Ss. Cyril and Methodius University in Skopje
FACULTY OF COMPUTER SCIENCE AND ENGINEERING

icit 2016

Proceedings of the 13th International Conference for Informatics and Information Technology

Held at Hotel Molika, Bitola, Macedonia
22-24th April, 2016

Editors:
Sasho Gramatikov
Magdalena Kostoska
Miroslav Mirchev

ISBN 978-608-4699-06-4

Conference on Informatics and Information Technology 2016
Web-site: http://ciit.finki.ukim.mk
Email: ciit@finki.ukim.mk

Publisher:
Faculty of Computer Science and Engineering, Skopje, Macedonia
Ss. Cyril and Methodius University – Skopje, Macedonia
Address: Rugjer Boshkovikj 16, P.O. Box 393, 1000 Skopje, Macedonia
Web-site: http://www.finki.ukim.mk/
Email: contact@finki.ukim.mk

Proceedings Editors:
Sasho Gramatikov
Magdalena Kostoska
Miroslav Mirchev

Technical editing: Sasho Gramatikov, Magdalena Kostoska, Miroslav Mirchev
Cover page: Vangel Ajanovski

Total print run:150
Printed in Skopje, Macedonia, 2016

ISBN: **978-608-4699-06-4**

CIP - каталогизација на публикација
Народна и универзитетска библиотека „Св.Климент Охридски“, Скопје

004.7:621.39(062)
004(062)

PROCEEDINGS of the 13th Conference on Informatics and Information Technology (13; 2016; Bitola)
Proceedings of the 13th Conference on Informatics and Information Technology: CIIT 2016, April, 22-
24, Molika, Bitola, Macedonia / editors Sasho Gramatikov, Magdalena Kostoska and Miroslav Mirchev. - Skopje : Faculty of Computer Science and Engineering, 2015. - 351 стр. : граф. прикази ; 30 см

Библиографија кон трудовите

ISBN 978-608-4699-06-4
1. Gramatikov, Sasho [уредник] 2. Kostoska, Magdalena [уредник] 3. Mirchev, Miroslav [уредник]

Preface

The annual Conference of Informatics and Information Technology was traditionally held for the thirteenth time in Bitola, Macedonia from 22th to 24th of April, 2016. The conference was organized by the Faculty of Computer Science and Engineering, Ss. Cyril and Methodius in Skopje, Macedonia.

The conference was comprised of five sessions including one student session. There were in total thirty regular presentations and forty five participants. As a key note speaker we had the pleasure to have Dr. Massimo Camplani from the University of Bristol, UK who gave a talk titled "Smart-home for health-care" and as an invited speaker Dr. Hristijan Gjoreski from the Jožef Stefan Institute, Ljubljana, Slovenia, who had a presentation on the topic "Ambient Intelligence (AmI) applications in healthcare domain".

The format of the conference allowed the participants to attend most of the talks that covered a diverse spectrum of research areas. The primary goal of the conference was to bring together researchers from different ICT areas, to facilitate multidisciplinary collaboration and promote young researchers.

The Editors,

Sasho Gramatikov, PhD

Magdalena Kostoska, PhD

Miroslav Mirchev, PhD

Editors

Sasho Gramatikov, PhD
Magdalena Kostoska, PhD
Miroslav Mirchev, PhD

Organizing Committee

Bojana Koteska, MSc
Monika Simjanoska, MSc

Program committee

Aleksandar Shurbevski, PhD, Kyoto University, Japan
Hristijan Gjoreski, PhD, Jozef Stefan Institute, Slovenia
Stojan Trajanovski, PhD, TU Delft/NAS group, Netherlands
Bojan Marinkovic, PhD, Mathematical Institute of the Serbian Academy of Sciences and Arts, Serbia
Haris Gavranović, PhD, International University of Sarajevo, BiH
Antun Balaz, PhD, Institute of Physics, Belgrade, Serbia
Sahra Sedigh Sarvestani, PhD, Missouri University of Science and Technology, USA
Boro Jakimovski, PhD, Ss. Cyril and Methodius University in Skopje, Macedonia
Dejan Gjorgjevikj, PhD, Ss. Cyril and Methodius University in Skopje, Macedonia
Kire Trivodaliev, PhD, Ss. Cyril and Methodius University in Skopje, Macedonia
Marija Mihova, PhD, Ss. Cyril and Methodius University in Skopje, Macedonia
Dimitar Trajanov, PhD, Ss. Cyril and Methodius University in Skopje, Macedonia
Ana Madevska Bogdanova, PhD, Ss. Cyril and Methodius University in Skopje, Macedonia
Lasko Basnarkov, PhD, Ss. Cyril and Methodius University in Skopje, Macedonia
Ivica Dimitrovski, PhD, Ss. Cyril and Methodius University in Skopje, Macedonia
Goce Ristanoski, PhD, National ICT, Australia
Andreja Naumoski, PhD, Ss. Cyril and Methodius University in Skopje, Macedonia
Sasko Ristov, PhD, Ss. Cyril and Methodius University in Skopje, Macedonia
Georgina Mircheva, PhD, Ss. Cyril and Methodius University in Skopje, Macedonia
Suzana Loshkovska, PhD, Ss. Cyril and Methodius University in Skopje, Macedonia
Biljana Stojkoska, PhD, Ss. Cyril and Methodius University in Skopje, Macedonia
Igor Mishkovski, PhD, Ss. Cyril and Methodius University in Skopje, Macedonia
Vesna Dimitrova, PhD, Ss. Cyril and Methodius University in Skopje, Macedonia
Vesna Dimitrievska Ristovska, PhD, Ss. Cyril and Methodius University in Skopje, Macedonia
Adrijan Božinovski, PhD, University American College Skopje, Macedonia
Vangel Ajanovski, PhD, Ss. Cyril and Methodius University in Skopje, Macedonia
Verica Bakeva, PhD, Ss. Cyril and Methodius University in Skopje, Macedonia
Nevena Ackovska, PhD, Ss. Cyril and Methodius University in Skopje, Macedonia
Simona Samardjiska, PhD, Ss. Cyril and Methodius University in Skopje, Macedonia
Zaneta Popeska, PhD, Ss. Cyril and Methodius University in Skopje, Macedonia
Aleksandra Popovska-Mitrovikj, PhD, Ss. Cyril and Methodius University in Skopje, Macedonia

Table of Contents

Talks

Papers

TALKS

Smart-Home for health-care: the Sphere Project

Massimo Camplani, PhD
University of Bristol, United Kingdom

Today's aging population and the rise in chronic health conditions is precipitating a shift toward empowering people to manage their care and well-being at home. In particular, advances in ambient assisted living (AAL) are providing resources that improve patients' lives, as well as informing necessary interventions from relatives, caregivers, and healthcare professionals. For smart home systems to be effective in detecting and managing health conditions, they must provide meaningful clinical data but also be desirable to their domestic users. This talk provides an overview of the the vision and the progresses of the Sensor Platform for Healthcare in a Residential Environment (SPHERE) project. The main aim of this project work is to build a generic platform that fuses complementary sensor data (wearable, video and environmental sensors) to generate rich datasets that support the detection and management of various health conditions. A particular emphasis will be given to video monitoring infrastructure of the SPHERE sensing platform.

Ambient Intelligence (AmI) applications in healthcare domain

Hristijan Gjoreski, PhD
Jožef Stefan Institute, Sloveniam

Ambient Intelligence (AmI) is an emerging discipline that brings intelligence to our everyday environments and makes those environments sensitive to us. The basic idea behind AmI is that by enriching an environment with technology (mainly sensors and devices interconnected through a network), a system can be built to take decisions to benefit the users of that environment based on real-time information gathered and historical data accumulated. AmI is a fast-growing multi-disciplinary area which builds upon the advances in multiple areas in computer science, such as: artificial intelligence, sensors, networks, pervasive/ubiquitous computing, and human computer interfaces. In this talk, four applications of AmI in healthcare domain will be presented, including:

- Human Activity Recognition. A system that recognizes the activities of a person that is wearing accelerometer(s) will be presented. The system recognizes the activities of the user by applying machine learning techniques to analyze the sensors data. Comparison of several algorithms on several sensor body locations will be presented (chest, thighs, ankles, wrists, etc.).

- Human Fall Detection. This application is about detecting when a person falls. This is achieved by analyzing data provided by wearable sensors, mainly accelerometers. Various algorithms and sensor locations will be discussed and experimental results will be presented. This application is especially useful for elderly people, which have tendency to fall, and sometimes are hurt and cannot call for a help. Having such system, an automatic call can be stablished once a fall is detected.

- Human Energy Expenditure Estimation. This application is about automatically estimating the energy expenditure of a person. A system will be presented that analyzes data from several wearable sensors (accelerometer, heart rate, galvanic skin response, etc.) and by using machine learning techniques it estimates the energy expenditure.

- Human Stress Detection. A system that continuously detects the stress level of a person will be presented. The system uses machine learning methods to analyze data from a wristband device and to detect the stress level.

FULL PAPERS

Overview of Decision Algorithms for HTTP Adaptive Streaming

Sasho Gramatikov
Faculty of Computer Science and Engineering, Skopje,
University "Ss. Cyril and Methodius", Skopje, Macedonia
Email: sasho.gramatikov@finki.ukim.mk

***Abstract*—Streaming video contents is one of the most rapidly growing services on the Internet. Because of the wide acceptance of this service, the heterogeneity of the devices for playing the videos and the limited quality of service that can be provided by the Internet, the necessity for streaming the videos with different qualities has become inevitable. Therefore, instead of streaming the entire video with same quality, the HTTP adaptive streaming approach encodes the videos with different qualities and then divides them in segments with duration of a few seconds. During the streaming, after each segment, the clients issue requests for a segment with quality that is appropriate for the available bandwidth. The quality of experience of the users largely depends on the quality level of the received video, the frequency of quality switches and the continuity of the video. Therefore choosing the appropriate quality of video is of a crucial importance for the adaptive streaming. This paper gives an overview of the most commonly used algorithms for adapting the streaming rate in HTTP adaptive streaming.**

Keywords—Video, Adaptive streaming, HTTP, Decision Algorithms, Quality of Experience

I. Introduction

The advances of the network technologies made the Internet suitable environment for offering Video on Demand and Live video streaming services with high quality. Therefore, the Video services dominate with the amount of traffic transferred over the modern Internet. However, the Internet cannot guarantee the Quality of Service (QoS) to meet the traffic the growing expectations of the users and their Quality of Experience (QoE). Although the existing transport and application protocols (RTP, RTCP, RTSP) give satisfactory results, they cannot overcome the issue of interruptions during watching the videos caused by the limited bandwidth of certain users.

One of the emerging solutions that has been widely accepted in the video market is the HTTP Adaptive Streaming (HAS). The basic idea behind this solutions lays in the fact that, most of the users would not mind changing the quality of the received video as long as there are no interruptions in the service. Hence, in the approach of HAS, each video is encoded with different qualities, ranging from the lowest to the highest, and each video quality is divided into chunks with duration of a few seconds and placed on a web server. In the process, a special meta-data manifest file is generated, containing all the relevant information related to the different qualities and the locations of the chunks. Whenever a client requests a video, it first receives the manifest, and than, based on the available bandwidth, requests a video chunk with suitable quality. Initially, the client downloads the chunks greedy, until the play buffer is loaded with sufficient number of chunks for watching the video. Afterwards, every subsequent chunk is requested based on the available bandwidth of the link between the client and the server. If at any time during the video transfer the quality of the bandwidth drops, the client requests a video with lower quality. The smaller video chunk requires less bandwidth, and therefore, prevents interruption of the stream for the price of worse received quality. If during watching the video the client decides to play back part of the video that was previously played with lower quality, the video can be played with higher quality provided that there are better network conditions.

The first request for the manifest, as well as the request for video chunks with different qualities are made using the same GET HTTP method used for web pages. This feature imposes the advantage of using ordinary web server for hosting the videos. Another advantage of HAS is the use of the HTTP as application protocol since it uses the port 80, which is allowed even in the networks with most restrictive polices. Furthermore, the decision on the video quality requested is taken by the client, thus alleviating the server from keeping track of the sessions of the clients. Therefore, instead of dedicated video server, an ordinary web server can be used for providing video service with adaptive quality. The main disadvantage of the DASH is that it requires storage space for hosting many chunks of different qualities of the same video.

Since the consumption of videos has not been limited to computer only, but also has been extended to massive use by mobile devices, the HAS approach has been implemented in different commercial players by many leading companies. They use the same principles, but define different structure and extension of the manifest file and the chunks of video contents. Thus, Apple develops the HTTP Live Streaming (HLS) [1] which works with their Quick Time player, Microsoft develops the HTTP Smooth Streaming (HSS) [2] recognized by their Silverlight application framework and Adobe develops the HTTP Dynamic Streaming (HDS) [3] that can be played on their Adobe Flash player. Other commercial players that implement these version of HAS also exist. Mostly, they are developed by companies offering on-line video services such as Youtube, Netflix, JWplayer, Wowza, MediaElement, Bitdash etc.

The main disadvantage of having many commercial imple-

mentations with similar principles is that they are not open-source and they all use different manifest file and segment formats with different duration. Therefore, in 2011 the HAS is standardized by MPEG into the open standard MPEG Dynamic Adaptive Streaming over HTTP (MPEG-DASH) [4]. The standard defines the format of the manifest file called Media Presentation Description (MPD), containing information about the quality versions of the videos, the URLs of their segments and other important parameters. The DASH standard is currently widely accepted by a large community of leading streaming companies which formed the DASH Industry Forum (DASH-IF) [5]. The forum works on continuous improvement and implementation of the standard towards creating uniformly accepted product. One of the important outcomes of the forum is the development of an open-source software packet containing JavaScript based DASH player and other useful tools for adaptation of the videos for DASH streaming. It is also worth to mention that most of the commercial players implement the DASH protocol.

II. Conventional decision algorithms for adaptive streaming

A key component of the HAS is the algorithm that the client uses for choosing the most adequate quality of every segment it request based on the network conditions and the status of its player. Although the main objectives of the HAS are the uninterrupted service and offering the best quality possible, their fulfilment leads to frequent changes of quality levels. In order to keep the clients satisfied, the algorithms should also aim to less frequent changes of quality levels, and when they are inevitable, they should achieve more even change without moderate changes of quality level. The basic parameter for decision making is the quality of the link from the client to the server, i.e., the available bandwidth calculated as the ratio of the downloaded segment size and its download time. Another important parameter is the buffer level, defined as the amount of playable video already downloaded from the server. The higher buffer levels are promising for requesting videos with better quality.

A. Rate adaptation algorithm

One of the simplest and conventional algorithms is the Rate Adaptation algorithm (RA) [6] where the authors use the download time of the last segment measured as the time difference between the first GET request for the segment and the time the segment is entirely downloaded as a main measure of the link quality. They define a ratio μ between the play time of the segment obtained from the manifest file and the segment fetch time and use it to take a decision for the quality level of the next segment.

Whenever $\mu > 1 + \epsilon$, i.e., the fetch time is shorter than the playback time for a predefined value, and the buffer level is higher than a predefined minimum value, the algorithm decides to request the next higher level. In the opposite case, when the ratio μ is lower than predefined critical value, the next requested level will be reduced. Unlike the previous case, where the level is increased step-wise, the reduction of the quality can be aggressive, depending on the measured segment fetch time. The new level is chosen in a way that it will not allow buffer underflow while it is being downloaded. In order to prevent buffer overflow, the algorithm also calculates the amount of idle time that it spends until it requests the next segment with the estimated quality. This value depends on the current buffer time, the predefined minimum buffer and the estimated quality level. While the algorithm provides smooth improvement towards better qualities, it may cause unnecessary fluctuations of the received quality due to short term link bandwidth variations.

B. Adaptive streaming of audiovisual content

To address the problem of frequent quality changes, the authors of [7] propose an algorithm for throughput estimation taking into consideration the past values of the real segment throughput T_e and estimated segment throughput T_e. At each step i, the estimated throughput of the i-th segment is calculated as $T_e(i) = (1 - \delta)T_e(i - 2) + \delta T_s(i - 1)$, where the parameter δ defines the influence of the real throughput in the calculation of the estimated throughput. Higher values of this parameter contribute to more accurate estimation, however, lower values contribute to smoother transition of the estimated values from step to step. For optimal experience, the authors define this parameter as a function of the normalized deviation of the real value relative to the estimated value. While the algorithm provides smooth playing on long term bandwidth changes, it still does not solve the problem of quality changes when there are shot term bandwidth variations.

C. Video quality adaptation scheme

Another, more advanced version of the RA algorithm is proposed in the video quality adaptation scheme for improving the QoE in HAS [8]. The key difference in this algorithm is that detecting different throughput of the downloaded segment from the current throughput is not enough for changing the quality level. The algorithm will increase/decrease the next quality level step-wise, only if the measured throughput is higher/lower than the currently downloaded segment and the buffer level is higher/lower than dynamically calculated buffer threshold B_{up}/B_{down}. The buffer threshold for upgrading the quality level B_{up} is calculated as the portion of the cumulative bit rates from the lowest to the rate one level higher than the current rate relative to all cumulative rates, multiplied by the maximum allowed buffer level B_{max}. The buffer threshold for downgrading the quality B_{down} is calculated similarly, with the difference that only the cumulative rates from the lowest to the current bit rates are considered. If the network conditions rapidly deteriorate, decreasing the quality level step-wise will deplete the buffer. Therefore, when the current buffer level is lower than the previous buffer level for predefined value, the algorithm aggressively lowers the quality of the requested segment.

III. Problems of conventional decision algorithms

The simulations and the experimental tests of the conventional algorithms give satisfactory results in variable network

scenarios. However, they are all conducted when a single client requests a video from the server. The analysis of multiple simultaneous request from different clients that share the same bottleneck link and compete for bandwidth show that the approach for measuring the TCP throughput by simply dividing the segment size by the fetch time lead to three problems: instability, unfairness and bandwidth underutilization [9]. The reason for these problems is that when the player is in steady state, its activity consists of consequent ON and OFF periods.

During the OFF periods, the player is requesting and downloading a segment, while in the OFF period it is waiting to issue new request after previously downloaded segment. If the player shares the same connection with other player their ON-OFF periods can be completely disjoint or they can overlap. In the first case, both players will receive the maximum available bandwidth due to the TCP nature of the connection. Both players overestimate their bandwidth because instead of receiving fair share of half of the capacity, they receive the entire link capacity. The instability comes to play when the ON periods star to overlap. In this situations, both players are downloading a segment, which will take longer time. Hence the players will estimate a lower throughput and switch to lower quality level. If the player calculates the idle time such that the next on state do not overlap with other on state of a player, then, after the segment is downloaded, it will estimate higher throughput. Thus, the player will frequently change the quality levels, leading to instability. The underutilization of the link is visible when the players can download a content with only a few qualities, e.g., low and high quality. If both players receive the content with different quality due to the fact that the link is not enough for simultaneous use of the high quality level, and then the algorithm of the player with higher quality mispredicts the throughput and switch to the low quality level, both players to use only the low profile. Thus, the available bandwidth will not be optimally utilised. The problems of unfairness, instability and underutilization can be caused by other factors in the real scenarios such as the number of competing players, the TCP dynamics, decision algorithms of the players and bandwidth fluctuations.

IV. Advanced decision algorithms

The problems incurred by the ON-OFF states of competitive players was a solid motivation for the work of many researches in the past years which resulted with more advanced algorithms with objective to provide stability, fairness and optimal bandwidth utilization.

A. FESTIVE

The Fair, Efficient and Stable adapTIVE (FESTIVE) algorithm proposed in [10] consists of three key steps executed periodically for each video segment. Since the algorithm is defined for the steady state of the player, initially, the player aggressively downloads the first segments with the same video quality in order to start playing the video.

The first step consists in bandwidth estimation. The same approach based on download time is used for calculating the throughput, with the difference that the estimated value is obtained as harmonic mean of the past 20 segments.

In the second step, called delayed bitrate update, the algorithm computes a reference bandwidth based on the results from the bandwidth estimation. This bandwidth is used to compute efficiency cost and stability cost. The efficiency cost for the next bitrate is best when the bitrate is equal to the reference bit rate. The stability cost is calculated based on the number of switches in a predefined time interval and on how close to the current bitrate is the reference rate. The algorithm gives more penalty when adding new level switch if there were previous switches within the predefined interval. The final cost is calculated by weighted sum of the stability and efficiency costs. The weights provide configurable decision making trading between the efficiency and stability. In the experimental results, the authors give precedence to the efficiency by factor of 12. The stability is further increased by gradual switch of bitrate level. When it comes to increasing the level, the bitrate k is allowed only after k steps. Contrary, when it comes to decreasing the bitrate, it is allowed after a single step.

In the last step, the algorithm uses randomized scheduling when choosing the time to request the next segment in order to avoid unfairness in the throughput estimation. The basic idea is choosing random critical buffer level at every decision step. If the current buffer level at the moment of decision is smaller than the randomly generated value, the next request will be scheduled immediately after the segment is downloaded. Otherwise, the request will be scheduled after the segment is downloaded and the buffer consumes the extra time relative to the critical randomly chosen buffer level.

B. PANDA

Although the previous algorithm has a more complex mechanism for estimating the bitrate, it still bases its measurements on the TCP download time of the segments during the ON periods. The downside of these measurements is that the system is not aware of the network conditions during the OFF periods. A solution for the ambiguous knowledge of the competing players in the ON-OFF periods is proposed in the Probe AND Adapt (PANDA) algorithm [11].

The authors use a more proactive mechanism for bandwidth measurement by actively probing the server with consequent increments of the sending data bitrate. The client simultaneously measures the TCP throughput using the conventional method and the target throughput by incremental probing. Then, it compares the difference and, if the conventional TCP rate the is higher than target bit rate, it estimates that the TCP overestimates the link quality in the ON periods and therefore the player uses quality level that is close to the target throughput. The opposite case, when the TCP throughput is lower than the target throughput, indicates link congestion and the target throughput is reduced proportionally to the obtained difference. The PANDA probing mechanism is similar to the TCP's congestion control, with the main difference that TCP uses round-trip time or packet losses to detect the congestion, while PANDA uses the reduction of the measured TCP throughput.

C. ELASTIC

The author of the fEedback Linearization AdaptIve STreamIng Controller(ELASTIC) [12] avoid the problem of ON-OFF periods by eliminating the idle waiting of the players to request a new segment. Thus, the players are constantly receiving segments and obtain more accurate measurements of the available bandwidth when more players are competing for the same link. The elimination of the OFF periods imposes that the algorithm has to decide only the quality level to request upon a segment complete download.

The inputs required for taking this decision are the estimated throughput, the current buffer length and a set-point buffer level that has to be achieved. By reaching the set-point buffer level, the requested video quality level matches the available bandwidth. In the proposed solution, the buffer length is modelled as non-linear differential equation. The change of the buffer level is defined as the difference of the buffer filling rate, expressed as the ratio of the estimated throughput obtained as harmonic mean from the past steps and the requested bit rate, and the buffer drain rate. The drain rate is a binary function that has value 1 if the player is playing or 0 otherwise.

Since the quality level of the requested video appear as a nominator in the dynamic of the buffer level, the authors use feedback linearisation technique to determine the next quality level that will lead the buffer to the set-point level and optimally use the available bandwidth.

D. Q-learning based algorithm

Another completely different approach is used in the Q-learning based algorithm [13]. This algorithm implements a learning technique where an agent has limited knowledge about the environment. The agent interacts with the environment, and for every action it takes, the state of the environment changes. The agent gets numerical reward for every action and state, which is used for the learning process. The moving gear of the algorithm is the Q-function which is used to measure the quality of the action taken when the system is in certain state. The function also defines a factor that determines the importance of the newly obtained data from the environment relative to the old values, and a discount factor which defines the importance of the future rewards. The function is calculated as a weighted sum of three aspect of the streaming quality: the current segment quality level, the oscillation of the quality levels and the buffer starvation.

V. Conclusions

This paper gives an overview of the existing algorithms for rate adaptation of the HTTP adaptive streaming. The algorithms' main goals are to achieve uninterrupted video streaming with the maximum quality of the segments that the network conditions allow, and to reduce the frequency of quality switches during the streaming. In the paper, the algorithms are classified as conventional and advanced based on whether they take the ON-OFF periods during the streaming of the segmented videos. The conventional algorithms measure the available bandwidth based on the time it takes to download a segment with know quality during the ON period when only one player is requesting a video from the server. However, in the real case scenarios, there are more players simultaneously competing for a shared link which cause unfairness, underutilization and instability of the streaming. The advanced algorithms use different approaches to address these issues and achieve real estimation of the available bandwidth and choose the quality level that will lead to stable streaming.

The video streaming on the Internet is a rapidly growing marked, and the HAS is confidently taking a leading position. In order to provide the best quality of experience to the users, many researches still take the design of advanced algorithms as a serious challenge.

Acknowledgment

The author thanks the Faculty of computer science and engineering at the Ss. Cyril and Methodius University in Skopje, under the SADMM ("Streaming and Analysis Decision Making for Multimedia Contents") project for financial support.

References

[1] (2016) Apple HTTP Live Streaming. [Online]. Available: https://developer.apple.com/streaming

[2] (2016) Microsoft Smooth Streaming. [Online]. Available: http://www.iis.net/downloads/microsoft/smooth-streaming

[3] (2016) Adobe HTTP Dynamic Streaming. [Online]. Available: http://www.adobe.com/products/hds-dynamic-streaming.html

[4] A. Vetro and I. Sodagar, "Industry and Standards The MPEG-DASH Standard for Multimedia Streaming Over the Internet," *IEEE MultiMedia*, vol. 18, no. 4, pp. 62–67, 2011.

[5] (2016) DASH Industry Forum. [Online]. Available: http://www.dashif.org

[6] C. Liu, I. Bouazizi, and M. Gabbouj, "Rate adaptation for adaptive http streaming," in *Proceedings of the Second Annual ACM Conference on Multimedia Systems*, ser. MMSys '11. New York, NY, USA: ACM, 2011, pp. 169–174. [Online]. Available: http://doi.acm.org/10.1145/1943552.1943575

[7] "Adaptive streaming of audiovisual content using MPEG DASH," *IEEE Transactions on Consumer Electronics*, vol. 58, no. 1, pp. 78–85, 2012.

[8] S. Kim, D. Yun, and K. Chung, "Video Quality Adaptation Scheme for Improving QoE in HTTP Adaptive Streaming," 2016.

