.,An Experience with Ontology-based Agent Clustering.,Retrieved from http://ceur-ws.org/Vol-18/12-visser.pdf, Accessed October 2017.

[10] Lee C., Wang M., Chen J., Ontology-based intelligent decision support agent for CMMI project monitoring and control. Retrieved from http://www.sciencedirect.com/science/article/pii/S0888613X07000783, Accessed October 2017.

[11] Yoo C. , Yoon J. , B. Lee, A unified model for the implementation of both ISO 9001:2000 and CMMI by ISO-certified organizations, The Journal of Systems and Software 79 (7) (2006) 954–961

[12] Jadex, Implementing a BDI-Infrastructure for JADE Agents, *Search of Inovation (Special Issue on JADE)*, vol. vol. 3, no. no. 3, pp. 76-85, 2003.

[13] Bellifemine F., Caire G., Greenwood D., Developing Multi-Agent Systems with JADE, 2007.

[14] A.S. Rao, M. Georgeff, BDI Agents: from Theory to Practice, In: Proceedings of the 1st International Conference on Multi-Agent Systems, San Francisco, CA, 1995, 312-319.

[15] Stancheva N., Stoyanova-Doycheva A., Stoyanov S., Popchev I., Ivanova V., "An Environment for Automatic Test Generation", Cybernetics and Information Technologies, Volume 17, ISSN (Online) 1314-4081, DOI: https://doi.org/10.1515/cait-2017-0025

[16] Stancheva N., Popchev I. , Stoyanova-Doycheva A., Stoyanov S. "Automatic generation of test questions by software agents using ontologies", 2016 IEEE 8th International Conference on Intelligent Systems (IS), 741 - 746, 2016.

Holistic Approach to Data Protection - Identifying the Weak Points in the Organization

Gaidarski, I, Kutinchev, P,
Institute of Information and Communication Technologies,
Bulgarian Academy of Sciences
i.gaidarski@isdip.bas.bg, kutinchev@isdip.bas.bg

Abstract: In this paper, we present a holistic approach to data protection. To effectively protect the organization's data, we must ensure that all of the data, both inside and outside organization, must be collected, classified and identification of the sensitive data is done. Our approach includes profiling of the participants in the information process, based on what data they are using or processing, how and where they are using or processing it. It allows us to identify the weak points in the organization and after risk assessment is done, appropriate measures are taken, by deploying appropriate security controls as Data Leak Protection (DLP) systems, which are designed to stop the data leakage from the inside to the outside.

Keywords: Data protection, Sensitive data, Data governance, Holistic Approach, Data Leak Prevention

1 Introduction

Today's organizations entirely depend from the information. The information is the most valuable asset, which organization owns. Customer data, financial and legal records, day-to-day correspondence – all that information is vital for the organization. The loss of that information can cause major damages.

Over recent years, organizations from practically all sectors have seen their sensitive data lost, stolen or leaked outside. These data loss incidents have cost millions in direct and indirect costs and have resulted in damages to brand and reputation of the organizations.

There are different types of data loss incidents, including theft of trade information, sale of customer details to external parties, loss of USB sticks, laptops or mobile devices. The majority of these incidents resulted from the actions of internal users or trusted third parties [1].

The data protection is therefore crucial and needs to be met with high priority. In order to accomplish this, a number of IT Security controls must be implemented, combining strategic, operational and tactical measures.

For the successful implementation of these measures every organization must give answers to few important questions:

- What sensitive data the organization hold?
- Where is the organization's data going?
- Where does the organization's sensitive data reside, internally and with third parties?

To answer them, a preliminary work must be done to identify all the sensitive data it possesses and use. When the sensitive information is identified, the next steps must be performed to protect the organization's sensitive data [2]:

1. To identify the threats to the sensitive data;
2. To take into account the associated risk and cost of losing that sensitive data;
3. To take into account the cost of the data and the protecting of that data;
4. To take into account the regulatory obligations for protecting the sensitive data;
5. To make an IT Security Policy for protecting the data and to provide usability of the organization's resources;
6. To develop a strategy for implementation of the IT Security Policy;
7. Maintain adequate security measures for protection of the sensitive data;
8. To make a strict policy for preventing the disclosure of the sensitive information to unauthorized parties;
9. To reduce the organization's risk and cost of compliance;

We will use data centric approach, which is holistic by its nature to protect the meaningful data of the organization.

2 Holistic Approach to Data Protection

For effective data protection, we employ a holistic approach in which we must consider all the possible data - both inside and outside the organization. To cover it we must have in mind the following considerations:

- Where does organization's data resides?
- What data the organization hold?
- Who is accessing the data?
- Where the data is going?
- What is the value of the organization's data.
- What are the organization's obligations for data protection - Regulations, standards and policies.
- Data protection risks and gaps.

- How to protect the data?
- How to prevent data leakage?

Holistic approach will ensure that the organization's most sensitive data is identified and effectively protected with combination of security controls and measures. The goal is to protect the organization's data keeping regulatory compliance and continuity of the business process.

3 Data Governance, Classification and Risk Assessment

To identify the organization's sensitive data, a data governance activities must be performed. With their help the organization will gain understanding what is it most important data, what sensitive data it holds, where does the sensitive data resides - both internally and with third parties and where is that data going.

The organization's data must be classified from a business perspective. It is not enough to just label the data as confidential, secret, private or public. There must be understanding for the specific types of data holded by the organization and use it to perform data protection, which can be customized for the specific business needs.

In general, there is no universal description of what is the organization's most important data, as it depends largely on the type of activities it has. By identifying what exactly data is vital to the organization's goals and day-to-day operations, it becomes clear which data it cannot afford to lose and must be protected.

When classifying data there must be considered the consequences of that data becoming available to unauthorized parties. It can be business disruption, damage to reputation and brand, regulatory violations and fines, loss of competitive advantage, direct financial losses and etc [2].

On Table 1 is shown some example categories for identifying common types of sensitive data.

Table 1. Example categories of sensitive data

Corporate data	Transaction data	Customer data	Personally identifable data
Price/cost lists Target customer lists New designs Source code Formulas Process advantages Pending patents Intellectual property Unreleased merger/acquisition plans and financial reports	Bank payments B2B orders Vendor data Sales volumes Purchase power Revenue potency Sales projections Discount ratios	Customer list Spending habits Contact details User preferences Product customer profile Payment status Contact history Account balances Purchase/transaction history	Full name Birthday Birthplace Biometric data Genetic information Credit card numbers National identification number Passport numbers Driver's license number

Legal documents Employee personal data		Payment/contract terms	Vehicle registration number Associated demographics Preferences

After the identification of the common data types, it is obligatory to assess the risk they carry with them.

Some considerations for assessing the risk of each data type include [3]:

- The value of internal data;
- Whether the data is protected by regulations;
- Impact on brand and reputation;
- Loss of competitive advantage;
- Direct impact to business partners and customers.

Where does organization's data resides? In general the data can reside virtually everywhere within the IT infrastructure of the organization. It can be stored on databases, cloud, servers or workstations, shared through the internal network or sent out through the internet, or carried on mobile devices. Once the most valuable data is identifyed, there must be determined where it resides within the IT infrastructure of the organization. There are two high-level ways to store the data:

1. Structured repositories: When the data is organized into structured repositories, such as relational databases.

The strategies for identifying structured sensitive data must include working with the business and IT staff to identify repositories that hold the sensitive data:

- Use of questionnaires to interview analysts, business process owners, application developers, database administrators, and other relevant parties;
- Business process walk-throughs;
- Reviews of the organization's documentation - application descriptions, data flow diagrams.