[9] S. Akhshabi, L. Anantakrishnan, A. C. Begen, and C. Dovrolis, "What Happens When HTTP Adaptive Streaming Players Compete for Bandwidth," *Proc. ACM Workshop on Network and Operating System Support for Digital Audio and Video (NOSSDAV'12)*, p. 6, 2012.

[10] J. Jiang, "Improving Fairness, Efficiency, and Stability in HTTP-based Adaptive Video Streaming with FESTIVE.pdf," pp. 97–108, 2012.

[11] Z. Li, X. Zhu, J. Gahm, R. Pan, H. Hu, A. C. Begen, and D. Oran, "Probe and adapt: Rate adaptation for HTTP video streaming at scale," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 4, pp. 719–733, 2014.

[12] L. De Cicco, V. Caldaralo, V. Palmisano, and S. Mascolo, "ELASTIC: A client-side controller for dynamic adaptive streaming over HTTP (DASH)," *2013 20th International Packet Video Workshop, PV 2013*, 2013.

[13] M. Claeys, S. Latré, J. Famaey, T. Wu, W. V. Leekwijck, and F. D. Turck, "Design of a Q-Learning-based Client Quality Selection Algorithm for HTTP Adaptive Video Streaming," *Aamas 2013*, pp. 30–37, 2013.

eHealth platform prototype for real-time biosensor data transfer

Magdalena Kostoska, Bojana Koteska, Monika Simjanoska, Martin Tashkoski
Faculty of Computer Science and Engineering
Ss. Cyril and Methodius University
Skopje, Macedonia
{magdalena.kostoska, bojana.koteska, monika.simjanoska}@finki.ukim.mk, tashkoskim@yahoo.com

Nevena Ackovska, Ana Madevska Boganova
Faculty of Computer Science and Engineering
Ss. Cyril and Methodius University
Skopje, Macedonia
{nevena.ackovska, ana.madevska.bogdanova}@finki.ukim.mk

Roman Golubovski
Faculty of Natural Sciences and Mathematics
Ss. Cyril and Methodius University
Skopje, Macedonia
roman.golubovski@t.mk

***Abstract*— The current state of the biosensors market offers a big spectra of technologies and prices. The professional biosensors used in professional facilities are still expensive. Our idea is to create a more accessible platform using less expensive biosensors. The main point is to offer the same functionalities of the professional platforms for better price – performance ratio. The proposed platform considers easily accessible biosensors which are used to monitor and process bio signals in real time and store the results in the cloud for further examination and processing.**

Keywords—biosensors; ehealth; platform; cloud

I. Introduction

In today's market we are faced with variety of biosensors. Their features vary in bio-impedance, communication protocol, power supply, price etc. The choice of appropriate biosensor is based on the context of the intended system and.

Our goal in platform usage is to monitor vital data signs and determine severity of patient state using triage protocol decision tree. One variation of the protocol is shown in Fig. 1. Based on the given decision tree the following parameters are important to determine the seriousness of the state: pulse rate, respiratory rate, SpO2 and blood pressure. Some of the recent researches also include the temperature parameter in the triage decision [1]. The idea is to determine severity of injuries in scenarios where medical aid is not always available or needs help by civilians or first aid medics. The scenarios include, but not limited to, mass disasters where communication can be limited. For that reason the communication protocols can vary.

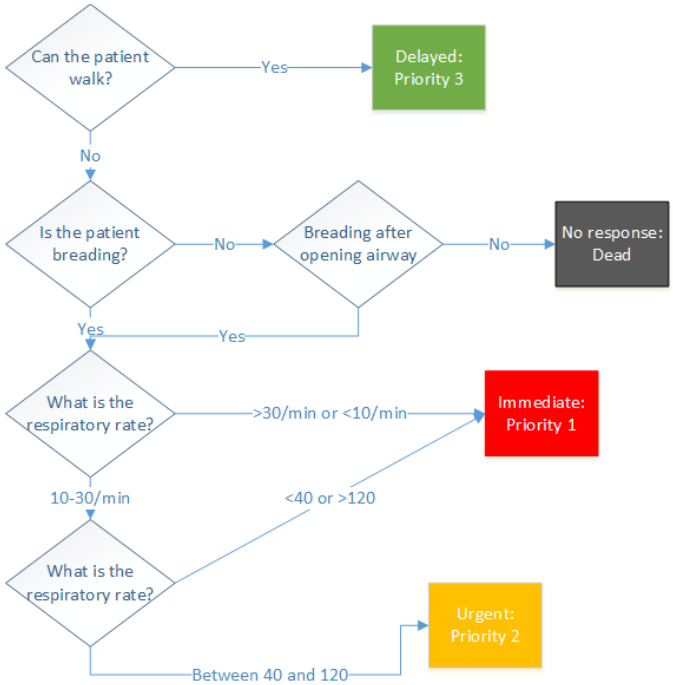


Fig. 1. Variation of triage decision tree

One of our goal was to determine the size of data needed to be transferred and to define the physical and logical constraints of the system. We have created controlled environment for initial testing.

The paper is organized as follows: in the next section we describe and compare available biosensors on the market, as well as published eHealth platforms. In Section III we described the used hardware components. Section IV gives overview of the platform. In section V we give the initial results and discuss them. Conclusive remarks are provided in Section VI.

II. Related Work

A. Biosensors

We have made a survey on biosensors available at the market. The results of the survey are shown in Table I.

TABLE I. OVERVIEW OF BIOSENSORS

Sensor	Communication	Bio-impedance	Power & Battery
Jawbone UP [2]	Bluetooth 4.0 BLE	Heart Rate, Respiration, Galvanic Skin Response (GSR), Skin Temperature	Battery life: Up to 7 days Charge time: ~60
Vitality [3]	N/A	ECG, Heart Rate and Heart Rate Variability, Blood Oxygen, Respiration, Core Body Temperature, Blood Pressure	N/A
Sensium Vitals [4]	Custom comm. hardware	Heart Rate, Respiration, Temperature	Battery life: Up to 5 days
HealthPatch MD [5]	Bluetooth 4.0 BLE	Single-Lead ECG, Heart Rate, Heart Rate Variability, Respiratory Rate, Skin Temperature, Body Posture, Fall Detection	Battery life: Greater than 2 days
QardioCore [6]	Bluetooth 4.0 BLE	ECG, Heart Rate, Heart Rate Variability, Respiratory Rate, Body Temperature, Galvanic Skin Response (GSR)	Battery life: Up to 2 days
PCARD [7]	Bluetooth 4.0	ECG, Skin Temperature	Battery life: Up to 3 days

B. eHealth Platforms

Numerous eHealth platform has been described in the literature. The initial platforms were more focused on quick interaction and alerting medical personnel [8] and introduction of distributed platforms in this area [9]. As the biosensors became more available on the market the researches started to target inclusion of eHealth platforms in personal homes [10, 11, 12].

With the availability of cloud platforms and their increased availability and popularity, the eHealth platforms started to use the benefits of the cloud. The researches include evaluation and in demonstration the usefulness of a cloud-based integrated health care system [13], development of intelligent management system based on cloud computing [14], increase of Quality of Service (QoS) using scalability in cloud [15] and development and integration of time critical eHealth services with DACAR platform [16].

III. Hardware Description

In this sections we will describe the hardware elements we use.

A. Cooking Hacks - e-Health Sensor Platform v2.0

The e-Health Sensor Shield V2.0 [17] created by Cooking Hacks allows Arduino and Raspberry Pi to interact with the shield and be used for biometric and medical applications where body monitoring is needed by using different sensors. It allows monitoring in real time the state of a patient and/or getting sensitive data in order to be subsequently analyzed for medical diagnosis. Biometric information gathered can be wirelessly sent using any of the 6 connectivity options available: Wi-Fi, 3G, GPRS, Bluetooth, 802.15.4 and ZigBee depending on the application.

1) Pulse and Oxygen in Blood Sensor (SPO2)

This sensor is used for noninvasive method of indicating the arterial oxygen saturation of functional hemoglobin. Oxygen saturation is defined as the measurement of the amount of oxygen dissolved in blood, based on the detection of oxygenated Hemoglobin (HbO2) and Deoxyhemoglobin (Hb). Two different light wavelengths are used to measure the actual difference: Deoxygenated hemoglobin (Hb) has a higher absorption at 660 nm and oxygenated hemoglobin (HbO2) has a higher absorption at 940 nm. Then a photo-detector perceives the non-absorbed light from the LEDs to calculate the arterial oxygen saturation. Fig. 2 depicts the usage of the sensor.

2) Electrocardiogram Sensor (ECG or EKG)

The Electrocardiogram Sensor (ECG or EKG) in this set is a single-lead ECG which uses 3 ECG electrodes. These basic leads acquire enough information for rhythm-monitoring, but are insufficient for determination of ST elevation (since there is no lead that gives information about the anterior wall). Fig. 3 depicts the usage of the ECG sensor.

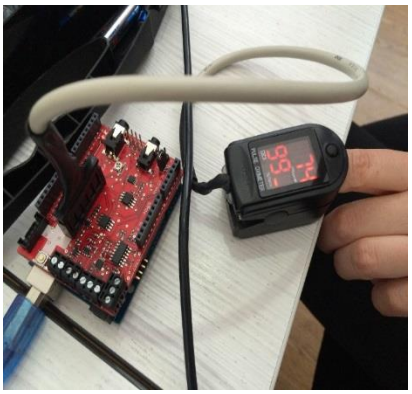


Fig. 2. Pulse/SpO2 sensor

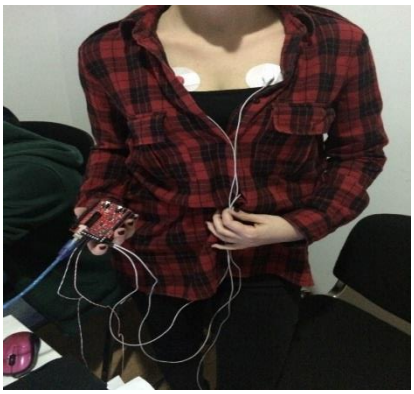


Fig. 3. ECG sensor

3) Airflow (breathing) or nasal / mouth airflow sensor

The Airflow (breathing) or nasal / mouth airflow sensor represents a single channel thermocouple sensor with a set of two prongs which are placed in the nostrils. The stay-put prongs (adjusted using flexible thread which fits behind the ears) position the sensor in the airflow path. Fig. 4 shows the usage of the airflow sensor.

Fig. 4. Airflow sensor

4) Body Temperature sensor

Body Temperature sensor is actually used by measuring a voltage and relating that to what the operating temperature of the sensor must be. The precision of the sensor can be improved by a calibration process. Calibration is a process of measuring voltage and resistance real values. In case of this sensor, the configuration file (library) contains default values for voltage (RefTension) and resistance (Rc, Ra, Rb). By measuring this values with multimeter and modify the library, greater accuracy can be obtained.

5) Patient position sensor (accelerometer)

The Patient position sensor (accelerometer) is used for monitoring of different patient positions (standing/sitting, supine, prone, left and right). This accelerometer is packed with embedded functions with flexible user programmable options, configurable to two interrupt pins. The accelerometer has user selectable full scales of ±2g/±4g/±8g with high pass filtered data as well as non-filtered data available real-time.

6) The Galvanic Skin Response Sensor (GSR - Sweating)

Skin conductance is a method of measuring the electrical conductance of the skin (which varies with its moisture level). Skin conductance is used as an indication of psychological or physiological arousal. The Galvanic Skin Response Sensor (GSR - Sweating) measures the electrical conductance between 2 points. It represents a type of ohmmeter. This sensor's precision can be also improved by calibration using a multimeter (since it is based on measuring a voltage).

7) Glucometer Sensor

Glucometer Sensor is used for determination of the approximate concentration of glucose in the blood by using blood sample. It uses disposable test strips that the meter reads and uses to calculate the blood glucose level in mg/dl or mmol/l.

8) Muscle/electromyography sensor (EMG)

Muscle/electromyography sensor (EMG) measures the electrical activity of skeletal muscles. It detects electrical potential generated by muscle cells when these cells are electrically or neurologically activated. This sensor uses three leads (MID, END and GND). It is designed by Advancer Technologies.

B. Arduino Uno

Arduino Uno microcontroller board is based on the ATmega328 (datasheet). The Uno board is the first in a series of USB Arduino boards, and it is used as reference model for the Arduino platform. It has 14 digital input/output pins, 6 analog inputs, a 16 MHz crystal oscillator, a USB connection, a power jack, an ICSP header, and a reset button. This microcontroller specifications are given in Table II.

The Uno can be programmed with the Arduino Software (IDE). The ATmega328 comes preprogrammed with a bootloader. It communicates using the original STK500 protocol. The bootloader can be bypassed and the microcontroller can be programed through ICSP (In-Circuit Serial Programming) header using Arduino ISP or similar.

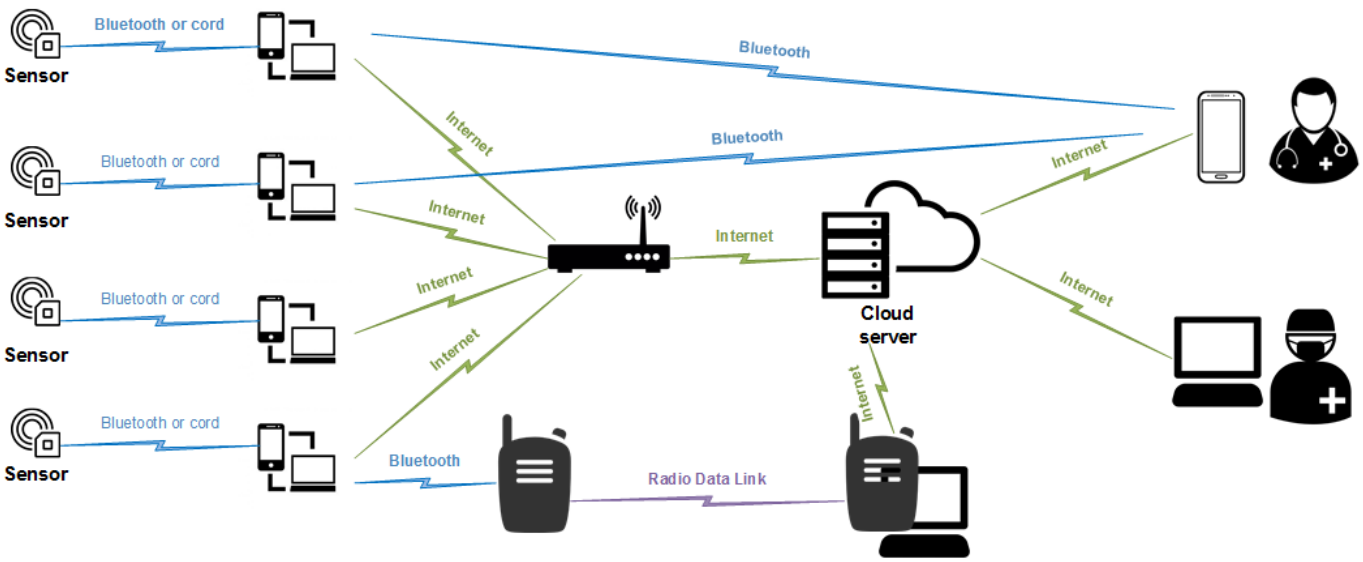


Fig. 5. eHealth platform concept

TABLE II. ARDUINO UNO SPECIFICATIONS

Characteristic	Value
Microcontroller	ATmega328P
Operating Voltage	5V
Input Voltage (recommended)	7-12V
Input Voltage (limit)	6-20V
Digital I/O Pins	14 (of which 6 provide PWM output)
PWM Digital I/O Pins	6
Analog Input Pins	6
DC Current per I/O Pin	20 mA
DC Current for 3.3V Pin	50 mA
Flash Memory	32 KB
SRAM	2 KB
EEPROM	1 KB
Clock Speed	16 MHz

IV. PLATFORM DESCRIPTION

The general concept of the platform consists of several elements: biosensors, local devices for communication with the sensors, connectivity of these devices to the cloud server and real-time monitoring of the available data by medical personnel. In general case, where Internet is available, the data is transmitted to the cloud and then distributed to the available doctors. In cases where Internet is not available the bio-data are transferred directly to the available medical personnel on site using Bluetooth technology. In cases of available HF radios with data link, these radios can also be used for data transfer. Fig. 5 depicts the general concept of the platform.

The solution can be adapted three different general scenarios, shown in Table III. The usage of devices and communication protocols is depicted in Fig. 6.

TABLE III. COMMUNIATION USAGE IN DIFFERENT SCENARIOS

Scenarios	Communication
Scenario 1: Internet available	**From biosensors to local devices**: Bluetooth or serial connection **From local devices to cloud**: Internet **From cloud to medics**: Internet
Scenario 2: No Internet available	**From biosensors to local devices**: Bluetooth or serial connection **From local devices to medics**: Bluetooth
Scenario 3: HF radios with data link available	**From biosensors to local devices**: Bluetooth or serial connection **From local devices to cloud**: HF radio data link **From cloud to medics**: Internet or HD radio link

Every type of communication impose some limits in transfer. While the Bluetooth connection has certain limitations, the HF data link has lowest data rate. The standard HF radios work with low bandwidth, use simplex mode and can be used as broadcast medium. Applications that use HF radios for communication should be able to perform data compression, group data (for minimizing the turnaround time) and avoid repeat transmissions, except when data is lost. The bandwidth of the radios vary between 256Kb/s (older models) up to 1 Mb/s. Due to the limitations in the link, the amount of data transfer is limited in scenario 3.

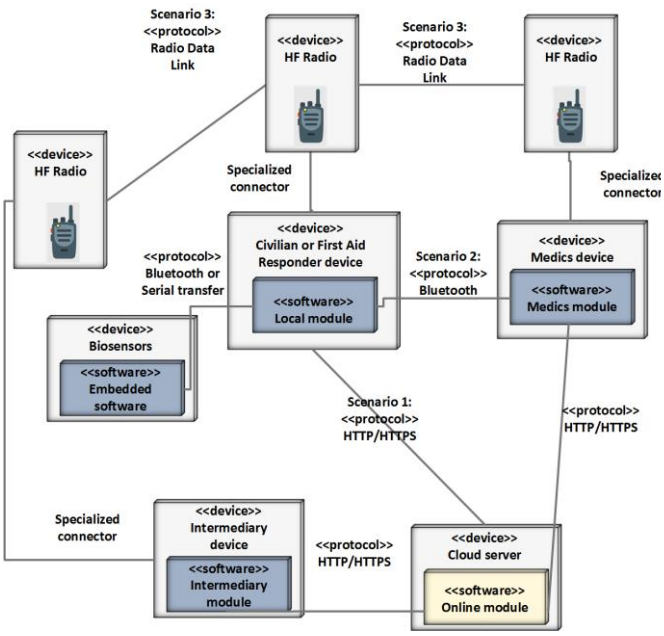


Fig. 6. Communication between elements in all scenarios

V. Results & Discussion

We set out initial platform using the Tonido cloud system [18], which represent a platform for private cloud. It offers access to files from a web browser, smartphone, tablet or DLNA enabled devices.

Our goal was to test the size of data that will be generated. We started by generating data using the SPO2 sensor with delay of 500ms (2Hz) between reads and the average result we obtained was data growth 2KB/minute. We also tested the same sensor with delay of 1ms (1000Hz) between reads and the data growth was on average 130KB/minute. We have also tested the size of data for generated by airflow sensor with delay of 1ms (1000Hz) between reads and we obtained average data growth of 40KB/minute.

We established that wearing the standard airflow sensor in many situations won't be practical, but this data can be extracted from raw data of ECG sensor. Fig. 7 shows the difference between raw and filtered ECG signal. We have also tested the size of data generated by ECG sensor and we obtained the following results:

- Readings with delay of 8ms (125Hz) between reads give average data growth of 30KB/minute
- Readings with delay of 5ms (200Hz) between reads give average data growth of 50KB/minute
- Readings with delay of 2ms (500Hz) between reads give average data growth of 110KB/minute
- Readings with delay of 1ms (1000Hz) between reads give average data growth of 180KB/minute

According to the results the total size of data using for vital signals (per person) has data growth of 40 KB/minute when using a sampling rate of 125 HZ and 310 KB/minute when using a sampling rate of 1000 HZ. According to the trauma surgeons a sample rate of 125 is adequate to use for triage decisions. Given the lowest connection available in the scenarios (usage of HF radios with only 256Kb/s data link) data from biosensors can be simultaneously transferred for more than 40 persons.

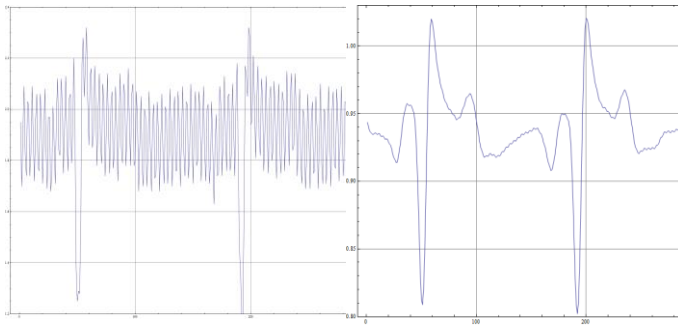


Fig. 7. ECG (raw signal – left, filtered signal – right) measurement

VI. Conclusion

The platform can be used for monitoring bio signals in home surrounding, as well as in situations where professional equipment is not available or there are no enough units available.

Given different scenarios the platform can support real-time data transfer for more than 40 persons in worst case scenario and much more in other scenarios. This means that the platform can be used for real-time monitoring and can significantly aid the process of prioritization in case of multiple injuries. One possible application is during the Medical Response to Major Incidents (MRMI).

Acknowledgment

The work presented in this paper was financed by SIARS (Smart I (eye) Advisory Rescue System), NATO SfP Project Number 984753.

References

[1] D. Claudi and G. E. Okudan, "Utility function-based patient prioritisation in the emergency department", European Journal of Industrial Engineering, Volume 4, Issue 1, DOI: 10.1504/EJIE.2010.02957

[2] Jawbone, https://jawbone.com/up/devices

[3] Vitality, http://clouddx.com/vitaliti/

[4] SensiumVitals, http://www.sensium-healthcare.com/

[5] HealthPatch MD, http://www.vitalconnect.com/healthpatch-md

[6] QardioCore, https://www.getqardio.com/

[7] PCARD, http://ipsitransactions.org/journals/papers/tir/2014jan/p5.pdf

[8] C. E. Chronaki et al., "An eHealth platform for instant interaction among health professionals," Computers in Cardiology, 2003, 2003, pp. 101-104, doi: 10.1109/CIC.2003.1291100

[9] Chuan Jun Su, Mobile multi-agent based, distributed information platform (MADIP) for wide-area e-health monitoring, Computers in Industry, Volume 59, Issue 1, January 2008, Pages 55-68, ISSN 0166-3615, http://dx.doi.org/10.1016/j.compind.2007.06.001.

[10] S. Junnila et al., "Wireless, Multipurpose In-Home Health Monitoring Platform: Two Case Trials," in IEEE Transactions on Information

Technology in Biomedicine, vol. 14, no. 2, pp. 447-455, March 2010. doi: 10.1109/TITB.2009.2037615

[11] M. Jaime et al., “A Home E-Health System for Dependent People Based on OSGI”, in Intelligent Technical Systems, pp. 117-130, 2009. doi: 10.1007/978-1-4020-9823-9_9

[12] I. Pau, F. Seoane, K. Lindecrantz, M. A. Valero and J. Carracedo, "Home e-health system integration in the smart home through a common media server," *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Minneapolis, MN, 2009, pp. 6171-6174. doi: 10.1109/IEMBS.2009.5334500

[13] E. Ekonomou, L. Fan, W. Buchanan and C. Thuemmler, "An Integrated Cloud-based Healthcare Infrastructure," *Cloud Computing Technology and Science (CloudCom), 2011 IEEE Third International Conference on*, Athens, 2011, pp. 532-536. doi: 10.1109/CloudCom.2011.80

[14] Lejiang Guo, Fangxin Chen, Li Chen and Xiao Tang, "The building of cloud computing environment for e-health," *E-Health Networking, Digital Ecosystems and Technologies (EDT), 2010 International Conference on*, Shenzhen, 2010, pp. 89-92. doi: 10.1109/EDT.2010.5496512

[15] J. Vilaplana et al., “The cloud paradigm applied to e-Health”, in BMC Medical Informatics and Decision Making, Vol. 13, Num. 1, 2013, pp. 1-10, doi: 10.1186/1472-6947-13-35

[16] L. Fan *et al*., "DACAR Platform for eHealth Services Cloud," *Cloud Computing (CLOUD), 2011 IEEE International Conference on*, Washington, DC, 2011, pp. 219-226. doi: 10.1109/CLOUD.2011.31

[17] Cooking hacks, “e-Health Sensor Platform V2.0 for Arduino and Raspberry Pi [Biometric / Medical Applications]”, available at: https://www.cooking-hacks.com/documentation/tutorials/ehealth-biometric-sensor-platform-arduino-raspberry-pi-medical

[18]

Daily weather conditions as indicator of air pollution in Skopje

Masha Bekjkovikj[1], Elena Markova[1], Kostadin Mishev[1], Dimitar Trajanov[1]
[1]Faculty of Computer Science and Engineering
Ss. Cyril and Methodius University
Skopje, Macedonia

***Abstract* — The effects of climate changes are emerging, especially in the last several years, enabling increased influence of the global warming upon various aspects of the entire ecosystem and its environment. In this paper, a correlation of several datasets of weather conditions and air pollution indications measured on a daily basis, gathered in Skopje, is made in order to represent dependencies between those measurements. Analyzing the information gathered for the particles that indicate the pollution in the air and their connection to the weather conditions, using Pearson correlation, strives to the conclusion that the parameters of the weather conditions such as changes in air pressure, cloud coverage, wind speed etc. are indicators of air pollution. Each significant correlation is represented with a graph due to more clear visualization of the results. The results of our research consider a weak relation among weather and air pollution parameters. Such consideration can be lately used as prediction of the air quality index depending on the current weather conditions. The results of such predictions can produce avoidance of negative effects that the air pollution can invoke to the human health.**

Keywords— air pollution, weather conditions, Pearson correlation, PM10,PM2.5,CO, Skopje,air quality

I. INTRODUCTION

Due to the emerging climate changes, various aspects of the entire ecosystem and its environment are affected. Such interactions are also notable in Republic of Macedonia, especially in its capital Skopje, considering its geographical position and population density. Several air quality measurement stations, are proving that the air quality indicators exceed the allowed limits, especially the PM2.5 and PM10 particles which are one of the most dangerous compounds of the air. It is remarkable that the index of air pollution is higher in winter than in summer period, making us curious about the potential factors of the air pollution. Many attempts throughout the years have been made in order to find the cause for the occurrence of these concentrations of particles that cause air pollution. People have become more aware of the rapid changes and increased health problems. Researchers strive to unveil the indicators of such occurrences in order to stop or soften their effects.