2. Unstructured repositories: When the data is stored in less controlled repositories such as workstations, laptops, removable devices and network shares. They are usually stored by end users without complying with any general rules and are inherently chaotic. In this case, the strategies for identifying the sensitive data also can include inquiry of business users and IT personnel, to provide useful information about the data stores. Due to the nature of the unstructured data, which can be stored in unpredictable locations, the usage of special Data Discovery tools are needed:

- Scanning of network segments to identify undocumented shared drives, servers and databases;
- Scanning of network shares, databases with different levels of access privileges and intranet sites;
- Scanning of user workstations and laptops to identify sensitive data stored on local drives.

These activities can be performed with the help of dictionaries and rules, especially designed for detection of sensitive data, based on the identified risk data types.

With the help of these tools can be identified data repositories with sensitive data and after that next steps can be performed to secure the data – mark them, delete or move to more controlled locations.

Where the data is going? Understanding where you hold sensitive data internally is essential, but it does not provide a complete view of where your data resides. To provide complete view of where the sensitive data resides it is not enought to know where the data resides. The organization must be aware where where and how it flows from the inside out direction. Special Monitoring tools can be used for complete picture of the data transfers.

Third-party data access. Very important understanding is what exactly sensitive information is accessible by and exchanged with third parties. Controls must be implemented to secure the third-party access, including:

- Secure data transmissions,
- Controlled access to company networks and data,
- Monitoring of third-party access to company resources,
- Third-party due diligence/information security assurance.

Profiling of the participants in the information process. After we have discovered and identified the sensitive data and the way if flows throughout the organization and to third parties, we must profile the participants in the information process, including internal and external for the organization people. It is from crucial importance for the organization to have complete picture of:

- What data they are using;

- What data they are processing;
- How they are using or processing it;
- Where they are using or processing it;
- Why they are using or processing it;
- Do they have the right to use it or process it;
- When they have access to the data;
- Is there some pattern of their action;
- Is there are some unusual activities, related to them.

That's extremely tough task and can be performed with the help of experienced business consultants, business owners and even psychologists. Profiling the staff with access to sensitive information can be critical for protecting and preventing its leakage. The risky employees can be restricted in their access and actions to sensitive information and even quarantined with zero access to it.

Regulations, standards and policies. With the knowledge of the sensitive data and the people who is using it, a Data Protection Policies must be developed and customized to document all the security requirements to the organization, IT infrastructure and the stuff, having in mind all possible Data Protection Regulations, Standards and Policies which concern that organization. Such standards are ISO 27000 [4][5], ISACA's COBIT [6][7], NIST "800 series" [8], special sector-specific regulations - the Gramm-Leach-Bliley Act (GLBA) [9] for the financial sector, Sarbanes-Oxley Act (SOX) [10][11] for US public companies, Health Insurance Portability and Accountability Act (HIPAA) [12] and Payment Security Industry (PCI) Data Security Standard (DSS) [13].

One of the most important regulations is the upcoming in 2018 EU General Data Protection Regulation (GDPR), issued by the European Commission, the European Parliament and the Council of Ministers of the European Union. The purpose of EU GDPR is to strengthen and unify the protection of personal data for individuals within the European Union [14].

4 Measures for Data Protection

There are different types of threats to the organization's data. They can be threats to the data's confidentiality, integrity or availability - data theft, data loss, data erasure or destruction and altering of the data. The threats can be result from different causes [1][18]:

- Intentional Actions: Intentional deletion of a file or program, theft, hacking, sabotage, malicious act – virus, worm, malware, ransomware, etc.;

- Unintentional Actions: Accidental deletion of a file or program, Misplacement of USB Flash drives or DVDs, Administration errors or other;
- Failure: Power failure, Hardware failure, Software crash/freeze, Software bugs, Data base corruption or many others;
- Disaster – fire, flood, earthquake, etc.

There is also the regulatory risks. Recently, the increased volume of incidents has resulted in renewed focus from regulators. Data protection requirements, as breach notification rules are becoming more strict, and the penalties – stronger. Thus, from the organization's perspective, reducing the risk of data loss reduces regulatory risk and helps to protect the company's brand, strategic business data and intellectual property [3].