PM10 particles or coarse dust particles are 2.5 to 10 micrometers in diameter and PM2.5 particles or fine particles are 2.5 micrometers in diameter or smaller. These particles appear as a result from crushing or grinding operations, all types of combustion, including motor vehicles, power plants, residential wood burning, forest fires, some industrial processes etc. It is shown that especially people with heart or lung diseases are affected by the appearance of these particles. These particles also have effect on the vegetation by clogging the stomatal openings of plants and interfering with photosynthesis functions.

In this paper, a correlation is shown to represent the connection between the occurrence of PM2.5 and PM10 particles and also CO emissions and weather conditions within the region of Skopje. The research consists of two different datasets. Each dataset contains daily measurements on an hourly basis for 2015. The measurements for the weather conditions are provided from the National Hydro meteorological Agency's measurement stations on several locations in Skopje. The second dataset provides information about the air quality and the quantity of CO, PM2.5 and PM10 concentrations gathered from several measurements stations installed in Skopje, specialized for following these parameters. Such data export, could provide interesting information for the distribution of the indicators quantity allowing us join with the parameters of the weather forecast.

The structure of the paper is the following: section II presents the related work, section III explains used data sources, while section IV describes the system architecture; section V presents the data analysis and finally the results are presented in Section VI. The conclusion and future work are presented accordingly in Section VII and VIII.

II. RELATED WORK

In the paper Health effects of outdoor air pollution [1] the consequences of air pollution are well explained. The behavioral effects are described, particularly as a result of exposure in early childhood, even in fetuses and newborns, and the consequences that result in hypertension in adults, whether [9] explains the acute health effects that air pollution has. Since the control of exposure is important from the very beginning of life, that was the motivation to secure the dataset for the particles of air pollution and to seek correlations, in order to see what results can be obtained to take some preventive measures. Air pollution also has effects on plants as well, which are described in detail in paper [7].

The weather conditions also have impact on the Cyanobacterial blooms. The presented work in [2] explains that the Cyanobacterial blooms are common in the Baltic Sea and that their development is highly influenced by weather conditions such as temperature and wind. This indicates that the weather conditions take a serious part of very different aspects of the world.

Another example which is well know is that the weather conditions effect on people's health and pain. The research in [3] evaluates the influence of the weather on pain in patients with rheumatic pain as well as correlates different climate variables with the patients' impression of weather sensitivity and assess correlations between pain and climate conditions on 5 days preceding and following painful episode. The results of the paper support the previously mentioned well known belief about the effect that weather conditions have on people's pain.

The research in [4] shows interesting results about the influence of weather conditions on physical activity in adolescents, where, for example, one of the conclusions is that the declines in physical activity can be explained by the declines during the winter season.

In [5] the influence of weather conditions on activity is also examined, but on insectivorous bats.

III. Data Sources

There are multiple measurement stations distributed in the region of Skopje providing different air quality indicators. Most of them, provide open access REST services which return real-time measurements. In our paper, we will use the services provided by:

- Ministry of environment and physical planning - The JSON service which provides measurement about CO, NO2, PM10, PM2.5, SO2, O3 air quality indicators providing data refresh each hour during the day. There are multiple measurement stations over Macedonia, but for purposes of this paper, we will use only the stations located in Skopje: Centar, Karposh, Lisitche, Gazi Baba and Rektorat.

- Measurement CO2 stations provided by the project Skopje Green Route - These measurement stations are placed on the most frequent crossroads in Skopje: Justice Palace, Red Cross and Faculty of Agriculture, providing measurement about CO2 air indicator refreshing the information each 5 minutes.

- Measurement station maintained by the "Laboratory of Eco informatics at Faculty of Computer Science and Engineering" providing information about the same air indicators like the measurements stations enabled by Ministry of environment and physical planning.

- Daily data log provided by the National Hydro meteorological Agency's measurement stations providing information about the current temperature, humidity, air pressure, wind speed and wind direction. By crawling such data from their web interface, the system obtains measurements on hourly level.

Some of the web services are RESTful and provide open URL location which can be accessed with GET requests. For the others ones, we use data crawlers which process the content of the web page and save the useful data in our local database. The log of all services is kept on our database which centralizes all information about all air quality indicators for all measurement stations. It runs scheduled processes which poll the JSON services asking for new fresh data from sensors. It appends timestamp to the measurement information and saves into MySQL database.

IV. System Architecture

The core of the system, shown on Figure 1, consists of a web crawler which gathers data from heterogeneous data sources. It represents a cron script which invokes the remote rest services, and returns the data to endpoint provided by the data log application. Such data is saved into MySQL database by invoking of the endpoint of a Java Spring application (Fig.1) providing possibilities of extending with other data sources. The process of data exchange between the service providers and out database is tempered on hourly level enabling synchronization of multiple heterogeneous data sets. Such data, can be easily edited and processed for forthcoming data analysis. Also, we provide user interface for visualization of continuous data series.

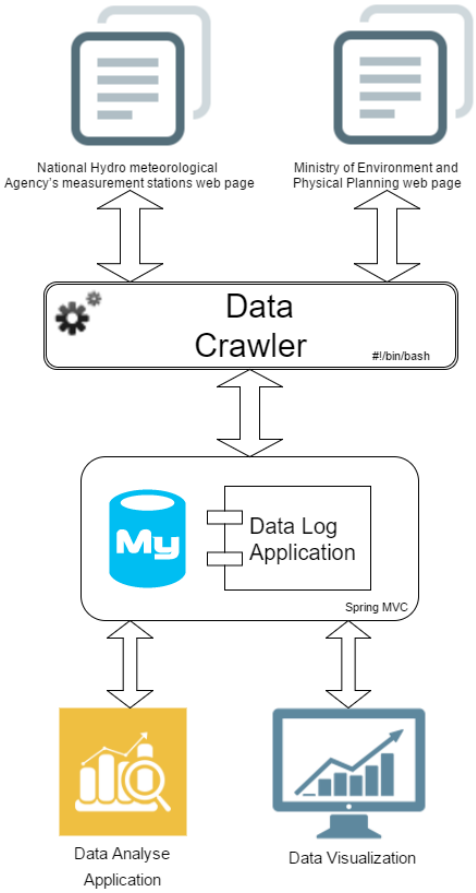


Fig 1. Schema of the system architecture.

V. Data Analysis

The collected data were merged into a common table with the date and time as a common parameter to create arrays as input parameters for the calculation of correlations. For data analysis and calculation of correlations, the scripting programming language Python [12] was used, through the interface PyCharm of JetBrains [13]. The data received in the collection stage and through writing SQL queries in a MySQL structure [13] led to structured time series that meet the requirements and can be used to calculate the correlation. Correlation coefficients were stored in flat files in csv format to facilitate the analysis in the python scripts.

Additional calculations were also made with python scripts in order to optimize the time of execution and data processing. The scipy module [15] was used because it has already implemented methods and functions used for analysis and data processing.

Pearson correlation coefficient is a measure of the linear correlation between two variables X and Y, giving a value between +1 and −1 inclusive, where 1 is total positive correlation, 0 is no correlation, and −1 is total negative correlation. It is widely used in the sciences as a measure of the degree of linear dependence between two variables [6].

VI. Results

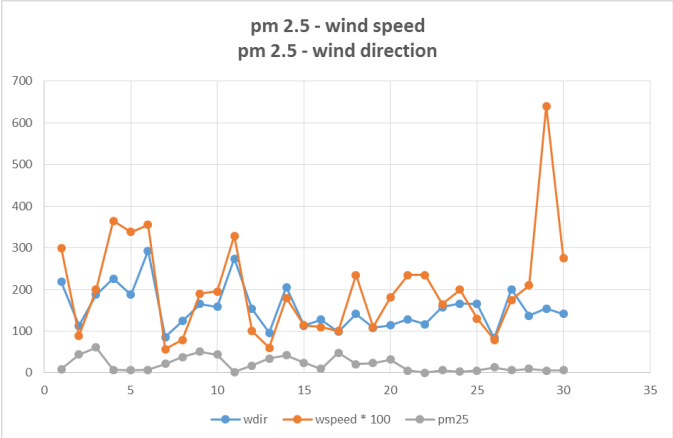


Fig 2. Graphic representation of dependencies of PM 2.5 particles with wind speed and direction

Parameter 1	Parameter 2	Correlation coefficient
Wind speed	PM 2.5	-0.29

Table 1. Measurement results for PM 2.5 particles, wind speed and direction

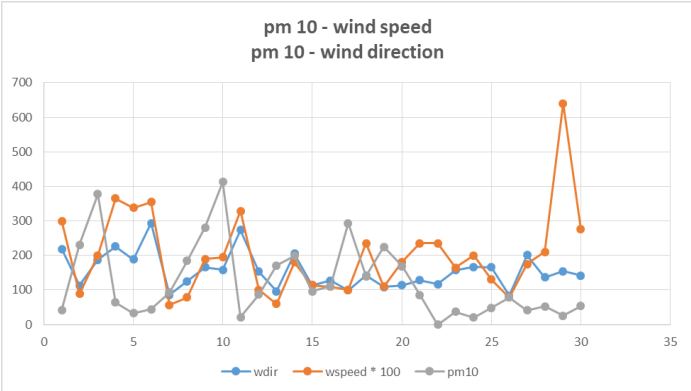


Fig 3. Graphic representation of dependencies of PM 10 particles with wind direction and speed

Parameter 1	Parameter 2	Correlation coefficient
Wind direction	PM 10	-0.34
Wind speed	PM 10	-0.30

Table 2. Measurement results for PM 10 particles, wind speed and direction

The results from the analysis of PM2.5 particles and it's correlations with wind speed and direction measurements show that these two parameters are in weak negative correlation. When using data to calculate the correlations on a daily basis between these two parameters some of the results show strong correlation as opposed to monthly based correlations due to the fact that daily measures are more accurate and monthly measures are a kind of an average value of parameter measures, but the overall conclusion is that generally PM2.5 particles are inversely proportional with wind speed and direction which means when there are higher measures of wind speed there is less concentration of PM2.5 particles in the air.

As Table 2 shows, the correlation coefficient between wind speed and direction and PM10 particles gives inverse proportional behavior of these two parameters.

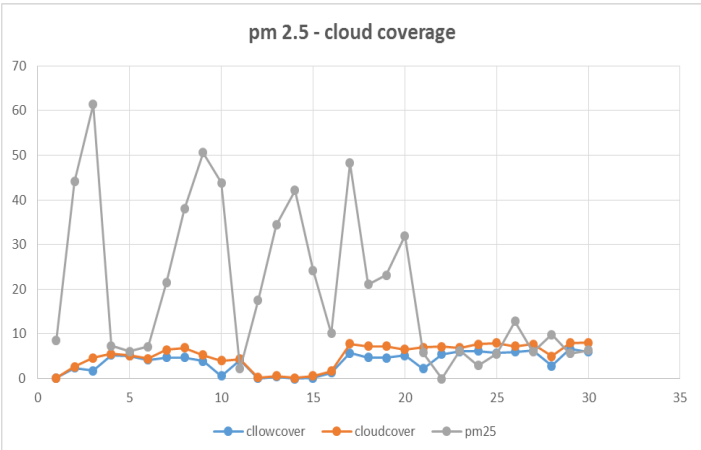


Fig 4. Graphic representation of dependencies of PM 2.5 particles with cloud coverage

Parameter 1	Parameter 2	Correlation coefficient

Cloud low cover	PM 2.5	-0.45
Cloud cover	PM 2.5	-0.47

Table 3. Measurement results for 2.5 particles with cloud coverage

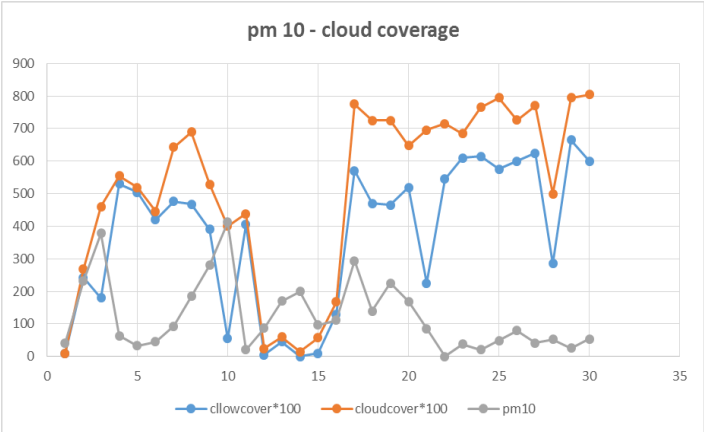


Fig 5. Graphic representation of dependencies of PM 10 particles with cloud coverage

Parameter 1	Parameter 2	Correlation coefficient
Cloud low cover	PM 10	-0.43
Cloud cover	PM 10	-0.47

Table 4. Measurement results for PM 10 particles with cloud coverage

Cloud coverage is an important weather condition parameter. The level of cloud concentration is often related to the occurrence of particles, especially during the autumn period in Skopje. The results from the analysis of PM2.5 particles and it's correlations with cloud coverage show that these two parameters are in weak negative correlation. The consistency of the daily as well as the monthly calculated correlation strive to the overall conclusion that PM2.5 particles are inversely proportional with cloud coverage concentration which means when there is higher concentrations of clouds there is less concentration of PM2.5 particles in the air.

As Table 4 shows, the behavior of PM10 particles resembles with the one of PM2.5 particles when influenced by the level of cloud coverage.

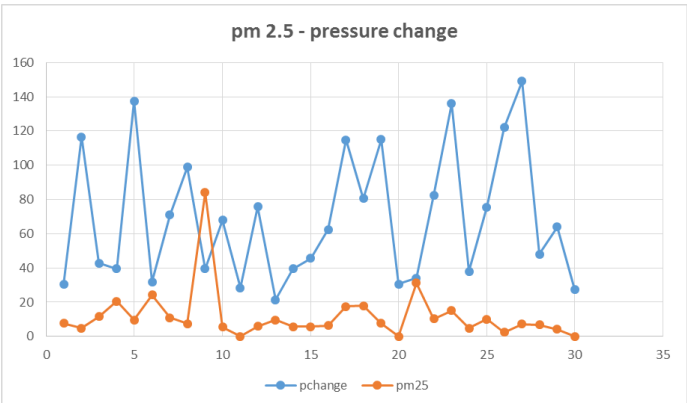


Fig 6. Graphic representation of dependencies of PM 2.5 particles with pressure change

Parameter 1	Parameter 2	Correlation coefficient
Pressure change	PM 2.5	-0.30

Table 5. Measurement results for PM 2.5 particles with pressure change

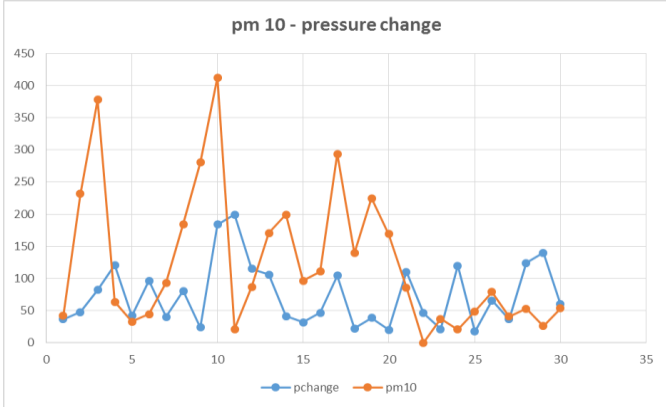


Fig 7. Graphic representation of dependencies of PM 10 particles with pressure change

Parameter 1	Parameter 2	Correlation coefficient
Pressure change	PM 10	-0.27

Table 6. Measurement results for PM 10 particles with pressure change

As Figure 6 shows, the changes in the pressure influence the occurrence of PM2.5 with a weak negative coefficient of correlation. When the indications of pressure change are low, there is increased concentration of PM2.5 particles in the air.

As Figure 7 shows, PM10 particles behave similarly with pressure changes thus it has a weak negative coefficient of correlation with pressure change.

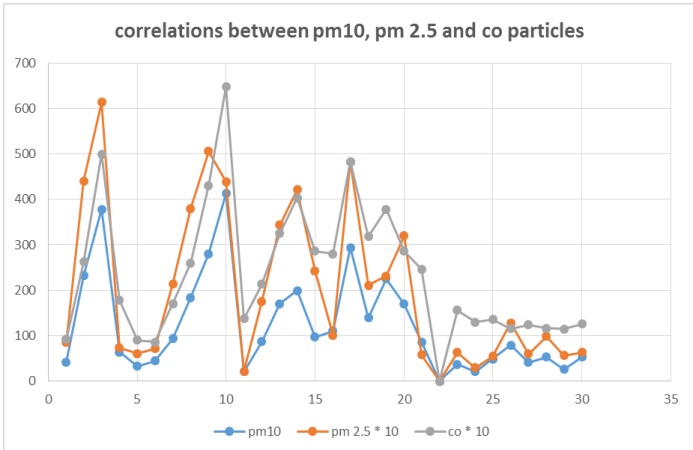


Fig 8. Graphic representation of dependencies between PM 2.5 particles, PM 10 particles and CO emissions

Parameter 1	Parameter 2	Correlation coefficient
PM 10	PM 2.5	0.89
PM 10	CO	0.93

PM 2.5	CO	0.86

Table 7. Measurement results for PM 2.5 particles, PM 10 particles and CO emissions

The occurrence of PM10 and PM2.5 particles, is tightly related (Figure 8). The interesting part in such analyze is the relation among such particles and CO emissions (0.93 and 0.86). Many researches [10] [11] show that most of the carbon monoxide emission come from transportation sources. The correlations between the emission and the particles show that it is very likely that in an environment polluted with high density of CO emissions, there is also high concentration of these particles, thus these parameters have high positive coefficient of correlation (Table 7).

VII. Conclusion

In order of strong correlation between PM10 and PM2.5 particles, the conclusion is that the presence of these two particles in the air is mutually dependent. This explains why these two particles are very similarly impacted by the parameters of weather conditions. Due to the fact that the dataset contains measurements on monthly basis, it is implied that these measures are somewhat an average value of measurements made on daily basis. Even so, the correlations between measurements of PM10 and PM2.5 particles, after the analysis, show that these two parameters have a strong positive correlation.

From the results shown on Figure 8 and Table 7, it is clear that the concentrations of CO emissions in the air is tightly related to the occurrence of PM10 and PM2.5 particles. This indicates that CO emissions are one of the main causes for the occurrence of these particles.

The other parameters of weather condition, such as wind speed and direction, changes in pressure and cloud coverage may not have as strong impact on the occurrence of the particle as CO emissions have, but they cannot be completely excluded as indicators of air pollution. If a correlation coefficient is consider to be weak on the scale of -0.3 to 0.3 and a strong coefficient has value of either greater than 0.6 or lower than -0.6, then the results shown in section VI classify these certain parameters to have somewhat medium influence on air pollution. Having measurements on a larger period of time and more strictly divided geographically, even only for the municipalities of Skopje, maybe better results would be obtained. The dataset analyzed in this paper, even though medium, these influences can be considered to be a worthy indicator of air pollution.

VIII. Future Work

It is well known that directly or indirectly many fields are influenced by weather conditions and air pollution, starting from environment, health and going further to people's behavior. In this paper we only present a small research on the dependencies these two factors have. If the datasets are widened and include more data and information not just for 2015 but other years also, some very interesting statistics can be pulled and analyzed. Having these datasets allows various different correlations and dependencies to be examined and this research can be expanded with more parameters, its time frame can be increased, more cities can be taken in consideration etc.

For example, given the results, strong correlations are noticed with the emission of CO gases. This implies the opportunity to expand the research and find correlation with other similar gases, for example CO2. The datasets for the CO2 could be gathered using the sensors set in the city.

Another work that could be done is to find correlation between parameters of the same dataset in case they do not correlate with the parameters from the other datasets.

Also, this research can be expanded and this generalized model for the city of Skopje can be made more specific for each and every municipality of Skopje which would give more accurate information because specific measurements for parameters are definitely more precise than generalized ones. The geographic characteristics of the city can be also used for finding correlations.

References

[1] Bascom, R., Bromberg, P.A., Costa, D.L., Devlin, R., Dockery, D.W., Frampton, M.W., Lambert, W., Samet, J.M., Speizer, F.E. and Utell, M., 1996. Health effects of outdoor air pollution. American journal of respiratory and critical care medicine, 153(2), pp.477-498.

[2] Kanoshina, I., Lips, U. and Leppänen, J.M., 2003. The influence of weather conditions (temperature and wind) on cyanobacterial bloom develoPMent in the Gulf of Finland (Baltic Sea). Harmful Algae, 2(1), pp.29-41.

[3] Strusberg, I., Mendelberg, R.C., Serra, H.A. and Strusberg, A.M., 2002. Influence of weather conditions on rheumatic pain. The Journal of Rheumatology, 29(2), pp.335-338.

[4] Bélanger, Mathieu, Katherine Gray-Donald, Jennifer O'loughlin, Gilles Paradis, and James Hanley. "Influence of weather conditions and season on physical activity in adolescents." Annals of epidemiology 19, no. 3 (2009): 180-186.

[5] Erickson, J.L. and West, S.D., 2002. The influence of regional climate and nightly weather conditions on activity patterns of insectivorous bats. Acta Chiropterologica, 4(1), pp.17-24.

[6] Sedgwick, P., 2012. Pearson's correlation coefficient. BMJ, 345, p.e 4483.

[7] Thomas, M.D., 1961. Effects of air pollution on plants. Air pollution, 239.

[8] National Hydro meteorological Agency, http://meteo.gov.mk/

[9] Seaton, A., Godden, D., MacNee, W. and Donaldson, K., 1995. Particulate air pollution and acute health effects. The lancet, 345(8943), pp.176-178.

[10] Wu, F., Wang, J., Chen, W. and Shuai, S., 2009. A study on emission performance of a diesel engine fueled with five typical methyl ester biodiesels. Atmospheric Environment, 43(7), pp.1481-1485.

[11] Stephens, R.D. and Cadle, S.H., 1991. Remote sensing measurements of carbon monoxide emissions from on-road vehicles. Journal of the Air & Waste Management Association, 41(1), pp.39-46.

[12] Python https://www.python.org/

[13] PyCharm https://www.jetbrains.com/pycharm/

[14] MySQL https://www.mysql.com/

[15] SciPy https://www.scipy.org/

Cross platform system integration using web services

Aleksandar Stojmenski, Vlatko Nikolovski, Ivan Chorbev, Vesna Dimitrova
FCSE
Skopje,Macedonia
{aleksandar.stojmenski, vlatko.nikolovski, ivan.chorbev, vesna.dimitrova}@finki.ukim.mk

***Abstract*—This paper presents a brief overview of the concepts for collaboration between various systems developed for the Faculty of Computer Science and Engineering in Skopje. Web technology such as the HTTP, originally designed for human-to-machine communication, is utilized for machine-to-machine communication, more specifically for transferring machine-readable data in web service formats such as JSON. A Central Authentication Service is being used for identity management and single sign-on for all integrated applications. Various guidelines and the whole process of integration of existing systems along with their interconnection and interoperability are covered. The interface that we are using in the integration of the multi-platform system pose no hard dependencies between the various applications, thus allowing easily integration and intercommunication protocols.**

Keywords—collaboration; systems integration; web services; cross-platform;

I. Introduction

System integration is a complex process where a cohesive platform is created from components that were not specifically designed to work together. Components of an integrated platform are often stand-alone systems that operate on different computer environments. This paper describes the platform integration of various applications and their interconnection and interdependability.

In order to collaborate between each other, these applications need to specify suitable protocols for exchanging data, as well as protocols for flow control. This paper gives an overview of the transformations of the data structures, the data itself and the impact of the different applications over the data. For this purpose, different design patterns are used to facilitate the communication between systems and to harmonize endpoint data formats that this paper examines. Different problems and their possible solutions are presented, regarding systems lifecycle, architecture, process, interface, synchronization and security. Each application endpoint exports OAuth2 security protocol functionalities for system authentication and authorization. Following the principles of the OAuth2 protocol, each server authenticates the users using bearer tokens. Furthermore, the communication protocol adopts the JSON data format as a primary exchanging throughput over a HTTP communication channel.

II. Background work

Although a lot of work and progress has already been done in the area of web services in the past years, efforts have been mostly focused on service description models and languages, and on automated service discovery and composition [1].

Web services were developed as a solution to (or at least as a simplification of) the system integration problem [2]. The main benefit they bring is that of standardization, in terms of data format (JSON), interface definition language (WSDL), transport mechanism (SOAP) and many other interoperability aspects. Standardization reduces heterogeneity and makes it therefore easier to develop business logic that integrates different (Web service-based) applications. Web services also represent the most promising technologies for the realization of service-oriented architectures (SOAs), not only within, but also outside companies' boundaries, as they are designed to enable loosely coupled, distributed interaction [3].

While standardization makes interoperability easier, it does not remove the need for design patterns that include adapters and mediators. Different Web services may still support different interfaces and protocols. For example, although two map or driving direction services may support JSON or XML and use SOAP over HTTP as transport mechanism, they may still provide operations that have different names, different parameters, and different business logic or protocols. In addition, other opportunities enabled by Web services have an implication in terms of adaptation needs. In fact, having loosely-coupled and B2B interactions imply that services are not designed having interoperability with a particular client in mind (as it was often the case with CORBA-style integration) [4]. They are designed to be open and possibly without knowledge, at development time, about the type and number of clients that will access them, which can be very large. The possible interactions that a Web service can support are specified at design time, using what is called a business protocol or conversation protocol. A business protocol specifies message exchange sequences that are supported by the service, for example expressed in terms of constraints on the order in which service operations should be invoked. Another studied solution is to make system integration with ActiveXML which utilities peer-to-peer interaction between nodes and specifies special data design and ActiveXML web services[5].