In our holistic approach we are using data centric security meta-models, which main focus is data. These models are result from Conceptual modeling of Information Security Systems [ISS] and concerns different aspects of the data and their protection – Figure1.

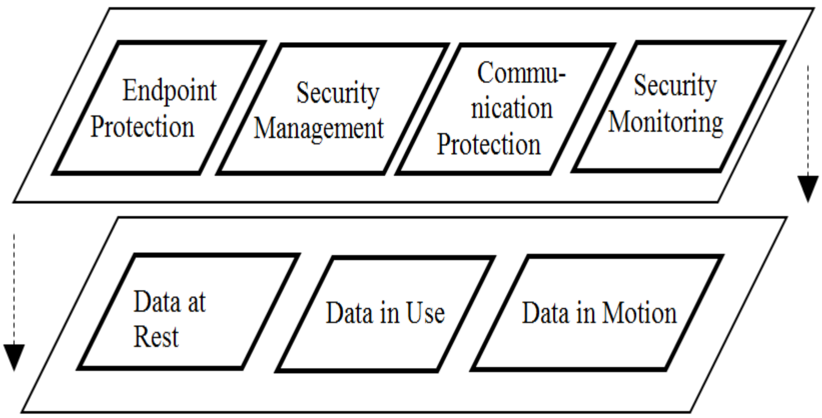


Figure 1. Meta-Models of ISS

With these security meta-models of ISS, we represent the all of the data in ISS – the protected sensitive data, the system management and configuration data and the data collected as part of the monitoring and analysis function of the ISS [15][17][18].

The meta-models can give us clear picture of what security measures and controls we must use in the ISS [19]:

For protecting Data-at-Rest and Data-in-Use we can use Endpoint protection controls:

- access control,
- passwords,

- antivirus,
- audit trails,
- physical measures.

For protecting Data-in-Motion we will use Communications protection controls:

- cryptographic techniques,
- network segmentation,
- perimeter protection,
- gateways,
- firewalls,
- intrusion detection,
- network access control,
- deep packet inspection,
- network log analysis;

Although security solutions like firewalls, unified threat management, intrusion detection/prevention systems can detect and/or check any threat to an organization, these tools are not effective when it comes to data-specific approaches. This is where Data Leak Prevention (DLP) solutions for dedicated data protection comes into the picture.

Data Leak Prevention systems are designed to prevent attempts to steal, modify, prohibit, destroy or obtain unauthorized access and usage of the data, without interrupting normal business processes. DLP can significantly reduce the leak of sensitive information, resulting from internal threats like human error, intentional action or outside breach. The main goal of DLP is to stop the data before it leaves protected environment of the organization. DLP solutions can provide very useful information for process of protection of the information:

- Identification of the violations, threats, risks and vulnerabilities to the data,
- Violations of security policies and procedures,
- Discovering and identifying all sensitive information in the organization.

DLP systems can be different types, depending on their focus area:

- With focus on servers, global communications and data channels of the organization. The DLP can control email servers, file transfers from file servers, and Internet traffic filtering;
- Focused on endpoints and local data channels – workstations, laptops, mobile devices - tablets and phones. Controlled channels include all possible physical ports, personal

emails, file transfer to cloud services and more – as DeviceLock DLP Suite, shown on Figure 2 [20].