III. Systems Architecture

The system architecture that is discussed in this paper consists of several different subsystems, all of them developed in different technologies. The CAS (Central Authentication Service) is a service developed in Java that provides single sign-on and identity management throughout the whole system architecture. The CAS service involves a back-end service, that does not have its own HTTP interface, but communicates

with a web application. The Service manager implements a protocol which is platform independent (JSON based). All applications and services in the system are communicating and synchronizing using this protocol. The service manager is implemented in C# Web Api and is used as a mediator for control messages exchange, storing permission access rules, identifying the status of the services (running, failed, blocked…) and enabling intra service communication.

The quintessential workflow starts when the client visits an application desiring to authenticate to it and the service manager is presented. The service manager then redirects to CAS to check if the user is authenticated for the requested service. CAS validates the client's authenticity by checking a username and password against a database with Active Directory users. If the authentication succeeds, CAS returns to the service manager passing along a security token. The service manager then validates the ticket by contacting CAS over a secure connection and providing its own service identifier and the ticket.

After obtaining the information about whether the user is successfully authenticated, the service manager checks the role of the user (professor, student, administrative staff etc.). If the user has a permission for the particular application, the service manager redirects the user to the application.

The service manager uses information from the iKnow system in order to serve information to the dependent applications. The information that is gathered is role specific. For example, if a student is being logged into a particular application, information about the number of ECTS credits awarded, the subjects that the student applied for the particular semester and general information (email, name, surname) are taken from the iKnow system. On the other hand, different types of information are provided, such as teaching subjects, number of students etc. The system triggers real time requests for each of those information's that are available instantly and up to date. Having those information's, the applications can easily merge the subjects with their students and professors. The discussed system architecture is shown on Figure 1.

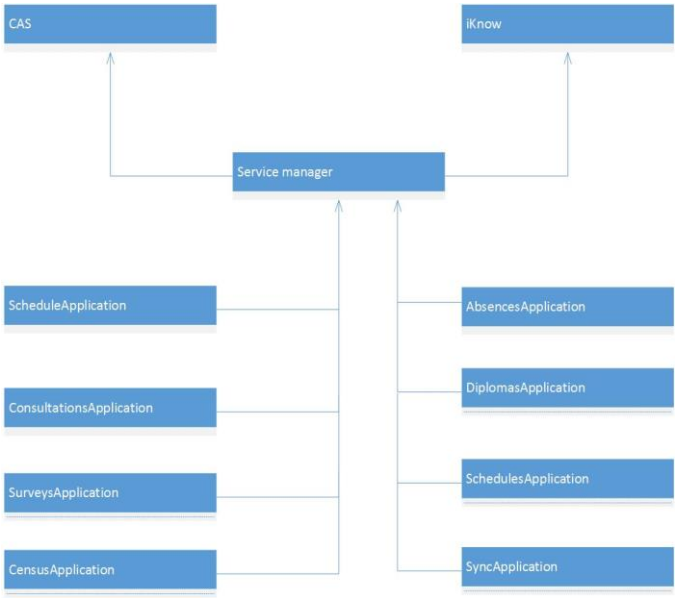


Fig. 1. Systems architecture

Technologies that are used in the system architecture include HAProxy[6] for load balancing, VMWare[7] for virtualisation, SQL server databases [8], WEB Api for interservice communication[9], Google Analytics[10] for traffic measurement, AngularJS[11] for frontend and WCF/WEB API services for communication between iKnow[12] and the different applications.

IV. Systems scalability

For handling numerous concurrent requests during exam applying and student enrollment, a load balancer is used. The load balancer is a device that acts as a reverse proxy and distributes network or application traffic across a number of servers. Behind the load balancer, applicative servers redirect the requests based on the workload of the server. The HAProxy product is used for load balancing because it supports more than 9 load balancing algorithms. The most common ones being:

- round-robin (for short connections, pick each server in turn)
- leastconn (for long connections, pick the least recently used of the servers with the lowest connection count)
- source (for SSL farms or terminal server farms, the server directly depends on the client's source address)
- uri (for HTTP caches, the server directly depends on the HTTP URI)
- hdr (the server directly depends on the contents of a specific HTTP header field)
- first (for short-lived virtual machines, all connections are packed on the smallest possible

subset of servers so that unused ones can be powered down)

We are using the round-robin load-balancing algorithm because it gives approximately equal distribution of incoming sessions. Another aspect of an efficient system scalability are the vertical and horizontal scaling paradigms, and how they can be optimized to provide the very best equitation between the performance and efficiency. When using vertical scaling on demand, physical resources (RAM, CPU) of the server are increased. When a server rich to a point where vertical scaling is insufficient or the vertical scaling buffer is overloaded, horizontal scaling is used. A cache is used that provides high throughput, low-latency access to commonly accessed application data by storing the data in memory. For an integrated cross-platform system as described in this paper, the most useful type of cache is distributed cache, which means the data is not stored on the individual web server's memory, but on other resources, and the cached data is made available to all web servers of an application (or other virtual machines that are used by the application). Redis[13] is used for distributed cache. The cache in Redis is persistent, which means that if one system fails or drops off the network, the session can be restored on another recovery server that will establish the distributed cache. This means that if one server is overloaded, additional virtual machines can be started on demand in order to maintain different sessions.

V. Use case scenario

At the beginning of the use case scenario for the Absences application, an employee in the Faculty of Computer Science and Engineering is authenticating via the CAS system. After successful authentication, the service manager checks if the user is in one of the following roles: teaching staff, administrative staff or computer center staff. Accordingly, the system shows an absences request form in which the employee writes the destination, purpose of the absence and other informations. After the request is submitted, the head of institute should approve the absence request and hand it over to the archive via the system. If the type of the absence is paid absence, after the leaving took place, the employee should submit the travel expenses in the system. After those are submitted, the system synchronizes with services offered by the National Bank of Republic of Macedonia (NBRM) to retrieve the up to date currency of the specific country in order to automatically convert the sum to Macedonian denars (MKD). When the currency is converted, the request is handed to the Vice-Dean for finances for approval. After the approval, the sum is paid to the employee via the Accountancy department of the faculty. The workflow of the Absences and the Declaration applications is shown on Figure 2.

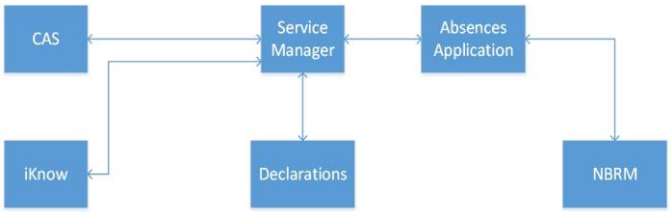


Fig. 2. Workflow for Absences and Declaration applications

For the Declarations application, the login with CAS and role identification is same as described above in this chapter. Only the teaching staff has access to this application. If the role is teaching staff, the application synchronizes with iKnow in order to retrieve the number of subjects and the number of students that are enrolled to each subject for the current semester. The application then generates declarations for payment in which each subject with the number of students is listed for the logged teacher.

VI. Usage statistics

Web analytics measurements for all services used by the Faculty of computer science and engineering are gathered from the Google Analytics service. On Figure 3 we can see that *the* number of sessions in the iKnow system is significantly higher from January to the beginning of February 2016. This is the period when students verify the past semester and enroll to the next semester of their studies. Using the Google Analytics we can identify critical events and prepare adequate number of servers in order to handle the expected incoming sessions.

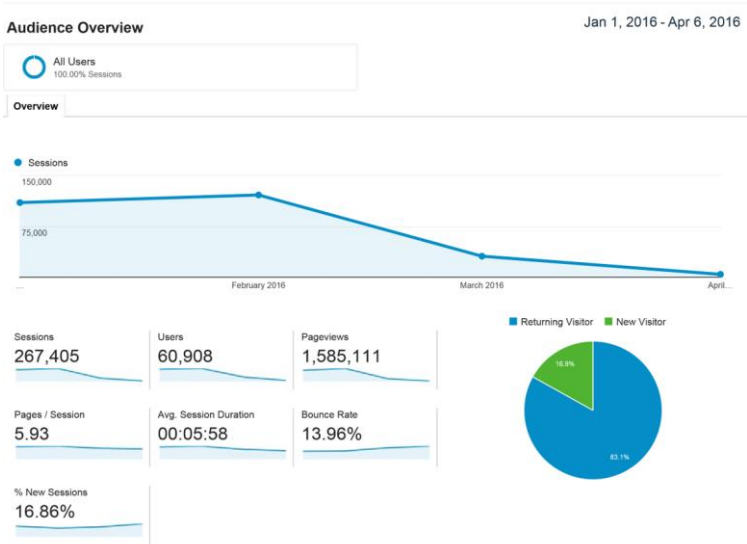


Fig. 3. Session analytics for iKnow system

Each of the application servers has 8GB of RAM, while 6.5GB of those are user usable. Based on the previous collected data each of the application servers can serve up to 13 000 concurrent user sessions, which leads to having approximately 0.5MB per user session. This number defines the hard threshold limit when the vertical scale transforms to a horizontal scale. On the other hand, the database server has 16GB RAM and 8 CPU cores. The working RAM that is used for running SQL Server itself is about 7GB which leaves 9GB of RAM for connections and queries. When the database server reaches 20 000 sessions, the SQL Server instance is experiencing deadlocks on the database tables for Courses and Exam Applications. Then we use transparent scale-out, with which an application does not need to know that the data is distributed in several databases across Microsoft SQL Server instances or servers. One of the benefits of transparent scale-out is that the application is simpler to code and maintain

because it has no information about the data location or partitioning. The primary feature of the SQL Server that allows scale-out is the distributed partitioned view. The mechanism used by SQL Server for the distributed partitioned view data access to a remote database is known as linked server. A linked server is a SQL Server database object that resides at an instance scope. The query engine has built-in object linking and embedding database (OLE DB) access. This allows data to be retrieved by the same basic mechanism from a remote storage engine as it is from the local SQL Server storage engine. With scale-out, SQL Server factors the query into smaller queries that run on different servers, and then all the results are aggregated. As with data-dependent routing (DDR), master and reference data would most likely be replicated across all servers to minimize data movement for join purposes. Depending on the complexity of the queries and how they are coded, there can be performance penalties when performing cross-server queries.

Another interesting analysis is the user flow through the iKnow system. As we can see on Figure 4, the most visited are the student pages for semester enrollment and exam application. Afterwards in the list of mostly viewed pages are the pages for checking grades information, personal information and sending document requests. All of the other pages that are present in the system, such as pages for professors, teaching assistants and student service are roughly viewed in 1% or less of the whole number of pages present in the iKnow system.

Page	Pageviews	% Pageviews
1. /Students/StudentsHome.aspx	246,778	15.79%
2. /Students/StudentsCourses.aspx	227,475	14.55%
3. /Account/Login.aspx?ReturnUrl=/	200,782	12.85%
4. /Students/StudentsSemesters.aspx	197,315	12.62%
5. /Students/ExamsAppliedByStudent.aspx	147,280	9.42%
6. /Students/ExamsPassedByStudent.aspx	142,457	9.11%
7. /Students/StudentsEdit.aspx	52,655	3.37%
8. /Account/Login.aspx?ReturnUrl=/Students/StudentsCourses.aspx	28,163	1.80%
9. /Students/StudentsRequests.aspx	26,224	1.68%
10. /Account/Login.aspx?ReturnUrl=/Students/StudentsHome.aspx	21,006	1.34%

Fig. 4. Page views analytics for iKnow system

VII. Conclusion

This paper presents a proof of concept protocol for intercommunication and synchronization between various applications developed by the Faculty of Computer Science and Engineering, Skopje. Using the JSON as a primary standard for data structures, the protocol provides a secure, platform independent composite system that unites various different applications under the same secure context. Furthermore, it provides a highly scalable architecture with built-in functionalities for reliability and management of the lifecycle of the composite applications. Use case scenarios of some of the applications are presented along with the relevant workflows. Usage statistics from the iKnow system in terms of scalability are discussed and the used overload session handling mechanisms are explained.

References

[1] Benatallah, Boualem, Fabio Casati, Daniela Grigori, Hamid R. Motahari Nezhad, and Farouk Toumani. "Developing adapters for web services integration." In *Advanced Information Systems Engineering*, pp. 415-429. Springer Berlin Heidelberg, 2005.

[2] G. Alonso, F. Casati, H. Kuno, V. Machiraju. Web Services: Concepts, Architectures, and Applications. Springer Verlag. 2004.

[3] B. Benatallah, F. Casati, and F. Toumani. Web services conversation modeling: A Cornerstone for EBusiness Automation. IEEE Internet Computing, 8(1), 2004.

[4] L. Bordeaux et al. When are two Web Services Compatible?. VLDB TES'04. Toronto, Canada. 2004. [CFPT03] C. Canal, L. Fuentes, E. Pimentel, J. Troya, A. Vallecillo. Adding Roles to CORBA Objects. IEEE TSE, 29 (3), 2003, IEEE Press

[5] Abitrboul, Serge, Omar Benjellourn, Ioana Manolescu, Tova Milo, and Roger Weber. "Active xml: Peer-to-peer data and web services integration." In*Proceedings of the 28th international conference on Very Large Data Bases*, pp. 1087-1090. VLDB Endowment, 2002.

[6] Tarreau, Willy. "HAProxy-the reliable, high-performance TCP/HTTP load balancer." *2011-8)[2013-4]. http://haproxy. lwt. eu* (2012).

[7] Rosenblum, Mendel. "VMware's virtual platform™." In *Proceedings of hot chips*, vol. 1999, pp. 185-196. 1999.

[8] Microsoft Sql Server <https://www.microsoft.com/en-us/server-cloud/products/sql-server/, 18.03.2016>

[9] Microsoft Web Api <http://www.asp.net/web-api, 19.03.2016>

[10] Plaza, Beatriz. "Google Analytics for measuring website performance."*Tourism Management* 32, no. 3 (2011): 477-481.

[11] Green, B. and Seshadri, S., 2013. *AngularJS*. " O'Reilly Media, Inc.".

[12] Lowy, J., 2007. *Programming WCF services*. " O'Reilly Media, Inc.".

[13] Sanfilippo, Salvatore, and Pieter Noordhuis. "Redis." *2011-03-19]. http://redis. io* (2009).

Maximizing Flow in Two-Terminal Multidimensional Flow Network

Martin Josifoski, Marija Mihova, Kire Trivodaliev
Faculty of Computer Science and Engineering
University "Ss. Cyril and Methodius"
Skopje, Macedonia
josifoski.martin@students.finki.ukim.mk, {marija.mihova, kire.trivodaliev}@finki.ukim.mk

Abstract — **Real-world systems ranging from communication to transportation are modelled using multistate flow networks. However these models fail to capture the complex layered structure of the systems having many different types of interaction among its constituents all of which contribute to its overall operation. In this paper we aim to introduce this complex structure considering a simplified two-terminal multidimensional flow network where all nodes are connected with multiple links. Once we make the definitions needed for this case, we formally define the problem of flow maximization for such a network and give the procedure for finding the solution.**

Keywords— Flow Maximization; Multidimensional Networks; Two-Terminal Flow Networks

I. INTRODUCTION

Many real-world systems, such as computer and communication systems, power transmission and distribution systems, and oil/gas production and transportation systems, can be modeled as multistate flow networks (MFNs) in which the state of each component (arc/node) can be regarded as being independent, discrete, finite, and multiple-valued. For example [1], a gas production and transportation system can be modeled as an MFN in which the flow, component, and state are defined as the gas, the compressor station comprising a certain number of compressor units in parallel, and the number of working units in the compressor station. In general a MFN should have the following properties [2,3]: a) the graph is connected and there are no self-loops; b) the capacity of each edge is a non-negative integer-valued random variable according to a given distribution; c) the capacities of different edges are s-independent; and d) all flows in the network obey the conservation law, i.e., the total flows into and from a node (not source and target nodes) are all equal.

The complexity of the aforementioned and many more real-world systems is such that their operation is layered and we can view these systems as being multidimensional with each dimension depicting a distinct interaction among constituents and all of them interplaying to fulfill the emergent function of the system [4,5].

A common scenario of a multilayered system is one where the same sets of nodes are connected via more than one type of link. Examples of such networks abound: individuals in a society are networking through numerous social relationships such as friendship, kinship, co-workership and via a multitude of communication channels such as online and offline contacts [6]. Critical infrastructure provides essential support for the functioning of modern society through concerted operations of multiple interlinked and interrelated networks such as energy production and supply, telecommunication, and transportation networks [7].

The aim of this paper is to introduce the multidimensionality as a property of the MFNs and moreover to determine a procedure for flow maximization in such a scenario. We deal with the problem of flow maximization in a two-terminal multidimensional flow network as we believe that this case can later on be generalized to problems closer to real-world systems.

The rest of this paper is organized as follows. Section 2 gives the basic definitions used. Section 3 introduces the multidimensionality, while in Section 4 we formally define the maximization problem and give the approach for solving it. This paper is concluded in Section 5.

II. BASIC DEFINITIONS

A two-terminal flow network is usually defined as a directed graph $G(V, E)$, but there are many flow problems that can be modeled by undirected graphs. In this section we give some basic definitions and concepts about undirected flow networks. These definitions are similar to the corresponding definitions for digraphs, but are adapted to undirected graphs.

A *two-terminal undirected flow network* is an undirected graph with two special vertices, a source s and a sink t ($s \neq t$), in which each edge $\{u, v\} \in E$ has a nonnegative capacity $c(\{u, v\}) \geq 0$. The function c is called *capacity function*. Shortly, we will denote such a capacity network by $G(V, E, c)$.

A *flow* in an undirected network $G(V, E, c)$ is a function $f : V \times V \rightarrow R^{+} \cup \{0\}$ that satisfies the following three constraints:

1. *Capacity constraint*: $0 \leq f(u, v) \leq c(\{u, v\})$, for all $u, v \in V$, i.e., the flow of an edge cannot exceed its capacity.

2. *Flow conservation*: for each $v \in V$,

$$f(V,v)-f(v,V)=\sum_{u\in V} f(u,v)-\sum_{w\in V} f(v,w)=\begin{cases}0, & v\notin\{s,t\}\\ |f|, & v=s\\ -|f|, & v=t\end{cases} \quad (1)$$

where $|f|$ is the value of the flow. In other words, the total flow in a node v, $f(V,v)$, must equal the total flow out of the node v, $f(v,V)$, $\forall v \in V\setminus\{s, t\}$; the flow leaving s and the flow entering t is equal to the value of the flow.

3. For all vertices u, $v\in V$, if $f(u, v) > 0$, then $f(v, u) = 0$. In other words, each flow uses a given edge only in one direction.
   Separately, it is assumed that if there is no edge $\{u, v\}$, i.e. $\{u, v\}\notin E$, then $f(u, v) = 0$.

A flow is a *maximum flow* if it has the largest possible value among all flows from s to t in a given capacity network [8].

A *cut* (S, T) of the undirected flow network $G(V, E, c)$ is a partition of V into two subsets S and $T = V \setminus S$, such that $s\in S$ and $t\in T$, as shown on Fig. 1. The *capacity* of the cut (S, T) is

$$c(S,T)=\sum_{u\in S}\sum_{v\in T} c(\{u,v\}) \quad (2)$$

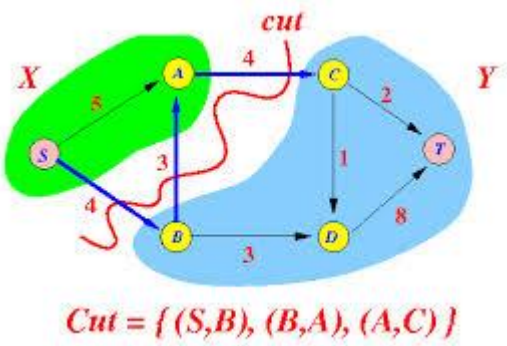


Figure 1. An example of a cut (S, T) of an undirected flow network

A *minimum cut* of a network is a cut whose capacity is minimum over all cuts. A known fact (in both directed and undirected networks) is that the capacity of the minimum cut is equal to the maximal flow of the network, and the capacity of each cut is smaller or equal to the maximal flow of that network. Moreover, for a given flow f,

$$f(S,T)=\sum_{u\in S}\sum_{v\in T} f(u,v)-\sum_{u\in S}\sum_{v\in T} f(v,u)=|f| \quad (3)$$

for any cut (S, T).

On Fig. 2 we show an example where the min cut has flow of 23, which is the maximal flow of the network.

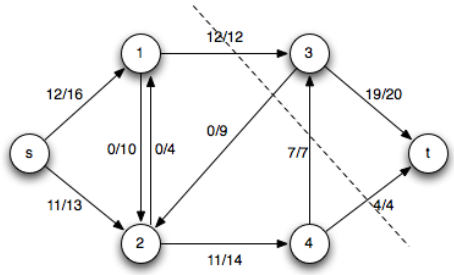


Figure 2. An example of a minimum cut of a network producing the maximal flow

III. MULTIDIMENSIONALFLOW NETWORK

As previously stated the novelty of this paper is the introduction of the multidimensional networks i.e. networks that have multiple edges between the same pair of nodes. We note here that we only view the scenario where the same set of nodes is connected via multiple links. Without losing generality we take that all nodes of the network are in such a set (in case of a link not being present we can assume a value 0 for it). We give the definitions of the basic concepts for such networks.

Definition 1. We define a two-terminal multidimensional flow network as a directed or undirected graph $G(V, E)$, with two special vertices, a source s and a sink t ($s \neq t$), in which each edge $(u, v)\in E$ ($\{u, v\}\in E$) has a nonnegative vector capacity $(c_1(u, v),\ldots,c_n(u, v))$, where $c_i(\{u, v\}) \geq 0$. The function $\vec{c}$ is called *vector capacity function*. Shortly, we will denote such a capacity network by $G(V, E, \vec{c})$. The flow of multidimensional flow network is defined similarly as in the one dimensional flow network, but instead of c we use $\sum c_i(u,v)$.

Definition 2. We define multidimensional cut the same as one dimensional cut. The *capacity vector* of the multidimensional cut (S, T) is

$$\vec{c}(S,T)=\sum_{u\in S}\sum_{v\in T} \vec{c}(\{u,v\}) \quad (4)$$

A *minimum multidimensional cut* of a network is a multidimensional cut such that there is no other cut vector with greater capacity vector (its capacity is minimum over all cuts).

Note that there may be more than one minimum multidimensional cut. Moreover there may be different minimum cuts with maximal flow. The set of all minimum multidimensional cuts will be called minimum multidimensional cut set (MMCS).

IV. Problem Formalization and Solution

We next formally define the problem of maximizing flow in a two-terminal multidimensional flow network.

Definition 3. Given a multidimensional flow network $G(V,E,\vec{c})$ and a positive vector $\vec{m}$ with the same dimension as $\vec{c}$, we define a linear combination capacity flow network as $G(V, E, \vec{m}^T\vec{c})$, (The function $\vec{m}^T\vec{c}$ is defined as $\vec{m}^T\vec{c} = \sum m_i c_i$).

Our problem is to find a vector $\vec{m}$, having $\sum m_i = 1$, such that the network $G(V, E, \vec{m}^T\vec{c})$ has the maximum possible maximal flow. We will refer to this flow as optimal (greater) normalized flow. The vector $\vec{m}$ that produces the optimal normalized flow will be referred to as the optimal linear combination.

The following proposition is clear.

Proposition 1. Given a multidimensional flow network $G(V, E, \vec{c})$, let (S, T) and (S1, T1) be the multidimensional cuts such that $\vec{c}(S,T) \le \vec{c}(S1,T1)$. Then $\vec{m}^T\vec{c}(S,T) \le \vec{m}^T\vec{c}(S1,T1)$.

Proposition 1 tells us that to calculate the optimal normalized flow it is sufficient to find only multidimensional minimum cuts, and calculate flows over all linear combinations of that vectors.

Proposition 2. Given a multidimensional flow network $G(V, E, \vec{c})$ with MMCS A, $\vec{\hat{m}}$ is the optimal linear combination iff for all other $\vec{m}$.

$$\min_{(Si,Ti)\in A} \vec{\hat{m}}^T\vec{c}(Si,Ti) \le \min_{(Si,Ti)\in A} \vec{m}^T\vec{c}(Si,Ti) \tag{5}$$

Moreover the optimal normalized flow is

$$\max_{\vec{m}} \min_{(Si,Ti)\in A} \vec{m}^T\vec{c}(Si,Ti) \tag{6}$$

Having in mind that each $\vec{m}^T\vec{c}(Si,Ti)$ is a linear expression, taking all minimum multidimensional cuts, we derive a system of such linear expressions. We need to maximize the expression with minimum value. To solve this problem we assume that the cut (Sk, Tk) has the minimum value. Then we have that for all other *i* different from *k*

$$\vec{m}^T\vec{c}(Sk,Tk) \le \vec{m}^T\vec{c}(Si,Ti) \tag{7}$$

Now, assuming that (Sk, Tk) gives the optimal solution we need to solve the problem for maximizing $\vec{m}^T\vec{c}(Sk,Tk)$ under the constraints as defined in (7) and $\sum m_i = 1$. This can be solved using the simplex method [9]. Solving this for all (Sk, Tk) in the MMCS *A*, we will take the maximal sub-solution over all solutions, and that will be our optimal solution.

V. Conclusion

In this paper we introduce the concept of multidimensionality that is inherent to many systems being modeled using multistate flow networks. We tackle the problem of flow maximization in a two-terminal multidimensional flow network. We first give an overview of the concepts in a single dimension system and then generalize them to a multidimensional one. We formally define the problem of flow maximization in the simplified environment we consider and give a formal solution to the problem.

Acknowledgment

This work is supported by the project “Computational analysis of human brain data”, financed by the Faculty of Computer Science and Engineering University “Ss. Cyril and Methodius” in Skopje, Macedonia.

References

[1] Aven, T. (1987). Availability evaluation of oil/gas production and transportation systems. *Reliability engineering*, *18*(1), 35-44.

[2] Yeh, W. C. (2011). An improved method for multistate flow network reliability with unreliable nodes and a budget constraint based on path set. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*,*41*(2), 350-355.

[3] Yan, Z., & Qian, M. (2007). Improving efficiency of solving d-MC problem in stochastic-flow network. *Reliability Engineering & System Safety*, *92*(1), 30-39.

[4] Kivelä, M., Arenas, A., Barthelemy, M., Gleeson, J. P., Moreno, Y., & Porter, M. A. (2014). Multilayer networks. *Journal of Complex Networks*, *2*(3), 203-271.

[5] Boccaletti, S., et al (2014). The structure and dynamics of multilayer networks. *Physics Reports*, *544*(1), 1-122.

[6] Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications* (Vol. 8). Cambridge university press.

[7] Gomez, S., Diaz-Guilera, A., Gomez-Gardeñes, J., Perez-Vicente, C. J., Moreno, Y., & Arenas, A. (2013). Diffusion dynamics on multiplex networks.*Physical review letters*, *110*(2), 028701.

[8] Cormen, H.T., Leiserson, E.C., Rivest, R.L., Stein, C,”*Introduction to Algorithms*,” Third edition. The MIT Press, 2009.

[9] Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. *The computer journal*, *7*(4), 308-313.

Towards Cryptanalysis of ARX Based Ciphers

Simona Samardjiska and Hristina Mihajloska
FCSE, "Ss Cyril and Methodius" University, Skopje, Republic of Macedonia
Email: {simona.samardjiska,hristina.mihajloska}@finki.ukim.mk

***Abstract*—Recently, Gligoroski et al. proposed a new authenticated cipher, the π-Cipher, that has entered the second round of the ongoing CAESAR competition for authenticated encryption ciphers. π-Cipher has an ARX based design, and therefore falls in the group of ciphers that are very difficult to analyze using standard techniques from symmetric cryptography.**

In order to analyze the differential properties of π-Cipher, we extend the framework proposed by Mouha et al. that models ARX operations as deterministic finite state automata (DFA). We model one part of the permutation function of π-cipher and show that the model provides practical tool for correctly analyzing more complex ARX functions. We show that without the proposed model there is a serious deviation in the probability estimates of the differentials in the cipher. Furthermore, we provide an algorithm that directly produces a minimized DFA in an automated manner. Previously, the matrices of the DFA were either derived theoretically or needed additional processing in order to be minimized.

Keywords–authenticated encryption, AEAD, ARX operations, differential cryptanalysis, ARX tools, π-Cipher

I. Introduction

Linear and differential cryptanalysis, since their emergence in cryptology, have proven to be the most viable tool for estimating the security of symmetric cryptographic primitives. First introduced by Biham and Shamir [1], differential cryptanalysis investigates how a difference in the input propagates through the cipher and influences the output. Serious deviations from uniformly random differentials implicate weaknesses in the cipher that can be used to devise distinguishes or key recovery attacks. For this reason, differential cryptanalysis has become a primer tool for gaining confidence in ciphers that are to be used in practice.

While there is a myriad of research done on the topic of differential cryptanalysis for Substitution-Permutation networks, very little is known about the differential properties of ARX based designs. ARX ciphers, using only the operation addition modulo 2^n, rotation and XOR, have become increasingly popular in the last few years, mainly because of their performance advantages and easy implementation not prone to errors. The mixture of these operations gives a good crypto primitive if the number of rounds is sufficient. However, the classical analysis suitable for SP networks does not apply here.

A. Common problems and erroneous results on ARX ciphers

Two groups of researchers from France [7] and Belgium [11] have pioneered in automated investigation of the differential properties of ARX ciphers, but so far, the tools are either dedicated or restricted to a small set of operations. The biggest problem in analyzing ARX ciphers comes from the size of the inputs that can be up to 64 or even 128 bits long as opposed to SP networks, where the S-boxes used, act on usually at most 8 bits. Therefore it is unfeasible to build differential tables. Also, the components of a complex ARX structure interact very differently depending on the particular design, so it is still difficult to come up with a general framework for compositions of several ARX operations. So far, a common strategy was to look at each operation individually, and then multiply the probabilities of each non-linear operation. This approach was shown to be erroneous in many differential attacks (for full reference see [7]).

B. Our Contribution

Recently, Gligoroski et al. [3] proposed a new authenticated cipher, the π-Cipher, that has entered the second round of the ongoing CAESAR competition for authenticated encryption ciphers [2]. π-Cipher has an ARX based design, arising from a particular quasigroup. From quasigroup theory perspective, the design was carefully made to reduce highly probable differentials. However, only a detailed cryptanalysis can prove the security of the cipher, and increase the confidence in it.

Our motivation being the differential properties of π-Cipher, we extend the framework proposed by Mouha et al. [11] in order to correctly analyze the cipher. We consider one segment of the quasigroup operation consisting of a parallel execution of four additions modulo 2^n and show how this part can be modeled as a deterministic finite state automaton (DFA). While the use of DFA was proposed by Mouha et al., so far it has not been used for more complex functions. This modeling provides a tool for accurate estimation of the probability of any differential of the segment, which is not possible using the known techniques. As a proof of concept we show that without the proposed model there is a serious deviation in the probability estimates. Furthermore, we provide an algorithm that directly produces a minimized DFA in an automated manner. Previously, the matrices of the DFA were either derived theoretically or needed additional processing in order to be minimized. This improvement makes the procedure feasible, and drastically less memory demanding.

II. The ARX structure of π-Cipher

π-Cipher is a proposal for an authenticated cipher with associated data for the ongoing "CAESAR" crypto competition. The scope of this competition is not just to seek for authenticated modes of operations for AES, but also for proposals of new ciphers that offer advantages over AES-GCM and are suitable for widespread adoption. The recent developments with the introduction of AES-NI instructions in latest Intel CPUs [5] made AES-GCM mode really efficient.

The core part of every sponge construction is the permutation function, and the whole security of the primitive relies on it. The design goal for the sponge construction in π-Cipher was to obtain a strong permutation, which for different values of the bit size ω of the words provides different features, i.e. to be very efficient when $\omega = 64$ and lightweight when $\omega = 16$.

π-Cipher has an ARX based permutation function which we denote as π function. It uses similar operations as the ones used in the hash function Edon-R [4] but instead of using 8–tuples here we use 4–tuples. The permutation operates on a b bits state and updates the internal state through a sequence of R successive rounds. The state IS can be represented as a list of N 4-tuples, each of length ω-bits, where $b = N \times 4 \times \omega$, i.e.,

$$IS = (\underbrace{(IS_{11}, \ldots, IS_{14})}_{I_1}, \ldots, \underbrace{(IS_{N1}, \ldots, IS_{N4})}_{I_N}). \quad (1)$$

The general permutation function π consists of three main transformations $\mu, \nu, \sigma : \mathbb{Z}_{2^\omega}^4 \rightarrow \mathbb{Z}_{2^\omega}^4$, where $\mathbb{Z}_{2^\omega}$ is the set of all integers between 0 and $2^\omega - 1$. These transformations perform diffusion and nonlinear mixing of the input. It uses the following operations:

- Addition + modulo 2^ω;
- Left rotation (circular left shift) $ROTL^r(X)$, where X is an ω–bit word and r is an integer, $0 \leq r < \omega$;
- Bitwise XOR operation $\oplus$ on ω–bit words.

Let $\mathbf{X} = (X_0, X_1, X_2, X_3)$, $\mathbf{Y} = (Y_0, Y_1, Y_2, Y_3)$ and $\mathbf{Z} = (Z_0, Z_1, Z_2, Z_3)$ be three 4-tuples of ω–bit words. Further, let $*$ be defined as:

$$\mathbf{Z} = \mathbf{X} * \mathbf{Y} \equiv \sigma(\mu(\mathbf{X}) \boxplus_4 \nu(\mathbf{Y})) \quad (2)$$

where $\boxplus_4$ is the component-wise addition of two 4-dimensional vectors in $\left(\mathbb{Z}_{2^\omega}\right)^4$.

Table I. An algorithmic description of the ARX operation $*$ for ω–bit words.

$*$ **operation for ω–bit words**		
Input: $\mathbf{X} = (X_0, X_1, X_2, X_3)$ and $\mathbf{Y} = (Y_0, Y_1, Y_2, Y_3)$ where X_i and Y_i are ω–bit variables. **Output:** $\mathbf{Z} = (Z_0, Z_1, Z_2, Z_3)$ where Z_i are $omega$–bit variables. **Temporary ω–bit variables:** $T_0, \ldots, T_{11}$.		
μ–transformation:		
1)	T_0	$\leftarrow ROTL^{r_{1,\omega,1}}(const_{1,\mu\omega} + X_0 + X_1 + X_2)$;
	T_1	$\leftarrow ROTL^{r_{1,\omega,2}}(const_{2,\mu\omega} + X_0 + X_1 + X_3)$;
	T_2	$\leftarrow ROTL^{r_{1,\omega,3}}(const_{3,\mu\omega} + X_0 + X_2 + X_3)$;
	T_3	$\leftarrow ROTL^{r_{1,\omega,4}}(const_{4,\mu\omega} + X_1 + X_2 + X_3)$;
2)	T_4	$\leftarrow T_0 \oplus T_1 \oplus T_3$;
	T_5	$\leftarrow T_0 \oplus T_1 \oplus T_2$;
	T_6	$\leftarrow T_1 \oplus T_2 \oplus T_3$;
	T_7	$\leftarrow T_0 \oplus T_2 \oplus T_3$;
ν–transformation:		
1)	T_0	$\leftarrow ROTL^{r_{2,\omega,1}}(const_{1,\nu\omega} + Y_0 + Y_2 + Y_3)$;
	T_1	$\leftarrow ROTL^{r_{2,\omega,2}}(const_{2,\nu\omega} + Y_1 + Y_2 + Y_3)$;
	T_2	$\leftarrow ROTL^{r_{2,\omega,3}}(const_{3,\nu\omega} + Y_0 + Y_1 + Y_2)$;
	T_3	$\leftarrow ROTL^{r_{2,\omega,4}}(const_{4,\nu\omega} + Y_0 + Y_1 + Y_3)$;
2)	T_8	$\leftarrow T_1 \oplus T_2 \oplus T_3$;
	T_9	$\leftarrow T_0 \oplus T_2 \oplus T_3$;
	T_{10}	$\leftarrow T_0 \oplus T_1 \oplus T_3$;
	T_{11}	$\leftarrow T_0 \oplus T_1 \oplus T_2$;
σ–transformation::		
1)	Z_3	$\leftarrow T_4 + T_8$;
	Z_0	$\leftarrow T_5 + T_9$;
	Z_1	$\leftarrow T_6 + T_{10}$;
	Z_2	$\leftarrow T_7 + T_{11}$;

An algorithmic definition of the $*$ operation over two 4–dimensional vectors $\mathbf{X}$ and $\mathbf{Y}$ for ω-bit words is given in Table I. The values of the rotation vectors $r_{1,\omega}$ and $r_{2,\omega}$ and of the constants $const_{i,\mu\omega}$, $const_{i,\nu\omega}$, $i = 1, 2, 3, 4$ used in the μ and ν transformations are given in the official documentation of the π-Cipher [**?**]. A graphical representation of the $*$ operation is given in Figure 1.

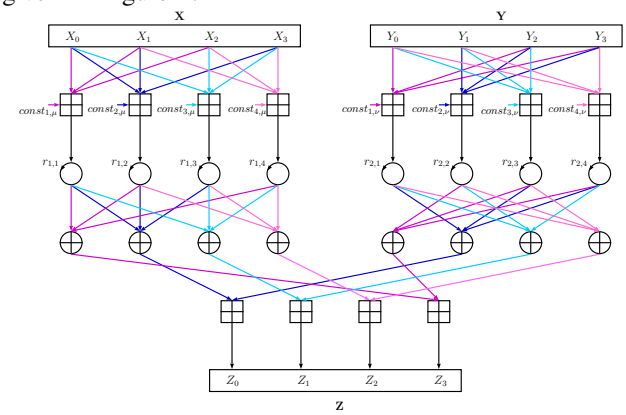


Figure 1. Graphical representation of the ARX operation $*$.

One round of the cipher is graphically described in Figure 2. In the figure, the diagonal arrows can be interpreted as $*$ operations between the source and destination, and the vertical or horizontal arrows as equality signs " $=$ ".

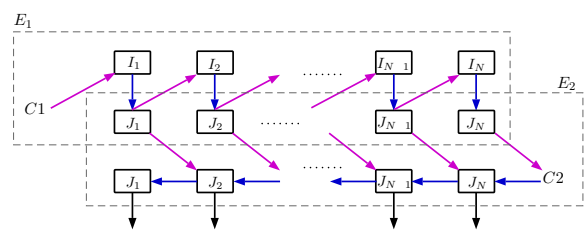


Figure 2. One round of π-Cipher

III. Differential properties of S-functions

S-functions (short for "State functions") were introduced by Mouha et al. in [11] where they provide a general framework for analysis of S-functions based on a DFA representation of the S-functions.

An S-function accepts n-bit words $a_1, a_2 \ldots, a_k$ and a list of states $S[i]$, $0 \leqslant i < n$ as input and produces an n-bit output b recursively as:

$$S[0] = 0,$$
$$(b[i], S[i+1]) = f(a_1[i], a_2[i], \ldots, a_k[i], S[i]) \quad (3)$$

Schematically, an S-function can be represented as in Fig 3 [11].

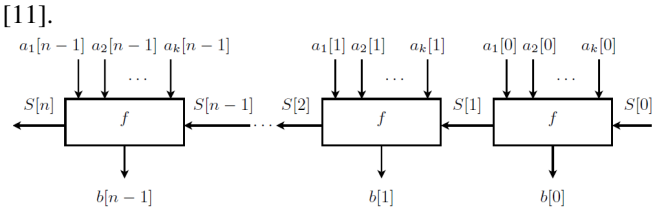


Figure 3. Representation of an S-function [11]

The operations of addition modulo 2^n and bitwise XOR are typical examples for S-functions. We will be interested in operations including addition of three words, so we provide a detailed description of the S-function representation for the output difference of modular addition with three inputs from [11].

Let x_1, y_1, z_1 be three n-bit words, and let $\Delta^{\oplus}x$, $\Delta^{\oplus}y$, $\Delta^{\oplus}z$ denote the differences for the three words respectively,

i.e. let

$$x_2 \leftarrow x_1 \oplus \Delta^{\oplus} x,\ y_2 \leftarrow y_1 \oplus \Delta^{\oplus} y,\ z_2 \leftarrow z_1 \oplus \Delta^{\oplus} z. \quad (4)$$

Then if we let

$$w_1 \leftarrow x_1 + y_1 + z_1,\ w_2 \leftarrow x_2 + y_2 + z_2, \quad (5)$$

$\Delta^{\oplus} w$ can be found as

$$\Delta^{\oplus} w \leftarrow w_1 \oplus w_2. \quad (6)$$

The differential probability $\mathrm{xdp}^{+}(\alpha, \beta, \gamma \rightarrow \delta)$ of the 3-input modular addition is defined as:

$$\mathrm{xdp}^{+}(\alpha, \beta, \gamma \rightarrow \delta) = \frac{|\{(x_1, y_1, z_1) : \Delta^{\oplus} x = \alpha, \Delta^{\oplus} y = \beta, \Delta^{\oplus} z = \gamma, \Delta^{\oplus} w = \delta\}|}{|\{(x_1, y_1, z_1) : \Delta^{\oplus} x = \alpha, \Delta^{\oplus} y = \beta, \Delta^{\oplus} z = \gamma\}|} = \frac{|\{(x_1, y_1, z_1) : \Delta^{\oplus} x = \alpha, \Delta^{\oplus} y = \beta, \Delta^{\oplus} z = \gamma, \Delta^{\oplus} w = \delta\}|}{8^n}.$$

On a bit level, in radix 2 representation the formulas (4),(5),(6) can be written as:

$$\begin{aligned}
c_1[0] &\leftarrow 0, && (7)\\
c_2[0] &\leftarrow 0, && (8)\\
x_2[i] &\leftarrow x_1[i] \oplus \Delta^{\oplus} x[i], && (9)\\
y_2[i] &\leftarrow y_1[i] \oplus \Delta^{\oplus} y[i], && (10)\\
z_2[i] &\leftarrow z_1[i] \oplus \Delta^{\oplus} z[i], && (11)\\
w_1[i] &\leftarrow x_1[i] + y_1[i] + z_1[i] + c_1[i], && (12)\\
c_1[i+1] &\leftarrow (x_1[i] + y_1[i] + z_1[i] + c_1[i]) \gg 1, && (13)\\
w_2[i] &\leftarrow x_2[i] + y_2[i] + z_2[i] + c_2[i], && (14)\\
c_2[i+1] &\leftarrow (x_2[i] + y_2[i] + z_2[i] + c_2[i]) \gg 1, && (15)\\
\Delta^{\oplus} w &\leftarrow w_1[i] \oplus w_2[i]. && (16)
\end{aligned}$$

Taking

$$\begin{aligned}
S_{xyz}[i] &\leftarrow (c_1[i], c_2[i]), && (17)\\
S_{xyz}[i+1] &\leftarrow (c_1[i+1], c_2[i+1]), && (18)
\end{aligned}$$

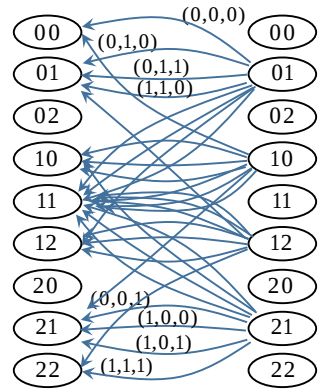
the S-function representing the output difference of modular addition of three n-bit words x_1, y_1, z_1 can be written as:

$$(\Delta^{\oplus} w[i], S_{xyz}[i+1]) = +_{xyz}(x_1[i], y_1[i], z_1[i], \Delta^{\oplus} x[i], \Delta^{\oplus} y[i], \Delta^{\oplus} z[i], S_{xyz}[i]) \quad (19)$$

Note that the carries $c_1[i]$ and $c_2[i]$ can take values from the set $\{0, 1, 2\}$.

The function (19) can be represented through bipartite graphs, one for each possible input-output difference $d[i] = (\alpha[i], \beta[i], \gamma[i], \delta[i]) \leftarrow (\Delta^{\oplus} x[i], \Delta^{\oplus} y[i], \Delta^{\oplus} z[i], \Delta^{\oplus} w[i])$. In a single graph, the vertices $(c_1[i], c_2[i]), (c_1[i+1], c_2[i+1]) \in \{(0,0), (0,1), (0,2), (1,0), (1,1), (1,2), (2,0), (2,1), (2,2)\}$ correspond to the states $S[i]$ and $S[i+1]$ respectively, and each edge is a possible transition for each triple (x_1, y_1, z_1). Each graph can be represented by its adjacency matrix, so in total the function (19) is completely defined by 2^4 square matrices of order 9. We denote these matrices by $A_{d[i]}$. For example, the graph of $(0, 0, 1, 1)$, and the corresponding matrix A_{0011} is given in Figure 4.

Combining n graphs for each bit position $0 \leqslant i < n$ of a differential $(\alpha, \beta, \gamma \rightarrow \delta)$, a larger graph is obtained, in which all paths from an initial state $(0, 0)$ to any final state $(c_1[n], c_2[n])$ satisfy the differential. Using a well known fact



$$A_{0011} = \begin{bmatrix} 0&1&0&1&0&0&0&0&0\\ 0&3&0&0&0&1&0&0&0\\ 0&0&0&0&0&0&0&0&0\\ 0&0&0&3&0&0&0&1&0\\ 0&3&0&3&0&3&0&3&0\\ 0&1&0&0&0&3&0&0&0\\ 0&0&0&0&0&0&0&0&0\\ 0&0&0&1&0&0&0&3&0\\ 0&0&0&0&0&1&0&1&0 \end{bmatrix}$$

Figure 4. The graph of $(0, 0, 1, 1)$, and the corresponding matrix A_{0011}. For clarity of the figure, only few of the edges are labeled with the corresponding inputs.

from graph theory, the number of paths can be calculated as a product of the corresponding adjacency matrices. As a result, the probability of a differential $(\alpha, \beta, \gamma \rightarrow \delta)$ can be calculated as:

$$\mathrm{xdp}^{+}(\alpha, \beta, \gamma \rightarrow \delta) = 8^{-n} L A_{d[n-1]} \ldots A_{d[1]} A_{d[0]} C \quad (20)$$

where $L = [1\ 1 \ldots 1]$ and $C^{\top} = [1\ 0 \ldots 0]$ are 1×9 vectors.

IV. Generalization to multi-output S-functions

Consider the following segment of the μ transformation, that we will call $\mu+$ transformation:

$$\begin{aligned}
T'_0 &\leftarrow X_0 + X_1 + X_2,\\
T'_1 &\leftarrow X_0 + X_1 + X_3,\\
T'_2 &\leftarrow X_0 + X_2 + X_3,\\
T'_3 &\leftarrow X_1 + X_2 + X_3.
\end{aligned},$$

or schematically the transformation given in Figure 5.

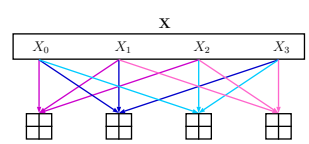


Figure 5. The $\mu+$ transformation

Modular addition is the only transformation performed in this part. However, for a given input difference, knowing the differential probabilities p_1, p_2, p_3, p_4 of each of the four modular additions, does not necessarily lead to the differential probabilities of the $\mu+$ transformation. Indeed, since the inputs in the four additions are not independent, except for very specific choices of the inputs, it is wrong to assume that the total probability can be computed as a product $p_1 p_2 p_3 p_4$. Unfortunately, as mentioned in the Introduction, due to lack of a efficient modeling of combinations of ARX operations, such assumptions are regularly made in cryptanalyses of ARX ciphers, leading to significant deviations of the estimated probability and erroneous attacks.

Therefore, we generalize the framework of Mouha et al.[11] and define a multi-output S-function, that will be used to estimate the differential probabilities of the $\mu+$ transformation.

A multi-output S-function accepts n-bit words $a_1, a_2 \ldots, a_k$ and a list of states $S[i]$, $0 \leqslant i < n$ as input and produces output $b_1, \ldots b_m$ of m n-bit words recursively as:

$$S[0] = 0,$$
$$(b_1[i], \ldots, b_m[i], S[i+1]) = f(a_1[i], a_2[i], \ldots, a_k[i], S[i]) \quad (21)$$

Now, let x_1, y_1, z_1, w_1 be four n-bit words, and let

$$x_2 \leftarrow x_1 \oplus \Delta^{\oplus} x,\ y_2 \leftarrow y_1 \oplus \Delta^{\oplus} y, \quad (22)$$
$$z_2 \leftarrow z_1 \oplus \Delta^{\oplus} z,\ w_2 \leftarrow w_1 \oplus \Delta^{\oplus} w. \quad (23)$$

Let

$$q_1 \leftarrow x_1 + y_1 + z_1,\ q_2 \leftarrow x_2 + y_2 + z_2, \quad (24)$$
$$r_1 \leftarrow x_1 + y_1 + w_1,\ r_2 \leftarrow x_2 + y_2 + w_2, \quad (25)$$
$$s_1 \leftarrow x_1 + z_1 + w_1,\ s_2 \leftarrow x_2 + z_2 + w_2, \quad (26)$$
$$t_1 \leftarrow y_1 + z_1 + w_1,\ t_2 \leftarrow y_2 + z_2 + w_2. \quad (27)$$

The output differences will be

$$\Delta^{\oplus} q \leftarrow q_1 \oplus q_2,\ \Delta^{\oplus} r \leftarrow r_1 \oplus r_2, \quad (28)$$
$$\Delta^{\oplus} s \leftarrow s_1 \oplus s_2,\ \Delta^{\oplus} t \leftarrow t_1 \oplus t_2. \quad (29)$$

Thus, similarly as for addition, the differential probability $\mathrm{xdp}^{\mu+}(\alpha, \beta, \gamma, \delta \rightarrow \phi, \chi, \psi, \omega)$ of the $\mu+$ transformation is defined as:

$$\mathrm{xdp}^{\mu+}(\alpha, \beta, \gamma, \delta \rightarrow \phi, \chi, \psi, \omega) = \frac{1}{16^n} \cdot |\{(x_1, y_1, z_1, w_1) : \Delta^{\oplus} x = \alpha, \Delta^{\oplus} y = \beta, \Delta^{\oplus} z = \gamma, \Delta^{\oplus} w = \delta, \Delta^{\oplus} q = \phi, \Delta^{\oplus} r = \chi, \Delta^{\oplus} s = \psi, \Delta^{\oplus} t = \omega\}|.$$

Using the definition of the states (17) of the S-function (19), we define the states of the $\mu+$ transformation as:

$$S_{\mu+}[i] \leftarrow (S_{xyz}[i], S_{xyw}[i], S_{xzw}[i], S_{yzw}[i]), \quad (30)$$
$$S_{\mu+}[i+1] \leftarrow (S_{xyz}[i+1], S_{xyw}[i+1], S_{xzw}[i+1], S_{yzw}[i+1]),$$

Hence, the multi-output S-function representing the output difference of $\mu+$ is:

$$(\Delta^{\oplus} q[i], \Delta^{\oplus} r[i], \Delta^{\oplus} s[i], \Delta^{\oplus} t[i], S_{\mu+}[i+1]) = {}_{\mu+}(x_1[i], y_1[i], z_1[i], w_1[i], \Delta^{\oplus} x[i], \Delta^{\oplus} y[i], \Delta^{\oplus} z[i], \Delta^{\oplus} w[i], S_{\mu+}[i]) \quad (31)$$

From the definition of a multi-output S-function, it is straightforward to see that these functions can be represented using DFAs in a similar manner as S-functions.

For the $\mu+$ function, there are 2^8 possible differences $d[i] = (\alpha[i], \beta[i], \gamma[i], \delta[i], \phi[i], \chi[i], \psi[i], \omega[i])$, and for a given difference, the number of possible states $S_{\mu+}[i]$ is 9^4. Therefore the matrices corresponding tho the possible differences are of order 9^4.