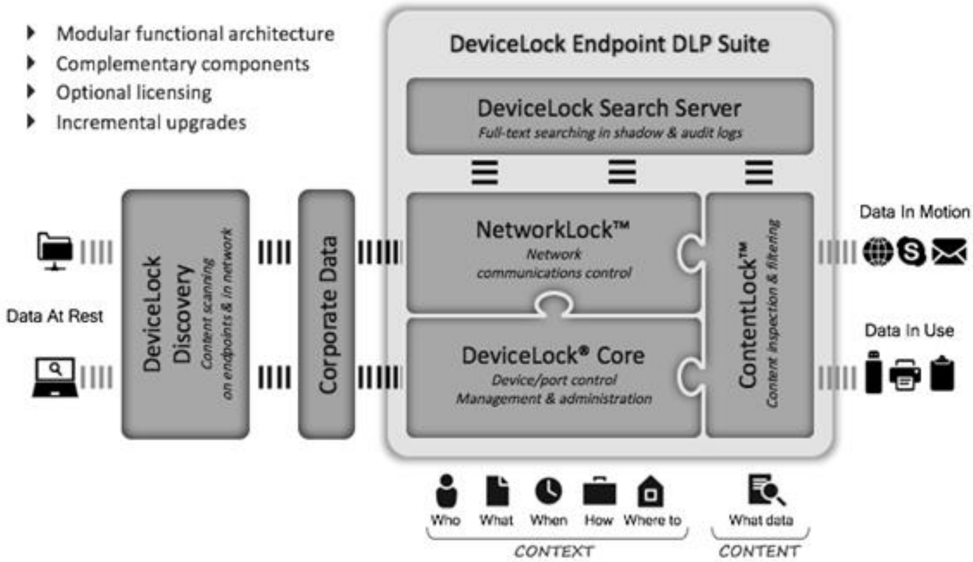
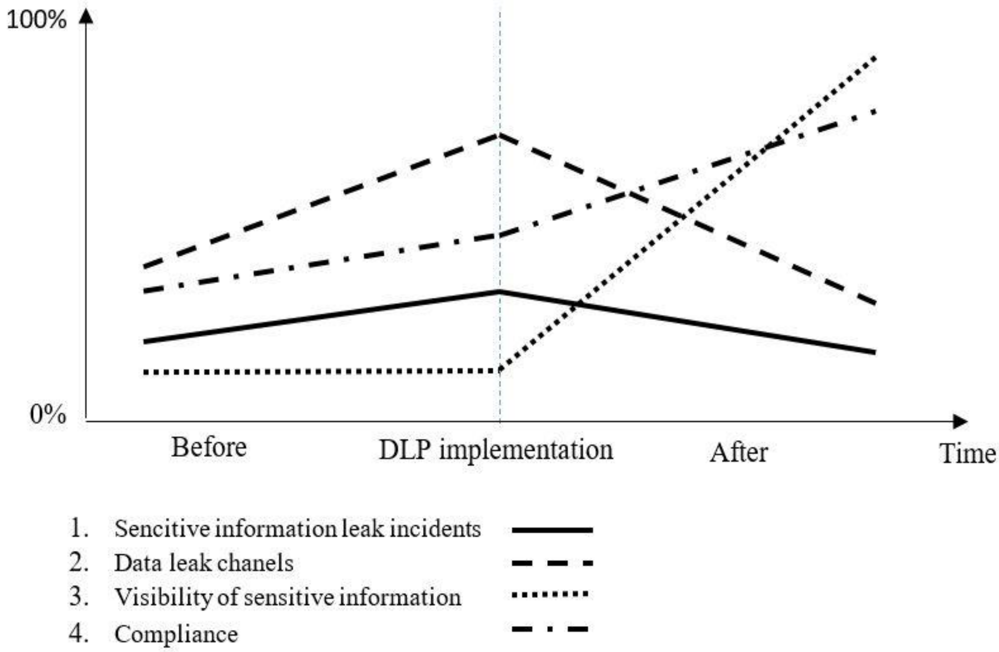


Figure 2. DeviceLock Endpoint DLP Suite

The implementation of DLP in organization can effectively protect the sensitive information from inside and outside threats and additionally brings the following results – Figure 3 [19]:

- Reducing the sensitive information leak incidents;
- Limiting data leak channels;
- Increasing the visibility of sensitive information, by the discovery function of the DLP (Data-in-Rest);
- Improving compliance with the internal security policies, legal regulations and privacy directives;