Since $\mu+$ is composed of four modular additions, we have the following theorem about the graph of a given input-output difference $d[i]$ (the proof will be provided in the extended version of the paper):

Theorem 1: Let $d[i] = (\alpha[i], \beta[i], \gamma[i], \delta[i], \phi[i], \chi[i], \psi[i], \omega[i])$ be a given input-output difference for the function $\mu+$. There is an edge between a state $S_{\mu+}[i]$ and a state $S_{\mu+}[i+1]$ in the graph of $d[i]$ for a given input (x, y, z, w) if and only if for the exact same input, there is an edge between (all of) the states:

- $S_{xyz}[i]$ and $S_{xyz}[i+1]$, in the graph of $(\alpha[i], \beta[i], \gamma[i], \phi[i])$ for the function $+_{xyz}$,
- $S_{xyw}[i]$ and $S_{xyw}[i+1]$, in the graph of $(\alpha[i], \beta[i], \delta[i], \chi[i])$ for the function $+_{xyw}$,
- $S_{xzw}[i]$ and $S_{xzw}[i+1]$, in the graph of $(\alpha[i], \gamma[i], \delta[i], \psi[i])$ for the function $+_{xzw}$,
- $S_{yzw}[i]$ and $S_{yzw}[i+1]$, in the graph of $(\beta[i], \gamma[i], \delta[i], \omega[i])$ for the function $+_{yzw}$.

The previous theorem completely determines the DFA of the function $\mu+$ through the DFAs of $+_{xyz}$, $+_{xyw}$, $+_{xzw}$, $+_{yzw}$. However, despite this connection, it is very hard to estimate the probability of a given difference $d[i] = (\alpha[i], \beta[i], \gamma[i], \delta[i], \phi[i], \chi[i], \psi[i], \omega[i])$ solely through the matrices of $+_{xyz}$, $+_{xyw}$, $+_{xzw}$, $+_{yzw}$. On the other hand, the probability can be precisely determined using the matrices for $\mu+$. Unfortunately, the size of the matrices makes this process highly impractical.

V. Indistinguishable states for $\mu+$

The DFA that we constructed for the $\mu+$ function has 9^4 states. However, it is possible that some of the states are indistinguishable which may lead to a smaller representation of the DFA, and consequently to smaller transition matrices.

For example, Lipmaa et al. [9] determine that the matrices for modular addition of two numbers are of order 2 (note that the DFA constructed in a similar manner as in the previous section would contain 4 states). Later, Mouha et al.[11] use a minimization algorithm [6] to reduce the number of states for (among other) modular addition of three numbers and show that in this case the number of distinguishable states is actually 4 (initially the DFA has 9 states).

Here we take another approach. Instead of first creating the matrices of order $9^4 = 6561$ and then reducing them to a smaller dimension using some minimization algorithm, we show how to initially avoid unnecessary distinction of states that are actually indistinguishable.

The following theorem holds (the proof will be provided in the extended version of the paper).

Theorem 2: Let

$$S_{\mu+}[i] \leftarrow (S_{xyz}[i], S_{xyw}[i], S_{xzw}[i], S_{yzw}[i])$$

be a state in the DFA of $\mu+$, where

$$S_{jkl}[i] \leftarrow (c_1^{(jkl)}[i], c_2^{(jkl)}[i]),$$
$$(jkl) \in \{(xyz), (xyw), (xzw), (yzw)\}$$

The following states are indistinguishable from $((c_1^{(xyz)}[i], c_2^{(xyz)}[i]), (c_1^{(xyw)}[i], c_2^{(xyw)}[i]), (c_1^{(xzw)}[i], c_2^{(xzw)}[i]), (c_1^{(yzw)}[i], c_2^{(yzw)}[i]))$:

- $((2-c_1^{(xyz)}[i], 2-c_2^{(xyz)}[i]), (2-c_1^{(xyw)}[i], 2-c_2^{(xyw)}[i]), (2-c_1^{(xzw)}[i], 2-c_2^{(xzw)}[i]), (2-c_1^{(yzw)}[i], 2-c_2^{(yzw)}[i]))$,
- $((c_2^{(xyz)}[i], c_1^{(xyz)}[i]), (c_2^{(xyw)}[i], c_1^{(xyw)}[i]), (c_2^{(xzw)}[i], c_1^{(xzw)}[i]), (c_2^{(yzw)}[i], c_1^{(yzw)}[i]))$.

Example 1: For instance, $((0,2),(1,1),(1,2),(2,2))$, $((0,2),(1,1),(0,1),(0,0))$, $((2,0),(1,1),(2,1),(2,2))$, are all indistinguishable from $((2,0),(1,1),(1,0),(0,0))$.
Using Theorem 2, it can easily be calculated that discarding the distinguishable states in the DFA of $\mu+$ leads to a smaller representation of only 2071 states.

Thus, since we a priori know the distinguishable states, we can directly create the DFA in a minimized form. This renders the minimization process obsolete, but also the storage (several gygabytes) and manipulation of extremely big matrices. Note that our strategy is applicable to the modular addition functions analyzed in [11], and to any multi-output S-function involving parallel execution of modular additions. Even more, it can be generalized to other types of S-functions by providing an appropriate theorem similar to Theorem 2.

We have implemented the algorithm for the $\mu+$ function, and leave other types of ARX functions for future work.

Example 2: Let

$$\alpha, \beta, \gamma, \delta \rightarrow \phi, \chi, \psi, \omega) = (0xF16A, 0xD26A, 0x8084, 0xAC4 \rightarrow 0x4FC4, 0x55D8, 0xF822, 0xF3DA)$$

Then, using the matrices of the $\mu+$ function we find that the probability of this differential is 0, i.e. this is an impossible differential. On the other hand, if we consider all of the modular additions separately, we obtain the following probabilities:

$$\mathrm{xdp}^+(0xF16A, 0xD26A, 0x8084 \rightarrow 0x4FC4) \approx 2^{-14},$$
$$\mathrm{xdp}^+(0xF16A, 0xD26A, 0xAC4 \rightarrow 0x55D8) \approx 2^{-14},$$
$$\mathrm{xdp}^+(0xF16A, 0x8084, 0xAC4 \rightarrow 0xF822) \approx 2^{-13},$$
$$\mathrm{xdp}^+(0xD26A, 0x8084, 0xAC4 \rightarrow 0xF3DA) \approx 2^{-16}.$$

This means that knowing only the differential probabilities xdp^+ of the components of the $\mu+$ function, it can not be deduced that a given differential is impossible. Note that, many of the erroneous attacks on ARX ciphers published so far, fail exactly at this point: the probability being wrongfully estimated, the attack is based on a differential that is not possible.

VI. Conclusion

In this paper we take a step towards a general framework for analysis of ARX ciphers. We model a combination of modular additions arising from the permutation function in π-cipher as a Discrete Finite Automaton. This modeling, introduced by Mouha et al., provides a tool for correct analysis of the differential properties of more complex ARX designs such as π-cipher.

As a next step, we plan to use this framework to provide a more concrete evidence of the conjectured resistance of π-cipher to differential cryptanalysis.

References

[1] E. Biham and A. Shamir. Differential Cryptanalysis of DES-like Cryptosystems. J. Cryptology, 4(1):372, 1991.

[2] D. J. Bernstein. Caesar: Competition for authenticated encryption: Security, applicability, and robustness. CAESAR web page, 2013. http://competitions.cr.yp.to/index.html.

[3] Danilo Gligoroski, Hristina Mihajloska, Simona Samardjiska, Hakon Jacobsen, Mohamed El-Hadedy, and Rune Erlend Jensen. Π-cipher v1. Cryptographic competitions: CAESAR, 2014. http://competitions.cr.yp.to/ caesar-submissions.htmls.

[4] Danilo Gligoroski, Rune Steinsmo degrd, Marija Mihova, Svein Johan Knapskog, Ljupco Kocarev, Ales Drpal, and Vlastimil Klima. Cryptographic hash function EDON-R. In 1st International Workshop on Security and Communi- cation Networks, pp. 85–95, Trondheim, Norway.

[5] Shay Gueron. Intels new aes instructions for enhanced performance and security. In Orr Dunkelman, editor, FSE, volume 5665 of Lecture Notes in Computer Science, pages 5166. Springer, 2009.

[6] D. Huffman. The synthesis of sequential switching circuits. Journal of the Franklin Institute, 257(3):161 190, 1954.

[7] Leurent, G.: Analysis of Differential Attacks in ARX Constructions. In: Wang, X., Sako, K. (eds.) ASIACRYPT. Lecture Notes in Computer Science, vol. 7658, pp. 226-243. Springer (2012)

[8] Leurent, G.: Construction of Differential Characteristics in ARX Designs Application to Skein. In: Canetti, R., Garay, J.A. (eds.) CRYPTO. Lecture Notes in Computer Science, vol. 8042, pp. 241-258. Springer (2013)

[9] H. Lipmaa, J. Wallen, and P. Dumas. On the Additive Differential Probability of Exclusive-Or. In B. K. Roy and W. Meier, editors, FSE, volume 3017 of Lecture Notes in Computer Science, pages 317-331. Springer, 2004.

[10] Lipmaa, H., Moriai, S.: Efficient Algorithms for Computing Differential Properties of Addition. In: Matsui, M. (ed.) FSE. Lecture Notes in Computer Science, vol. 2355, pp. 336-350. Springer (2001)

[11] N. Mouha, V. Velichkov, C. De Canni'ere, and B. Preneel. The Differential Analysis of S-Functions. In A. Biryukov, G. Gong, and D. R. Stinson, editors, Selected Areas in Cryptography, volume 6544 of Lecture Notes in Computer Science, pages 36-56. Springer, 2010.

Dataflow DSP Filter for ECG Signals

Ervin Domazet, Marjan Gusev and Sasko Ristov
Ss. Cyril and Methodius University, Faculty of Computer Science and Engineering,
1000 Skopje, Macedonia
e-mail: ervin_domazet@hotmail.com, {marjan.gushev, sashko.ristov}@finki.ukim.mk

***Abstract*—Electrocardiogram (ECG) signal analysis and interpretation is achieved by Digital Signal Processing (DSP). ECG signals usually are accompanied by a lot of noise coming from several sources, including the noise from environment (electrical switching power or other related sources) or the internal noises generated by the human breathing physical movement or similar sources. DSP filters are essential in eliminating the noise and extracting the essential characteristic signal. Afterwards, the main ECG features can be detected and analyzed for further determination of the complex heart condition. This processing of the ECG signals is based on detecting the hidden information and the subtle deviation of the heart rhythm to alternating changes of the wave amplitude. In some specific cases, real time analysis of heart signals can save lives. Due to intensive data processing, sequential algorithms are insufficient to run in real-time, especially when a cloud data server processes thousands of data streams coming from remote wearable ECG sensors.**

In this research, we focus on parallelizing the sequential DSP filter for processing of heart signals on dataflow cores. Dataflow Computing is a completely different paradigm of computing than conventional CPUs, where instructions are parallelized across the available space, rather than time. It is a revolutionary way for High Performance Computing (HPC) solutions. Data streams are optimized by utilizing thousands of dataflow cores, providing order of magnitude speedups. We consider using Maxeler Systems for dataflow computing. The performance of the parallelized code will be compared to that of the sequential code. Our analysis shows speedups linear to the kernel size of the filter.

***Index Terms*—DSP, ECG, Heart Signal, Parallelization, Dataflow Computing, Maxeler.**

I. Introduction

Electrocardiogram (ECG) is a stream of electric impulses generated by the beating heart muscle. They are detected by electrodes placed on human skin, by measuring the electric potential that reaches the skin surface [1]. ECG stream holds cardiovascular condition of the patient. Figure 1 reveals general representation ECG signal with its representative P, QRS and T waves.

Fig. 1. General Representation of ECG Signal.

Interpretation of an ECG stream is essential for a better quality of life. Interpretation along with signal analysis is achieved by Digital Signal Processing (DSP) [2], [3]. Nevertheless, ECG signals are exposed to noise that stem from several sources, varying from environment (electrical switching power, radio waves or other related sources) to the internal noises generated by the human breathing physical movement or similar sources.

Precise interpretation and analysis of the ECG signal can be achieved by eliminating the noise. Essential data preprocessing phase is conducted by the DSP filters. Hereinafter, the main ECG features can be detected and analyzed for further determination of the complex heart condition. Processing of the signal is based on detecting hidden information and the subtle deviation of the heart rhythm to alternating changes of the wave amplitude [4].

Lugovaya [5] had focussed on revealing the efficiency of an ECG signal for identification, when compared to the three efficient biometric methods, i.e identification based on fingerprints, iris or retina, and face. Her experimental results showed that the rate of correct identification was 96%, which gave an insight for considering ECG signal as a new biometric characteristic. What is more important, she was successful in showing the persistence of an individual's ECG characteristics over time (slow

and gradual variations on ECG signal). This in turn makes it possible to detect subtle deviations of the heart rhythm and alternating changes of the wave amplitude.

In some specific cases real time processing of the ECG signal can save lives. Due to intensive data processing, sequential algorithms are insufficient to run in real-time, especially when a cloud data server processes thousands of data streams coming from remote wearable ECG sensors.

In this research, we focus on parallelizing the sequential DSP filter for processing of the heart signals on dataflow cores. The DSP filter is used for preprocessing of the ECG data, in order to eliminate noise from the ECG signal. Based on the noise components of the ECG signal, several filtering methods are available, such as Low pass, High bass and Bandpass filter.

Dataflow Computing is a completely different paradigm of computing than traditional CPUs. Instructions are parallelized across the available space, rather than time. It is a revolutionary way for High Performance Computing (HPC) solutions[6], [7]. Data streams are optimized by utilizing thousands of dataflow cores, providing order of magnitude speedups. Maxeler systems are used for dataflow computing [8]. The performance of the parallelized code is compared to that of the sequential code. Our analysis shows speedups linear to the kernel size of the filter.

The paper is organized as follows Section II presents DSP filters and convolution operators. In Section V, we explain the way we have parallelized the DSP filter by using Dataflow computing, namely Maxeler systems. Results of timings and performance figures of the parallelized code versus the original sequential code are presented in Section IV. Finally, the paper is concluded with a discussion and future work in Section V.

II. Digital Signal Processing

Digital Signal Processing (DSP) is the act of manipulating signals with intention varying from filtering, measuring to producing or compressing analog signals. As the power of computers radically increased during the last decades, so does the power of the DSP [3]. DSP had made a tremendous impact on science and engineering, by providing methodologies to deal with the most powerful technologies.

DSP had revolutionized many fields in science and engineering. There are many industrial sectors benefiting from the advancements on the DSP field such as Medical, Military, Space and Telephone. Electrocardiogram analysis, diagnostic imaging, voice and data compression, radars, secure communication, telephone signal filtering are among the range of revolutionized fields.

The focus of our research is on ECG signal filtering, with the intention to remove the noise that stem from several sources. The commonly used method in DSP filtering is the *convolution*, as one of the most important techniques of signal processing. It is defined as a mathematical operation that combines the *input stream* and the *impulse response* in order to generate a new *output stream*. In case of a filter, the impulse response is known as a *filter kernel*.

Each value of the output stream in digital signal convolution is represented as the sum of input stream multiplied by set of weight coefficients, which define the *impulse response*. The impulse response is the signal that results when a delta function (unit impulse) is the input in the DSP filter.

Denote by $f(i)$ the weight (filter kernel) coefficients in the range $i \in \{-\infty, +\infty\}$ if it is an infinite response filter. We will use finite response filters and the weight coefficients $h(i)$ in the range $i \in \{0, \ldots, M-1\}$, where M is the filter length. Let the input stream consists of elements $x(i)$ and the output stream of elements $y(i)$, for $i = 0, \ldots$. The convolution, as a mathematical operation can be expressed by (1).

$$y(i) = \sum_{j=0}^{M-1} h(j)x(i-j) \tag{1}$$

During this research, we have used the three classic filters to eliminate the noise: Low-Pass, High-Pass and Band-Pass filters.

A. Low-Pass Filter

Low-pass filters are designed to thoroughly weaken all the frequencies above the cutoff frequency, known as a *stopband*, while passing all frequencies below the *passband* [3]. These filters are composed of stream of data items. All samples of

the output stream are in fact a weighted average of the input with the adjacent points of low pass filter. A simple low-pass filter is presented in Figure 2.

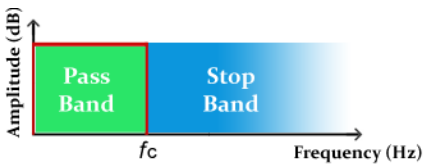


Fig. 2. Frequency response of a simple Low-pass filter.

B. High-Pass Filter

A high-pass filter has opposite characteristics of the low-pass filter. The effect of the filter is to weaken the frequencies below the *cutoff frequency* whereas passing all frequencies above the cutoff frequency.

As in the case of Lowpass filter, the output is generated with a weighted average of the adjacent input stream. The response characteristics of a simple high-pass filter is presented in Figure 3.

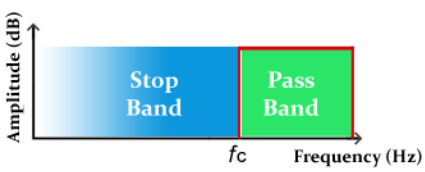


Fig. 3. Frequency response of a simple High-pass filter.

C. Band-Pass Filter

A band-pass filter is a composition of the high-pass and low-pass filters. This type of filter passes certain ranges of frequencies and rejects the frequencies of the remaining region. The frequency response of a simple band-pass filter is shown in Figure 4.

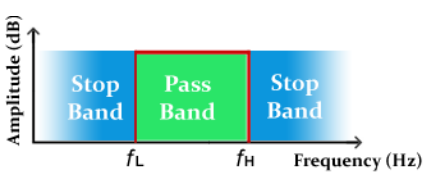


Fig. 4. Frequency response of a simple Band-pass filter.

III. Parallelization for Dataflow Computing

Algorithm 1 presents the sequential version of convolution of a one dimensional input with a kernel. The complexity of the algorithm depends on the input and kernel stream length, i.e $O(nm)$. When run on a CPU, the flow is sequential, meaning that the inner loop length depends on the kernel size. This flow is visualized in Figure 5.

Algorithm 1 Filtering algorithm

```
procedure CONVOLUTION(in, kernel, out)
    i ← 0
    while i < inputSize do
        sum ← 0
        j ← 0
        while j < kernelSize do
            sum ← sum + in[i − j] ∗ kernel[j]
            j ← j + 1
        out[i] ← sum
        i ← i + 1
    return out
```

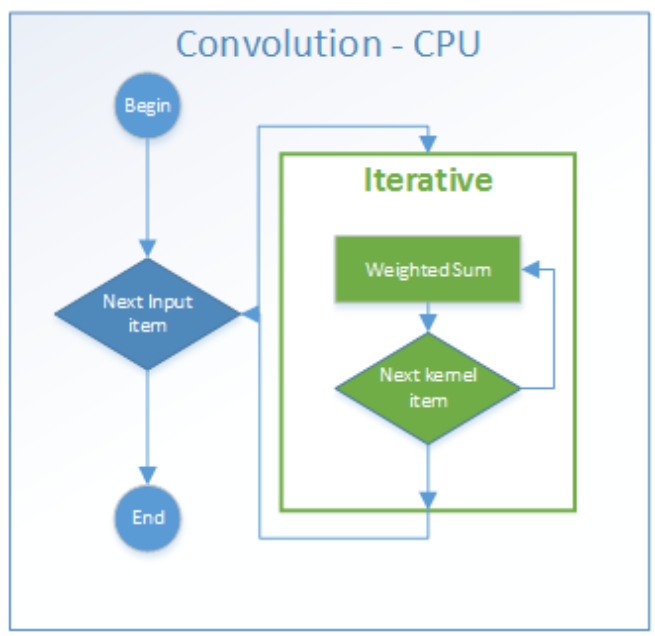


Fig. 5. Flow of sequential filtering algorithm.

In CPU computing, iterations are parallelized across the available time, and performed sequentially.

Dataflow is a completely different computing paradigm compared to the traditional CPUs. Here, the instructions are parallelized across the available space, rather than time. Using this key feature, we have achieved parallelization of kernel computing via Dataflow cores. In this manner, iterative kernel computation is massively parallelized. Depending on the kernel size, up to thousands of dataflow cores can be utilized synchronously, providing a speedup with a higher order of magnitude. The proposed solution is visualized in Figure 6.

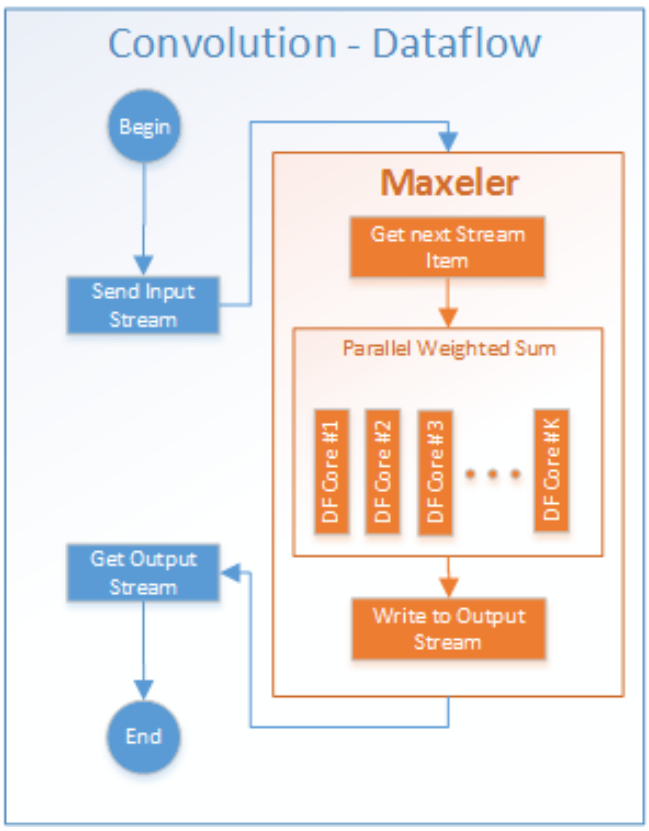


Fig. 6. Parallel Dataflow Computation algorithm.

IV. Tests and Results

The sequential code is tested on an 8-core Intel(R) Xeon(R) X5647, 2.93 Ghz system with 12GB of memory. On the other hand, parallelized code is tested on a Maxeler simulator. Five different kernel sizes are tested, and vice versa, for various length ECG input signals.

A. Functional Verification

To verify the functional characteristics of the execution of the sequential and parallel algorithms we have provided several experiments. The input was a short sequence of 500 samples of an ECG signal with all characteristic P, Q, R, S and T waves, as presented in Figure 7.

Figure 8 presents the effect of applying a low pass filter on the ECG signal. One can notice that the 50Hz noise is eliminated.

The effect of the high pass filter is presented in Figure 9. The effect of this filter is elimination of the baseline drift, caused by breathing and other physical movements.

The effect of the band pass filter as a combination of a low pass and high pass filter is presented in Figure 10. This filter eliminates the baseline drift, caused by breathing and other physical movements and also the 50Hz noise caused by the electricity.

We have verified both the sequential and parallelized solution and obtain identical results.

B. Speed-up Analysis

Speed-up is calculated by (2), where T_s is the time required to process the sequential algorithm, and T_p is the time required to process the parallel algorithm with p cores. Since the system clock on the sequential machine is much higher than the system clock on the parallel machine, we will compare the number of sequential steps N_s (operations required by the sequential algorithm) and the number of processing steps N_p (operations required by the parallel algorithm), by considering the sequential system clock C_s and parallel device's system clock C_p.

$$S_P = \frac{T_s}{T_p} = \frac{N_s}{N_p}\frac{C_p}{C_s} \tag{2}$$

Note that for each input signal, the speedup values are calculated by using T_s and T_P values for the same input configuration. Hence, the main reason for using these values is to compare the performance of the sequential code on CPU and the parallelized code on Maxeler Dataflow Engine (DFE).

Sequential running code has mainly two phases: the initialization and processing phases. Let the input stream contains N elements and the filter kernel M elements. On initialization, the input stream and filter kernel are transfered from the memory to the CPU by a total of $N + M$ memory access

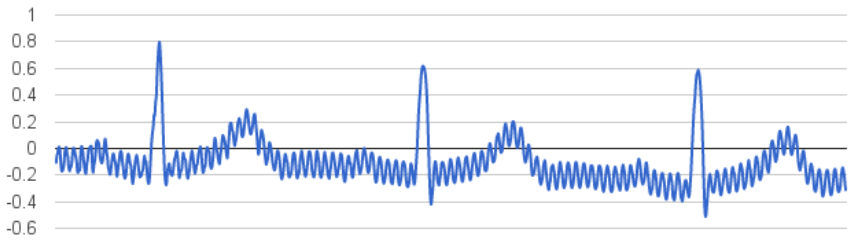


Fig. 7. A segment of an ECG signal with several QRS complexes.

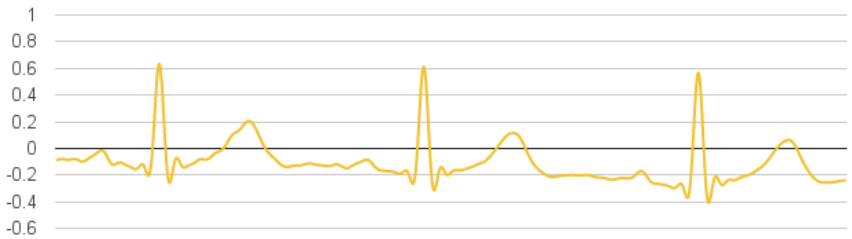


Fig. 8. The ECG signal filtered with a low pass filter of 30Hz.

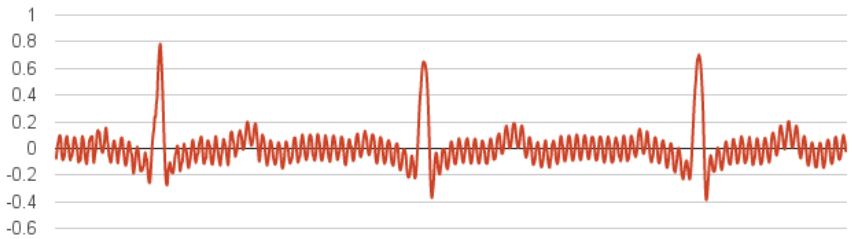


Fig. 9. The ECG signal filtered with a high pass filter of 0.5Hz.