***F**igure 3. Generalized results from DLP implementation*

5 Conclusions

Our holistic approach addresses not only the organizations data. It inludes profiling of the participants in the information process, based on what data they are using or processing, how and where they are using or processing it. It allows us to identify the critical places, staff and activities, thus discovering the weak points in the organization. After risk assessment is done, appropriate measures may be taken, by deploying appropriate security controls as Data Leak Protection (DLP) systems, which are designed to stop the data leakage from the inside to the outside. It can not only limit the data leaks, but improve the compliance with the internal security policies, legal regulations and privacy directives.

6 Acknowledgement

This research was supported by the Bulgarian FNI fund through the project "Conceptual Modeling and Simulation of Internet of Things Ecosystems (KoMEIN)".

References

[1] Rhodes M., (2013). Information Security The Complete Reference, 2nd Edition, The McGraw-Hill.

[2] Ernst and Young (October 2011). Data Loss Prevention - Insights on governance, risk and compliance, DLP conceptual model_EY_Data_Loss_Prevention.pdf
[3] Ernst and Young (June 2011).The evolving IT risk landscape - The why and how of IT Risk Management today.
[4] Hintzbergen, J, Hintzbergen K. (2010) . Foundations of Information Security Based on ISO27001 and ISO27002. Van Haren.
[5] ISO 27001 official page: https://www.iso.org/isoiec-27001-information-security.html, May 2017
[6] IT Governance Institute. COBIT Security Baseline: An Information Survival Kit. 2nd ed. IT Governance Institute, 2007
[7] COBIT resources: http://www.isaca.org/COBIT/Pages/default.aspx, May 2017
[8] NIST Special Publications (800Series): http://www.csrc.nist.gov/publications/PubsSPs.html, May 2017
[9] Gramm-Leach-Bliley Act (GLBA) resources: https://www.ftc.gov/tips-advice/business-center/privacy-and-security/gramm-leach-bliley-act, May 2017
[10] Anand, S. Sarbanes-Oxley Guide for Finance and Information Technology Professionals. Wiley, 2006
[11] Sarbanes-Oxley Act (SOX resources: https://www.sec.gov/about/laws/soa2002.pdf
[12] Beaver, K, Herold R.. The Practical Guide to HIPAA Privacy and Security Compliance. 2nd ed. Auerbach, 2011
[13] PCI Security Standards: https://www.pcisecuritystandards.org/pci_security/, May 2017
[14] EU General Data Protection Regulation official page: http://ec.europa.eu/justice/data-protection/reform/index_en.htm, May 2017
[15] IEEE Std 1471, IEEE Recommended Practice for Architectural Description of Software-Intensive Systems (2000).
[16] ISO/IEC/IEEE 42010:2011 - Systems and software engineering - Architecture description Iso.org. 2011-11-24. Retrieved 2013-08-06
[17] Hilliard R., Aspects, Concerns, Subjects, Views.. OOPSLA'99 Workshop on Multi- Dimensional Separation of Concerns in Object-Oriented Systems
[18] Tudor, Killmeyer J., (2001) Information security architecture: an integrated approach to security in the organization, CRC Press LLC.
[19] Gaydarski I, Minchev Z., (2017) Conceptual modeling of information security system and its validation through DLP systems, 9th International Conference on Business Information Security (BISEC-2017), 18th October 2017, Belgrade, Serbia
[20] DeviceLock www.devicelock.com/products/, August 2017

decision support
control systems engineering
data mining
algorithms
human cognition
intelligence
data management
knowledge
parallel processing
big data security
systems analysis
big data
data analysis
knowledge management
intelligent control systems
artificial intelligence
innovations
operations research
distributed processing
process control
data engineering
data processing
soft computing

ISSN: 2367 - 6450

Institute of Information and Communication Technologies - Bulgarian Academy of Sciences
"John Atanasoff" Union on Automatics and Informatics, Bulgaria