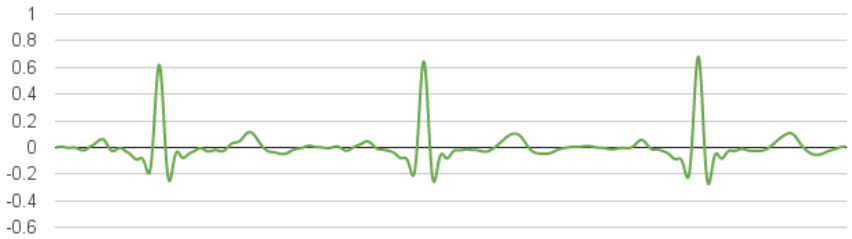


Fig. 10. The ECG signal filtered with a band pass filter between 0.5Hz and 30Hz.

operations and the N output elements are written in the memory. Processing phase requires $N * M$ multiplications and $N * M$ additions. Assuming that each memory access, multiplication and addition requires 1 processing step, the relation that shows the total number of processing steps is presented in (3).

$$N_s = 2NM + 2N + M \tag{3}$$

The number of operations for the parallel algorithm is calculated differently. In addition to processing, the dataflow engine needs to transfer data from memory to device and return the results back, which is equal to a total of $N + N + M$ memory access operations for the input and output stream, and the filter kernel. The dataflow engine performs operations on N samples concurrently in a pipelined manner, so the processing takes N processing steps plus the pipeline length of the number of operations, which is equal to the kernel length M. Note that the summation can be realized in a tree parallel organization, which will take only $log_2 M$ steps, but since we expect that $N >> M$ it will not affect the final result. So the total processing steps is expressed by (4).

$$N_p = 2N + M + N + M = 3N + 2M \tag{4}$$

Table I presents the number of operations and calculated speedup for our experiments where the input contains 100.000 samples and kernel length was 100, 500, 2000, 5000 and 7500. The sequential machine clock was 2.93 GHz and the dataflow Maxeler device system clock 400 Mhz. The speedup increases with the length of the kernel size.

V. Discussion and Conclusions

This work contributes Maxeler Dataflow parallelization for noise filtering of ECG heart signals. The provided parallel solution takes advantage of thousands of Dataflow cores. Their size is increasing linearly (according to the maximum number of available space) depending on the kernel size used.

Results obtained by executing the sequential algorithm and the parallel dataflow algorithm show that the obtained results are identical for low-pass, high-pass and band-pass filters.

TABLE I. Speed-up analysis as the kernel size increase.

No.	Platform	Input Size	Kernel Size	Num. of Opers.	S_P
1	CPU	100000	100	20200100	9.19
	DFE			301000	
2	CPU	100000	500	100200500	45.45
	DFE			301000	
3	CPU	100000	2000	400202000	179.72
	DFE			304000	
4	CPU	100000	5000	1000205000	440.47
	DFE			310000	
5	CPU	100000	7500	1500207500	650.18
	DFE			315000	

The analysis shows that the speedup is proportional to the filter length. In this research we have experimented with the Hamming window and the Blackman window with length of 100 and 200 elements to obtain relatively good results.

This research is the first step of the ECG signal processing with the intention to extract hidden information. As the next work we plan to parallelize the Wavelet Transformation sequential code for Maxeler Dataflow.

As future work, we plan to carry out more tests on Maxeler DFE with higher available space. We believe that the parallelized code will scale linearly with increasing filter size.

References

[1] P.D.Khandait, N. Bawane, and S.S.Limaye, "Article: Features extraction of ecg signal for detection of cardiac arrhythmias," *IJCA Proceedings on National Conference on Innovative Paradigms in Engineering and Technology (NCIPET 2012)*, vol. ncipet, no. 8, pp. 6–10, March 2012, full text available.

[2] A. Antoniou, *Digital signal processing*. McGraw-Hill Toronto, Canada:, 2006.

[3] S. W. Smith, *Digital signal processing: a practical guide for engineers and scientists*. Newnes, 2003.

[4] R. Acharya, S. M. Krishnan, J. A. Spaan, and J. S. Suri, *Advances in cardiac signal processing*. Springer, 2007.

[5] T. S. Lugovaya, "Biometric human identification based on ecg," 2005.

[6] J. A. Sharp, *Data flow computing: theory and practice*. Intellect Books, 1992.

[7] M. J. Flynn, O. Pell, and O. Mencer, "Dataflow supercomputing," in *Field Programmable Logic and Applications (FPL), 2012 22nd International Conference on*. IEEE, 2012, pp. 1–3.

[8] "Maxeler dataflow computing," https://www.maxeler.com/technology/dataflow-computing/, last visited on 26.03.2016.

Intelligent File Tracking System

Aco Latkoski,
School of Computer Science and Information Technology
University American College Skopje
aco.latkoski@gmail.com

Veno Pachovski, PhD
School of Computer Science and Information Technology
University American College Skopje
pachovski@uacs.edu.mk

Adrijan Bozhinovski, PhD
School of Computer Science and Information Technology
University American College Skopje
bozinovski@uacs.edu.mk

Abstract — **In the current period of the digital age, technology is taking an even bigger part in people's everyday lives. Although a large percentage of the information and documentation is stored as electronic (digital) documents and files, there are areas where physical files are still used. This paper presents a system that manages these physical files as well as digital files within an organization. All the components of the system are explained in order to get a better understanding of its architecture. More information is presented for the radio frequency identification technology that the system uses for managing the physical files. The benefits from this system as well as future plans to introduce big data to the system are also presented.**

Keywords — intelligent tracking system, file tracking, data warehouse, RFID tags, RFID tagging

I. Introduction

One of the industries that requires storing physical files and objects is the medical industry. All Health Records must be stored in systems that have controlled access mechanisms that restrict and record accesses to that information. Database based systems must have their own built-in access controls. Paper based records must require the logging of all record retrievals and the clear recording of the name of any person making any entry onto the record [1].

There are many systems today that are designed for storing and managing digital files but the real challenge is the managing and storing of the physical files and objects.

The process of storing physical files poses several challenges, especially if the number of files is quite large. Some of those are: insufficient storage space, managing costs, efficiency, security, speed of delivery for the requested files, finding missing (or misplaced) files etc.

The physical files must be sorted (arranged) in some order in the warehouse. A feature of the record series is chosen as the basis for the arrangement. This feature is the most easily identified by determining how various types of records are requested. Features may include subject, a name associated with the record, a number which identifies the record, and/or a title. It is best to use an existing feature rather than to create an abstract one (like Metadata). The most common arrangements are: Alphabetic, Chronological (Folders are arranged by sequential date order) and Numerical [2].

Because of the continuous adding of new files, there will be time when there is not enough space left on some shelves to store additional files. Because of this, the staff have to store the files to some other shelf and with this the order is broken. Another big problem is finding the active files for the patients that are spread throughout the hospital. The file follows the patient as s/he moves trough the hospital, but it could happen that the file be misplaced or forgotten in some office along the way. Because of problems like this, the tasks connected with managing and storing large number of physical files could be very frustrating and insufficient.

This is where an intelligent file tracking system falls in place. The system is composed of hardware and software components that work together. The hardware component of the system is composed of a Radio-Frequency Identification (RFID) sensor network, hand-held RFID readers and Servers for hosting the software solution. The software component of the system is composed of a web portal that is the main interface from which users interact with the system, Android and iOS apps that run on the hand-held RFID readers, a data warehouse to gather and store all the data and an Analytics module with a reporting tool set (MS SQL/SSRS), enabling standard report template generation and production [3].

II. Real-Time Location Systems

A real-time location system (RTLS) can be considered, metaphorically, as an indoor Global Positioning System (GPS). However, satellite based GPS-es do not work indoors. This is due to a variety reasons, predominantly, because of the buildings themselves – the stone, concrete, wiring and plumbing shield the interior from satellite signals from the outside. Therefore, real time location systems have been developed using a variety of signal modalities: radio-frequency, infrared, and ultrasound. These categories break down further into functional modalities: active, passive or hybrid. Each form has its strengths and weaknesses. Passive Radio-frequency identification (RFID) is perhaps the most widespread form of RTLS [4].

III. Radio-Frequency Identification

Radio-Frequency Identification (RFID) is not a new technology. It was first used over sixty years ago by Great Britain to identify aircraft in World War II and was a part of the refinement of the radar. It was during the 1960s that RFID was first considered as a solution for the commercial world. The first commercial applications involving RFID followed during the 70s and 80s. These commercial applications were

concerned with identifying some assets inside a single location. They were based on proprietary infrastructures.

The third era of RFID started in 1998, when researchers at the Massachusetts Institute of Technology (MIT) Auto-ID Center began to research new ways to track and identify objects as they moved between physical locations. This research, which has a global outlook, centered on radio frequency technology and how information that is held on tags can be effectively scanned and shared with business partners in near real time [5].

An RFID device consists of a small chip and an antenna, as shown on Figure 1.

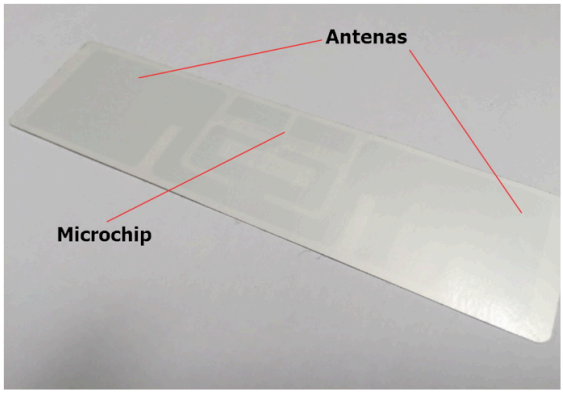


Figure 1: Components of a RFID tag

The RFID device serves the same purpose as a bar code or a magnetic strip on the back of a credit card or ATM card; it provides a unique identifier for that object.

Rather than using light to collect or read a number from a bar code, radio waves are used to read a number from the RFID tag. RFID therefore does not need line-of-sight to operate. Using radio means that the tag no longer has to be visible on the object to which it is attached; the tag can be hidden inside the item or box that is to be identified and still be read. This minimizes or eliminates the need for a person to have to present the reader to the tag, as it can now be fixed to a wall, for example. As the item is passed by the reader, it will be read automatically, thus incurring potentially large savings in labor costs or substantial increases in the throughput of scanned items [5].

Another feature of RFID is the ability to read many tags simultaneously. It is not necessary to present each tag to the reader separately (as is required for barcodes); instead, all tags within the range of the reader can be read almost simultaneously as they pass the reader. Again, there is a huge savings potential in not having to manually present the reader to each item to be identified. Furthermore, data can also be written to the tag - a feature that is not possible with barcodes. This feature has tremendous implications for IT systems [5].

RFID tags are further broken down into two categories:

- Active RFID Tags are battery powered (Figure 2). They broadcast a signal to the reader and can transmit over the greatest distances (100+ meters). Typically, they can cost £5–£20 or more and are used to track high value goods like vehicles and large containers of goods. Shipboard containers are a good example of an active RFID tag application [5].

Figure 2: Active RFID tag

- Passive RFID Tags do not contain a battery. Instead, they draw their power from the radio wave transmitted by the reader, and because of this they are much more compact (Figure 3). The reader transmits a low power radio signal through its antenna to the tag, which in turn receives it through its own antenna to power the integrated circuit (chip). The tag will briefly converse with the reader for verification and the exchange of [1] data. As a result, passive tags can transmit information over shorter distances (typically 3 meters or less) than active tags. They have a smaller memory capacity and are considerably lower in cost (less than £1), making them ideal for tracking lower cost items [5].

Figure 3: Design of passive RFID

On Table 1 the advantages and disadvantages of the passive and active RFID tags are shown [6].

Table 1: Advantages and disadvantages of the passive and active RFID tags

	Active RFID	Passive RFID
Power	Battery operated	No internal power
Required Signal Strength	Low	High
Communication Range	Long range (100m+)	Short range (3m)
Range Data Storage	Large read/write data (128kb)	Small read/write data (128b)
Per Tag Cost	Generally, $15 to $100	Generally, $0.15 to $5.00
Tag Size	Varies depending on application	"Sticker" to credit card size
Fixed Infrastructure Costs	Lower – cheaper interrogators	Higher – fixed readers

IV. The RFID Sensor Network and Hand-Held RFID Readers

In order for the RFID tags to be efficiently tracked, an RFID sensor network needs to be built throughout the hospital. These sensors are typically deployed at "choke points" across the hospital, providing "read points" of item movement through key areas. On Figure 4 an example of a RFID scanner is shown, that is located above one of the main passages in the hospital.

Figure 4: An RFID Scanner located above a hospital door

Hand-held RFID readers are mobile gadgets that allow the users to conduct search operations across the enterprise. An iOS or Android app is usually installed on these mobile readers, which allows the users to perform much more operations like Tag /Find/Audit. Because of these additional capabilities, these RFID devices are very powerful tools used by the operators during their everyday tasks.

A hand-held RFID reader is shown on Figure 5.

Figure 5: A hand-held RFID reader

A hand-held RFID reader is mostly used for locating missing files. Usually when some file is marked as missing, the operator first searches the database to locate the last known location of the file. When the last location is determined, the operator will go to the location and will activate the search functionality of the mobile RFID reader. The reader then starts to transmit signals based on how far it is from the file itself. As the operator starts to get closer to the file, the signal (usually a beeping sound) becomes faster and faster, so the operator can easily find the missing file. Thus, the time to locate the missing files is greatly reduced.

V. The Software Components

The Intelligent File Tracking System is composed of several software components.

The web portal is a web base application that is the main interface for the users of the system. Trough this program (Figure 6), the users can manage the digital versions of the files as well as the metadata about the physical files that are stored.

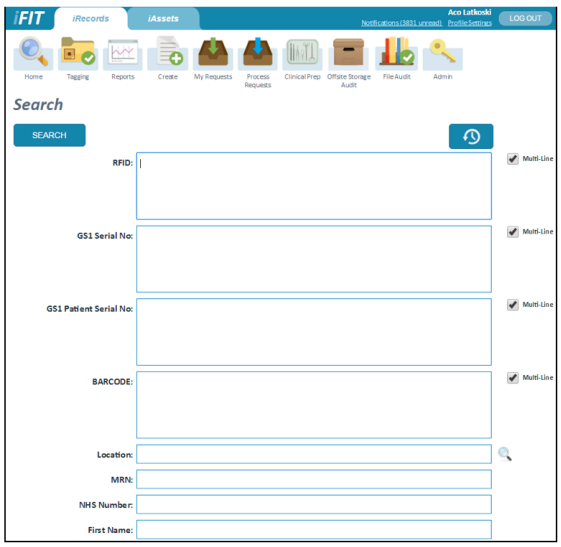


Figure 6: User interface of the web portal

Android and iOS applications are used as main interfaces with the system when the users are using the mobile RFID readers, and do not have access to a computer.

The Database/Data warehouse is the center storage of all data contained in the system. This is very important because the data are not dispersed on multiple places and are thus much easier to manage.

The Analytics module is another very important part of the system because it combines all types of data and metadata to get useful information that the users can benefit from. This module offers different kinds of dashboards with active Key Performance Indicators, or KPIs (Figure 7) and also provides a possibility for ad-hoc reporting as well as standard reports to meet the need of the different users that use the system.

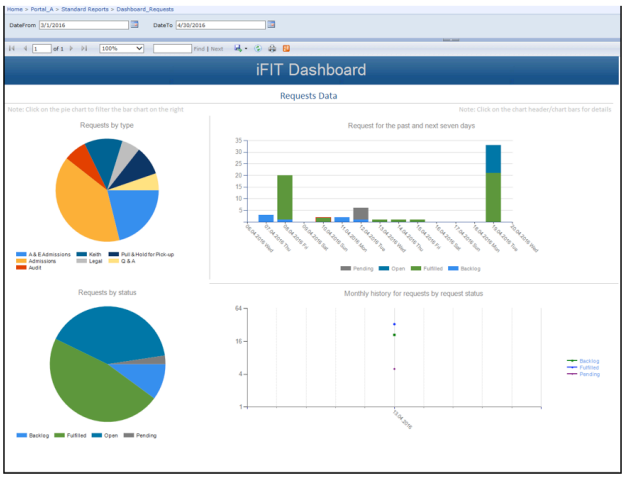


Figure 7: A snapshot of a dashboard from the Analytics Module

For example, it could provide summary data in a dashboard, so the manager of the department can easily obtain information about the things s/he is interested in, without loosing much time. In another instance, some user that works in the warehouse can get a different kind of report that contains all the details about the daily operations in the warehouse itself.

VI. How the System Works

This intelligent files tracking system empowers the health records staff with modern logistics management technology to efficiently track and manage the physical records throughout the organization. This system relies on the use of barcodes and RFID chips to track, search and store the physical files.

RFID passive tags are placed on high-volume items (health records, drugs, supplies, specimens etc.) while RTLS (Real Time Locating Systems) and RFID active tags (utilizing the hospitals Wi-Fi network) are used for low volume tracking, such as patients at risk and key assets that require remote monitoring.

All the files are marked with barcode and passive RFID tags. Because of this, the files do not need to be sorted in some specific way, like for example the alphabetical order. When some file is needed, the operator will check the file on the portal, and s/he will obtain exact info about its location, including department, shelf number, and shelf row. With this, the operator can quickly find and extract the file from the warehouse.

When some file needs to be put back to the warehouse, the operation is even simpler, because the user just needs to find some empty shelf and put the file on the shelf. The file will be stored in the database with the exact new location.

When the file is pulled out of the warehouse and is moved around the hospital, its movement is automatically tracked using the fixed RFID scanners located throughout the hospital. If there is a need for the file to be located, the user can easy locate the file using its last known location from the scanners. When the file needs to be retrieved for some reason, the operators can use the hand-held RFID readers to track the file to the exact place in the room. The RFID reader will scan for the file within the room and will provide information regarding the exact place of the file as shown on Figure 8.

Figure 8: Use of the Hand-Held RFID Reader

VII. Benefits from the System

Using an intelligent file tracking system, the operation of finding missing files is greatly simplified. The organization can have multiple benefits from using this system. Using the data from the existing cases where this system is implemented, the following benefits can be listed [3]:

- 80% labor reduction in check-in files and pre-sorting of the files. Using this system, the files do not need to be sorted in any special order (example: alphabetical) so the arrangement and check-in of the files is much more efficient.
- All the data are stored in a centralized Data warehouse. Because of this, the problems related with data integration, data security and data redundancy are minimized.
- 40% labor reduction in time and effort to pull the files from the storage location (warehouse).
- With the help of a hand-held RFID reader, the user can locate lost files within a radius of 10 meters and 360 degrees around the device. All this is done

without upsetting potential residents in the room. This significantly increases retrieval speed and efficiency of finding missing files and reduces the time for this operation up to 80%.

- 40% labor reduction in archiving of dormant (no longer in use) records.
- The warehouse of files library usually gets an additional 20% extra space because of the new way of storing files that do not require additional place for tags and markings of the files and shelves.
- 40% labor reduction in health records administration i.e.: merging of existing records, handling of duplicate and temporary records
- 25% labor reduction in overall management of the records services department.

VIII. Future Work

A. Introducing Big Data

Big data is being generated by everything around us at all times. Every digital process and social media exchange produces it. Systems, sensors and mobile devices transmit it. Big data is arriving from multiple sources at an alarming velocity, volume and variety [7].

In the current system, all the data from the fixed RFID sensors and the Mobile RFID readers are not gathered and processed. The future plan is to store all this info and use it to gather useful insights.

The plan is that this be implemented using the ELK Stack. The Elastic Stack (ELK Stack) — stands for Elasticsearch, Logstash and Kibana — is consisted of open source projects that help the user to take data from any source, any format and search, analyze, and visualize them it in real time [8].

B. Elasticsearch

Elasticsearch is a distributed, open source search and Analytics engine, designed for horizontal scalability, reliability, and easy management. It combines the speed of search with the power of Analytics via a sophisticated, developer-friendly query language covering structured, unstructured, and time-series data [8].

C. Logstash

Logstash is a flexible, open source data collection, enrichment, and transportation pipeline. With connectors to common infrastructure for easy integration, Logstash is designed to efficiently process a growing list of log, event, and unstructured data sources for distribution into a variety of outputs, including Elasticsearch [8].

D. Kibana

Kibana is an open source data visualization platform that allows the users to interact with their data. Kibana brings components for visual representation of the data that can be combined into custom dashboards that can help to share insights from the data [8].

IX. Conclusion

Today, the technology allows us to simplify and improve many aspects of our daily lives. Many tasks that were complicated and time consuming, can be simplified and improved.

This paper listed the challenges when working with physical files and presented the system for overcoming these challenges. The RFID technology together with the data warehouse are the core components of this file tracking system. All other components are created to support them and together to bring simple and effective solution that the users can use.

Using this system, the hospitals can get numerous benefits due to automated process and the ability to handle electronic, as well as the physical records. Such a system can deliver even greater benefits if implemented in countries like Macedonia, where most of the processes in the hospitals are done manually.

References

[1] Salford Royal NHS Foundation Trust, "Confidentiality Code of Practice," April 2014. [Online]. Available: http://www.srft.nhs.uk/EasysiteWeb/getresource.axd?AssetID=3577&type=full&servicetype=Inline. [Accessed March 2016].

[2] The University of Washington, "Developing or Improving a Filing System," 28 September 2009. [Online]. Available: http://f2.washington.edu/fm/recmgt/filesmanagement/plan. [Accessed March 2016].

[3] 6PM, "Hospital Management - iFIT," 6PM, 2015. [Online]. Available: http://6pmsolutions.com/products/hospital-management/ifit. [Accessed March 2016].

[4] J. E. Stahl, "Real Time Location Systems in the Hospital Environment," May 2012. [Online]. Available: https://healthmanagement.org/c/hospital/issuearticle/real-time-location-systems-in-the-hospital-environment. [Accessed March 2016].

[5] Holloway, Simon, "RFID: An Introduction," Microsoft, June 2006. [Online]. Available: https://msdn.microsoft.com/en-us/library/aa479355.aspx. [Accessed March 2016].

[6] AtlasRFID, "Active vs. Passive RFID," Jovix, [Online]. Available: http://atlasrfid.com/jovix-education/auto-id-basics/. [Accessed March 2016].

[7] IBM, "What is Big Data?," IBM, [Online]. Available: www.http://www.ibm.com/big-data/us/en/. [Accessed March 2016].

[8] Elastic, "The Elastic Stack | Make Sense of Your Data," Elastic, 2015. [Online]. Available: https://www.elastic.co/products. [Accessed March 2016].

Advantages of using the Wizard-of-Oz approach in assistive robotics for autistic children

Ana Tanevska Nevena Ackovska
Faculty of Computer Science and Engineering (FCSE)
Skopje, Macedonia

***Abstract*—Socially Assistive Robotics (SAR)) is one of the fastest-growing fields in the research area focusing on using robots as assistive technology. One target group that benefits greatly from the rehabilitative interaction with robots are the children suffering from Autism Spectrum Disorder (ASD), a condition that impairs normal social and cognitive development. Over the past fifteen years there's been a lot of progress in treating young autistic children by including some form of robotic interaction in their therapy. With the rapid growth of SAR as humanoid robots become more commercially available than ever, two major approaches emerge – the autonomous implementation, where the robot interacts independently with the ASD child; and the Wizard-of-Oz (WOZ) approach, which involves a human supervisor adjusting the robot's responses. This paper elaborates on the two different methods of using humanoid robots in therapy, gives an overview of several such SAR projects in the world and presents the work we did here implementing the WOZ approach in a research experiment with children from the ASD spectrum and the NAO humanoid robot.**

Keywords—socially assistive robotics, autism spectrum disorder, human-robot interaction, NAO humanoid robot, wizard-of-oz approach

I. Introduction. Autism and robotics

Autism represents a whole spectrum of cognitive disorders with similar characteristics and a wide (and occasionally unpredictable) range of symptoms [1]. Individuals with ASD are occasionally prone to some learning difficulties, have impaired or non-existent social skills and communication abilities, and often develop repetitive actions and rituals.

In the majority of cases, a diagnosis for ASD doesn't automatically imply an intellectual disability (ID) diagnosis as well – autistic people still have the capability to learn new skills, engage in hobbies and often have talents for mathematics, music, natural sciences, programming, etc. The learning difficulties these people face stem not from a stunted intellect, but from the lack of a normal socio-emotional development, which impairs their communication and behavior in the classroom and in the workplace and makes it very difficult to acquire new skills [1].

The everyday activity with which ASD individuals (and children in particular) have the greatest amount of difficulty is ordinary interpersonal communication, owing to the fact that autistic people struggle with giving and interpreting social cues and are less able to handle their emotions and react to other people's emotions. Failing to learn social communication in childhood later reflects in the ASD individual's life in more than one way.

The inability to interact with the teacher and the other children in the classroom means that the autistic child will have problems studying and communicating in school, which in turn implies that the same person will labor to lead a well-adjusted life. This goes to show that, while not a life-threatening disease, ASD still has a major detrimental effect on the child's normal development, which is why it's very important to start with therapy as early as possible.

The DSM-5 criteria [2] for early autism diagnosis categorizes the ASD symptoms into two main groups – *deficits in communication and social interaction*, which manifest themselves as difficulties with initiating and maintaining social relationships; and *occurrence of restricted/repetitive actions, interests and behaviors*, which apart from acting as a social restraint can often interfere with the everyday tasks an autistic child has to do.

Following the DSM-5 criteria, the problems that the ASD children face in the learning environment can be sorted in several categories [3]:

- issues with attention span – manifesting as either heightened, intense interest in some specific area of the curriculum, or as inability to focus on the material
- lack of well-developed speech and expression
- problems with abstract thoughts/concepts – most common problematic concepts for young ASD children are colors, numbers, belonging (mine, yours, his/hers)
- problems with social cognition, interaction and imitation – one of the main problems in interaction for ASD children is their missing imitation skill, and the inability to mime other people in facial expressions and actions defer
- difficulties with planning/organizing/task solving

One of the main causes for their communication difficulties is the fact that autistic children absorb a lot more information from a person than non-autistic people do. While interacting with someone, ASD children are presented with a plethora of

social and sensory cues from their conversation partner – facial expressions, voice pitch/inflection, constant gesticulations, etc.

To make the learning process easier for ASD children, the teacher needs to present the child with the least number of information apart from the learning materials, while at the same time strive to keep their pupil's attention.

Roboticists and pediatricians have been teaming up since the late 1990s in order to consider some alternative forms of ASD therapy involving assistive technology. The idea for using robots in rehabilitation stems from the fact that ASD children benefit from using safe, non-threatening, friendly objects while playing or learning [4][5].

The earliest records of robots being used as assistive technology included some forms of mobile robotic toys, which tried to engage the children's interest while attempting to teach it something as well. Notable examples from this era include the AuRoRA study [6] involving robotic platforms and dolls like Labo-1 [6] and Robota [8][9] which focused on involving the children in some low-level imitation games; and Sherbrooke's mobile robotic toys study [10] which used four types of robotic toys, each one engaging the children by some specific way of interaction.

The robot toys marked a great success in those early experiments, as the children felt safe and comfortable in their presence, received a significantly smaller number of social cues from them, and after being in the toys' vicinity for a while, even didn't mind the presence of other people in the room. These experiments gradually evolved over different types of robot toys and platforms, until the eventual implementation of humanoid robots.

II. Implementing Humanoid Robots in Therapy for Children with Autism

Although ASD children still show consistent success when playing/learning with robot toys, the limited number of actions the toys have and the lack of proper learning modules mean that there is a need for more sophisticated platforms. The earlier experiments with robotic dolls proved that autistic children enjoy humanoid robots and find them friendly, so with the release of humanoids for commercial use in the past several years the research area of SAR reached its peak.

When implementing humanoid robots in therapy, there are two main methods of implementation, depending on the way the robots respond to the children and who else participates in the rehab therapy.

A. Wizard-of-Oz implementation

The WOZ approach is the easier one for implementing, as it doesn't rely completely on the robot for processing the input data, but it also involves a human component – a researcher or a medical professional who carefully monitors the child's reactions to the robot and adjusts the responses accordingly. A WOZ implementation can be partial, where the human participation is just occasionally tweaking the robot's responses, or it can be complete (i.e. teleoperation), where the robot is completely controlled (every single action) by a hidden human.

Successful SAR studies based on total teleoperation include the experiments with the Kaspar robot [12], where an imitation turn-taking game takes place between two ASD children, while a third person controls Kaspar's movements via a remote; the experimental therapy with the humanoid robot Troy [13], where two clinicians are involved, one in controlling Troy with a remote while monitoring the child's reactions, and the other trying to initiate interaction between the child and the robot; as well as several studies with the humanoid robot NAO [11][14], where the modules are programmed prior to the classes, but are started via smartphone or a laptop connected to the robot.

While most of the studies above had chosen the WOZ approach for the great amount of flexibility it offers, it's important to note that some research experiments and projects implement in them some aspects of the WOZ method due to the lack of storage space available on robots, such as the SAR study with the Lego NXT robots [15], where the module for conversation with the children is entirely teleoperated and the sound files are sent on cue from a computer connected with the robot.

B. Autonomous implementation

Talking of autonomy in this research area indicates that the robot would be the only one responsible for its actions and responses to the child. This approach implies a much higher level of AI, consisting of advanced modules for processing sensor inputs and translating them into potential for actions. However, the responding actions of the robot are part of a predefined set, created before the beginning of the therapy sessions, which in turns means that the autonomous implementation is very limiting and not adjustable in real-time, so it's the better approach for a fixed set of exercises.

C. Nikolopoulos et al. [15] give a good elaboration on the neccessities when starting with autonomous implementation, such as having a familiarization phase between the robot and the child prior to programming the modules, and using proximity and eye gaze as a factor when settling on a fixed set of exercises. Another factor to consider when opting for this implementation is the increased amount of sensor data that will need to be processed, which can be countered by having a very strict, limited number of modules, such as in the graded cueing study with the NAO robot [16], where an imitation game was made incredibly specific by having a fixed amount of positions the child could do, and by connecting the robot with a Kinect sensor to follow the child's poses so there'd be no need of a human supervisor.

The case in robotics is always made for using an autonomous implementation, as it is the highest form of AI and it always showcases the researchers' skills. However, in SAR the most important aspect isn't the autonomy of the robot and its skillset, but how well the robot responds to the child's needs and how well they're matched in therapy.

It's very important to note that even for a trained human professional with years of medical expertise it can sometimes be very difficult to correctly anticipate the child's reaction to something or to adjust the learning plan/module, let alone to predict it all beforehand and program it. This shows that having someone present who would monitor the robot and adjust the commands is more beneficial to the SAR therapy and brings better results.

III. Used Materials and methods

A. Technology

The robot selected as the most suitable one for this project was Aldebaran's NAO humanoid robot. NAO has already been used in several autism therapy experiments in the world, and with its neutral, friendly face, LED eyes, bright colors and its height matching that of a human toddler it satisfies all of the criteria for autism robots [17].

The modules were all programmed in Choreographe, but in order to implement the commands in Macedonian we used human voice, recorded and then edited to sound more robotic in Audacity. Since our experiment was based on the WOZ approach, we put more focus on the adjustability and flexibility of the exercises, and kept the amount of sensor information we were processing from the NAO to a minimum – in this case the only sensors who occasionally served as input were the tactile sensors on NAO's head.

B. Subjects

The target group for our experiment were young children (between the ages of 7 and 11) on the autistic spectrum, and the selection of candidate subjects was coordinated by the non-profit organization for assistive technology "Open the Windows" (OtW). The final list of subjects consisted of eight of their ASD patients, from which one was female and seven were males. Additionally, four of the subjects were without developed speech and one was diagnosed with cerebral palsy (CP) as well. All of the children involved had been previously visiting the OtW center for rehabilitative therapy involving assistive technology, but none of them had had any previous contact with robots and/or robotic toys/platforms.

The first meeting with the NAO robot went under heightened supervision and in the presence of both the medical professional in charge of the ASD children and their parents, on the off chance that some of the children could be disturbed by the presence of the robot and would be in need of calming down from familiar people.

The frequency of the modifications to the modules depended greatly on the severity of the children's reaction to NAO – a few of them required constant monitoring and modifications, while with some there was just a slight change or removal of a module at the beginning and then the child was left on its own to interact with the robot in the following exercises.

C. Experimental_setting

Since the experiment already meant the children would be faced with something new (which is always challenging for ASD individuals), the setting had to be in a familiar and safe environment, so the exercises were held in the rooms in the center where the children usually had rehabilitative classes.

Touch plays a big role with some autistic children, especially when seeing a new object for the first time, so to minimize the chance of damage to the robot, NAO was seated on a table and the children faced it, while the chairs were removed and the table was cleared from anything else other than the robot. The eight weekly sessions of exercises took place in the usual timeslots for therapy for each of the children, either during the first or the last 10-15 minutes from their normally scheduled class.

We wanted to stay consistent with the abovementioned groups of learning difficulties, so in coordination with the center's staff who were responsible for the children, we developed eight modules in total, combined in four groups depending on the area they targeted. These exercises are elaborated below:

- Communication and social skills interaction exercises:
  - *firstTimeMeet* – the introduction module for the first week, NAO introduces itself, says it's glad to be here, extends a hand for greeting and asks the child to introduce themselves.
  - *helloHandshakeGreeting* – this was the standard module for greeting, performed at the beginning of each class from the second week onward, NAO says hello and shakes hand with the child.
  - *headTouchWave* – the first familiarization exercise with the robot, the children get to hear NAO talk and see how it responds to touch (NAO greets the child and asks them to pet its head, and if they do it, it changes the eyes' color and waves his arm).
  - *goodbyeWave* – this was the module used at the end of each class, NAO says goodbye to the child and thanks them for playing with it.
- Body and belonging awareness exercises:
  - *faceBodyParts* – NAO first demonstrates its own body and face parts, then asks the child to do the same.
- Spatial awareness and orientation exercises:
  - *moveArms* – this module and the next one were usually played after the *faceBodyParts* one, after the child has been familiarized with NAO's body. In this module NAO moves its arms up and down, urging the child to follow, in a gym-like fashion.

- *moveHead* – similarly to the previous module, this one too is presented as a gym exercise, with NAO moving its head left/right and forward/backward and engaging the child to follow.

- Concept of colors exercises
  - *eyesColours* – by changing the LED color in its eyes, NAO teaches the child the different primary colors and then engages the child to select the right card (i.e. the card matching the current eye color).

IV. Results and discussion

Evaluating the results in a SAR project is always a challenging task, as the results are more often of a qualitative instead of quantitative nature, and this is most evident in the SAR projects that focus on ASD children. The factors in evaluating our success were the time each child took to perform the set of exercises, and the amount of modifications of the modules for the duration of the experiment. In the context of this paper, we will discuss the more relevant of the criteria – the amount of modifications per child.

An important side note – the names of the ASD children who participated in our study are protected for ethical reasons, owing to the sensitive nature of our research. As we discuss the results below, each of the children will be referred to by their initials.

Out of the eight children participating, two dropped out of the experiment after the first couple of classes. T.N.'s main issue was the language obstacle, since he didn't speak Macedonian, and as such couldn't understand the commands NAO gave him. Even though we modified the exercises by either having NAO gesticulate more clearly, or by having his mother translate NAO's commands, it still wasn't enough to breach the language obstacle, and after the first two classes he stopped coming.

On the other side, I.A. had no problem with understanding NAO, but he wasn't able to overcome the initial fear he had from the robot. The modifications here were with the four exercises that had the NAO perform pronounced movements, since that had the greatest negative impact on I.A., but that didn't change the effect NAO had on him, so he dropped out of the experiment as well.

The performances of all of the participating subjects are given in Table 1. It shows the comparison of the children's performance in their first class with NAO and their last one. The durations given in the Table 1 are expressed in minutes.

As one could observe in Table 1, the other six children all finished the experiments, and with most of them (five out of six) we also noted a decrease in the time it took for them to finish the set of exercises, as well as lessened fear from the robot and increased curiosity and wish to interact. Here, the modifications made to the exercises were less noticeable, and most of them were done directly on the spot (with one notable exception being A.K, who didn't need any modules changed).

A.A. required most of the modifications on account of him having CP, so the three exercises focusing on body movements were partially altered (more precisely, A.A. had loss of mobility in his left arm, so the exercises which demanded arm movement were removed while the rest were left in).

TABLE I. Comparison of performance

ASD child	duration of one set before modification	# of exercises modified and/or removed out of 8 total	duration of one set after modification
A.S.	5-10	2/8	<5
A.K.	10-15	0/8	5-10
A.A.	>20	3/8	10-15
F.U.	10-15	1/8	10-15
P.I.	10-15	2/8	5-10
I.N.	10-15	2/8	15-20
I.A.	>20	4/8	/
T.N.	15-20	6/8	/

With A.S. and P.I. there were also some minor alterations, but they were simply lowering or increasing the number of repetitions per module respectively, as A.S. executed the exercises successfully on her first try, but with more repetitions got distracted by NAO's robotic voice and ignored the commands in favor of hearing NAO speak over and over again. On the other hand, P.I. was the exact opposite and could only focus on what NAO was telling him to do after several attempts.

Finally, F.U. and I.N. also finished the set of modules successfully with only some modifications required, but the disparity between that and their increasing set durations is owed to their changeable attitude to NAO – F.U. to the end didn't develop a genuine interest for the robot and as such wasn't motivated to play with it and do the modules faster, whereas I.N. was susceptible to very changeable moods – some days he was eager to do the exercises and others he didn't want to go anywhere near NAO.

V. Discussion and conclusion

The whole experimental setup, as well as the results we obtained, show that it is a great challenge to work with ASD children. The mood swings, the lack of real understanding what is interesting for them or not, the inability to comprehend what makes them scared or comfortable can sometimes be a great obstacle in obtaining the desired outcome.

And yet, maybe the desired outcome is not what we should truly strive for. Perhaps getting different outcomes is providing us with the opportunity to discover more suitable ways to improve our performance and more importantly – to learn how to help each ASD child with their individual needs and issues.

Ultimately, as the study showed and as we ourselves discovered – the work with ASD children and robots is rewarding. The majority of our subjects improved their time for exercises completion with each repetition. From the six children that endured the whole duration of the experiment, with five of them we noticed an improvement in the time they took to finish the exercises. This is what marked the whole two-month experiment as a successful one, and what will hopefully motivate other researchers as well to take interest in the still-developing area of socially assistive robotics.

ACKNOWLEDGMENT

The authors would like to thank the non-profit organization for assistive technology “Open the Windows” for helping us in finding the subjects as well as granting us use of their center during the two months of the experiment, and are very grateful to their staff as well as the children and parents who participated in this study.

The authors also extend their thanks to the robotics laboratory of the Faculty for Computer Science and Engineering (FCSE) and to the professor Andrea Kulakov for providing us with unlimited access to the NAO robot for the duration of the research experiment.

The work presented here is partially supported by the Faculty for Computer Science and Engineering.

REFERENCES

[1] M. S. Sweeney, The Complete Mind: How it develops, how it works and how to keep it sharp, Washington DC: The National Geographic Society, 2009

[2] L. Carpenter, “DSM-5 Autism Spectrum Disorder, Guidelines & Criteria Exemplars,” 2013.

[3] Teaching Students with Autism: a Resource Guide for Schools, Special Programs Branch of British Columbia Ministry of Education, 2000.

[4] A. Duquette, F. Michaud, and H. Mercier, “Exploring the use of a mobile robot as an imitation agent with children with low-functioning autism”, in Autonomous Robots, 24(2), pp. 147–157, 2008

[5] B. Robins, K. Dautenhahn, and J. Dubowsky, “Does appearance matter in the interaction of children with autism with a humanoid robot?”, in Interaction Studies, 7(3), pp. 479–512, 2006.

[6] I. Werry and K. Dautenhahn, “Applying Mobile Robot Technology to the Rehabilitation of Autistic Children,” in: Procs SIRS99, 7th Symp on Intelligent Robotic Systems, Coimbra, 1999.

[7] P. Sengers, “Do the Thing Right: An Architecture for Action-Expression,” in: Autonomous Agents 98, Minneapolis, 1998.

[8] A. Billard, K. Dautenhahn and G. Hayes, “Experiments on Human-Robot Communication with Robota, an Imitative Learning and Communication Robot Doll,” in: Workshop "Socially Situated Intelligence", SAB98 conference, Zurich, 1998.

[9] K. Dautenhahn, B. Robins, A. Billard and R. T. Boekhorst, “Robotic assistants in therapy and education of children with autism: can a small humanoid robot help encourage social interaction skills?,” Universal Access in the Information Society, 4(2), pp. 105-120, 2005.

[10] F. Michaud and K. Theberge-Tumel, “Mobile Robotic Toys and Autism,” Socially Intelligent Agents, pp. 125-132, 2002.

[11] S. Shamsuddin, H. Yussof, L. Ismail, S. Mohamed, F. Hanapiah and N. Zahari, “Humanoid Robot NAO Interacting with Autistic Children of Moderately Impaired Intelligence to Augment Communication Skills,” Procedia Engineering, no. 41, pp. 1533-1538, 2012.

[12] H. Admoni, M. Mataric and B. Scassellati, “Robots for Use in Autism Research,” Annual Review of Biomedical Engineering, no. 14, pp. 275-294, 2012.

[13] M. A. Goodrich, M. Colton, B. Brinton and M. Fujiki, "A case for low-dose robotics in autism therapy," Human-Robot Interaction (HRI), 2011 6th ACM/IEEE International Conference on, Lausanne, 2011, pp. 143-144.

[14] D. Conti, S. Di Nuovo, S. Buono, G. Trubia and A. Di Nuovo, "Use of robotics to stimulate imitation in children with Autism Spectrum Disorder: A pilot study in a clinical setting," Robot and Human Interactive Communication (RO-MAN), 2015 24th IEEE International Symposium on, Kobe, 2015, pp. 1-6.

[15] C. Nikolopoulos et al., "Robotic agents used to help teach social skills to children with Autism: The third generation," RO-MAN, 2011 IEEE, Atlanta, GA, 2011, pp. 253-258.

[16] J. Greczek, E. Kaszubski, A. Atrash and M. Matarić, "Graded cueing feedback in robot-mediated imitation practice for children with autism spectrum disorders," Robot and Human Interactive Communication, 2014 RO-MAN: The 23rd IEEE International Symposium on, Edinburgh, 2014, pp. 561-566.

[17] N. Giullian, D. Ricks, A. Atherton, M. Colton, M. Goodrich and B. Brinton, "Detailed requirements for robots in autism therapy", Conference Proceedings-IEEE International Conference on Systems, Man and Cybernetics, pp. 2595-2602

Evaluation of the fuzzy nearest neighbors classifier in the prediction of the binding sites of proteins

Georgina Mirceva, Andreja Naumoski, Andrea Kulakov
Department of intelligent systems
Faculty of computer science and engineering, Ss. Cyril and Methodius University in Skopje
Skopje, Macedonia
georgina.mirceva@finki.ukim.mk, andreja.naumoski@finki.ukim.mk, andrea.kulakov@finki.ukim.mk

***Abstract*—The knowledge about the functions of protein structures is of high importance because it can be used for designing new drugs that could improve human life. With the improvement and development of new technologies, the number of protein structures that are determined significantly rises. But the knowledge about the structure of proteins is not useful if they are not functionally annotated. Therefore, plethora of algorithms have been proposed for functionally annotating protein structures. Some of the algorithms annotates protein structures by using the features of the binding sites of protein structures that are the regions where interactions occur. For that purpose, also there is need for development of algorithms for predicting the binding sites of protein structures. In this paper we use the fuzzy nearest neighbors classifier for prediction of the binding sites of proteins. Detailed experimental evaluation is performed and the experimental results are presented in this paper.**

Keywords—protein function; protein binding site; fuzzy nearest neighbours classifier

I. INTRODUCTION

One of the most important compounds of the cells in living organisms are proteins since they participate and control large number of processes. By using the knowledge gathered about their functions and the interactions where they participate, we can influence these processes in the cells of living organisms. Namely, we can stimulate or deactivate processes. With the rapid improvements in the technology, many methods are proposed that are used for determining the structure of protein molecules. However, it is not useful if the knowledge gathered with these technologies is used for discovering the functions of the protein structures, which is very important for drug design. For this reason, many research groups work on development of fast and accurate computational methods for determining protein functions. In the literature, a plethora of algorithms can be found that could be used for discovering the functions of protein molecules. One group of methods [1] aim to discover protein functions by identifying the homologous proteins of the inspected protein that should be functionally annotated. These methods are based on the fact that proteins that have common predecessor are very likely that would have same functions. Based on the same fact, it is believed that the protein functions are determined based on the characteristics of the stable regions of the protein structures that do not change during evolution of the protein structures. This is used in a second group of methods [2], which are based on identifications and inspection of the conserved parts of protein structure. Third group of methods [3] use protein-protein interaction networks and apply various methods for graph analysis in order to determine protein functions. Forth group of methods [4] aim to identify the binding sites of protein structures and then discover protein functions based on the characteristics of the identified binding sites. In this paper, our focus is on the forth group of methods.

In order to identify the protein binding sites, first we have to extract the characteristics of the residues of the protein molecule. In this paper we extract four features of the amino acid residues: Accessible Surface Area (ASA) [5], depth index (DPX) [6], protrusion index (CX) [7] and hydrophobicity [8]. Next, the protein binding sites are predicted by using some classification method. In the process of classification, for each amino acid residue we determine if the inspected residue is component of a binding site or not. In this step, numerous classification methods could be used. In [9], we have already performed protein binding sites prediction by using various classification methods. During evolution of protein structures, some mutations in their structures are introduced, and we must be sure that those mutations would not have high influence in the decision process, as it is with the classical classification methods. For that purpose, we have already introduced several fuzzy based methods for predicting protein binding sites. First, we applied the fuzzy decision trees (FDTs) [10], and then we also introduced the fuzzy pattern trees (FPTs) [11] for protein binding sites prediction. In [9], additionally we performed selection of the characteristics of the amino acid residues in order to improve the predictions. Besides these fuzzy based classifiers that induce some type of tree, also many other fuzzy based classifiers can be found in the literature. In [12], we introduced the fuzzy nearest neighbors classifier [13] for determining the binding sites of protein structures, and in this paper we make additional experimental analysis in order to evaluate the prediction power of this method.

The rest of this paper work is organized as follows. First, in section 2 we give explanation how the characteristics of the amino acid residues are extracted. Then, in the same section we give detailed description of the fuzzy nearest neighbors classifier. In section 3, we perform evaluation of our method, and then in section 4 we give some final conclusions and also we identify several directions for further improvement.

II. Our Method

Our method has two stages. In the first stage, first the characteristics of the amino acid residues are extracted. Next, in the second stage the fuzzy nearest neighbors classifier [13] is applied in order to determine whether the inspected amino acid residue is part of binding site or not.

A. First Stage: Extraction of the Characteristics of the Amino Acid Residues

In this research we use four characteristics of the amino acid residues, which are: Accessible Surface Area (ASA) [5], depth index (DPX) [6], protrusion index (CX) [7] and hydrophobicity [8].

The Accessible Surface Area (ASA) gives evidence about the surrounding of a given residue. ASA is usually calculated by using the rolling ball algorithm [5], where a probe sphere with radius R = 1.4 Å is rolled over the protein surface, and with the rolling ball algorithm we estimate how much of the area of each atom could be touched with this probe sphere. In this research we predict the binding sites of protein structures, and it is obvious that the residues that a deeply buried in the protein interior could not be part of a binding region, therefore we do not consider those residues. In this way, we significantly reduce the number of samples and also improve the prediction power. In order to provide this filtering, we take into consideration the residues that have no less than 5% of area that couldn't be touched by the probe sphere [14]. Amino acid residues contain several atoms, so after calculating the ASA for each atom, later we aggregate these values by summing them.

Depth index (DPX) is introduced in [6] and it gives evidence how deep a given atom is from the protein surface. The depth index of an atom is equal to the Euclidean distance between that atom and the closest accessible atom (atom that has ASA greater than zero). After calculating the depth indices for all atoms, then we aggregate them by averaging.

Protrusion index (CX) is proposed by the same authors in [7] and it gives information how dense is the volume around each atom. CX is a ratio between the empty (non-occupied) volume and the occupied volume around the examined atom. In order to estimate the occupied volume, first the number of non-hydrogen atoms are counted, and then this number is multiplied with the average volume of an atom. The empty volume is calculated as a difference between the total volume of the inspected sphere with some given radius, and the occupied volume. In this way, the protrusion index for each atom is estimated, and then the protrusion index of a given residue is calculated by averaging the protrusion indices from all atoms that constitute that residue.

Hydrophobicity is the last characteristic that is used in this research. It gives evidence about the hydrophobic properties of the amino acids. Namely, with this characteristic we present how likely is a given amino acid to be closer to the interior of the protein or to its surface. In the literature, there are different scales for hydrophobicity, and in our research we use the scale introduced in [8], which is the most widely used scale.

B. Second Stage: Fuzzy Nearest Neighbors Classifier

In the second stage, we apply the fuzzy nearest neighbors (FNN) classifier proposed in [13] in order to identify which amino acid residues of the examined protein could be part of binding region. As the name says, the fuzzy nearest neighbors classifier is motivated from the k-nearest neighbors classifier (KNN), which is based on the classical set of theory.

The FNN algorithm [13] works as follows. First the k nearest neighbors are identified, thus forming the set of nearest neighbors NN. Then, the test sample t is classified by maximizing the following function

$$\frac{\sum_{x \in NN} S(x,t)C(x)}{\sum_{x \in NN} S(x,t)}, \tag{1}$$

where with $S(x,t)$ we denote the similarity between a given neighbor x and the test sample t. In equation (1), with $C(x)$ we denote the class membership function for the examined class. $S(x,t)$ is calculated by

$$S(x,t) = \frac{1}{D(x,t)^{\frac{2}{m-1}}}. \tag{2}$$

In equation (2), with $D(x,t)$ we denote the distance from the inspected test sample t and its neighbor x, and m determines the extent of this distance. In this paper, m is set to 2, thus from equation (2) $S(x,t)$ is calculated as $1 / D(x,t)^{2}$.

As we mentioned before, in this research we use the gradual class membership function proposed in [14] in order to overcome the disadvantage of the classical KNN classifier. Let C be the set that contains the samples from the inspected class and n_C be the number of samples from that class, while K is the number of nearest neighbors. The gradual class membership function $C(x)$ is defined as

$$C(x) = \begin{cases} 0.51 + 0.49\frac{n_C}{K}, & x \in C \\ 0.49\frac{n_C}{K}, & x \notin C \end{cases}, \tag{3}$$

where K is the number of nearest neighbors.

III. Experimental Results

In this research we use the same dataset that was used in [12]. For each sample we have four descriptive characteristics, which are the ASA, DPX, CX and hydrophobicity that were described in section 2. Besides these descriptive characteristics, we also have one class attribute that shows if the inspected residue is part of binding site, and therefore it obtains values 0 and 1. The descriptive characteristics are extracted based on the information about the protein tertiary structure obtained from the protein pdb files, while the value for the class attribute is obtained from the BIND database [15].

In the evaluation we consider only the most representative chains of protein molecules. Namely, the dataset is formed from the amino acid residues from the protein chains that have less than 20% sequence similarity. For measuring the sequence

similarity between two inspected protein chains, we use the criteria introduced in [16]. We want to obtain a test dataset that is the most representative, so from these chains, those that have less than 10% sequence similarity are used for forming the test dataset, and the other chains are used for forming the training dataset. As we described in the previous section, the amino acid residues that are far from the surface of the protein and that cannot be reached by the probe sphere also cannot be part of a binding region. Therefore, we take into consideration only the surface amino acid residues, which are the residues for which at least 5% of their surface area could be touched by the probe sphere. For the test dataset, we make an additional filtering, by considering only the chains that are part of the B1549 test dataset that we used in [9]. In this way, the test dataset contains 277735 samples, from which only 16.42% are residues that are part of binding site. Similar class distribution is also found for the training samples, since only 13.58% of the residues are part of binding region. The distribution of the class attribute is not uniform, therefore in order to avoid preferring of the dominant class, we balance the training dataset by down sampling it up to 27% of its size without making replacement of the samples.

In this work we use the implementation provided by the authors of the fuzzy nearest neighbors classifier [13]. The parameters are set to the default values, as they are defined in the software. We made experiments by considering different number of nearest neighbors *K* and we use the Euclidean distance for measuring the distance between two samples. We use four evaluation measures that are commonly used for estimating the prediction power of classification models, which are Precision, Recall, F1 measure and AUC-ROC. On Fig. 1, Fig. 2, Fig. 3 and Fig. 4 we present the experimental results obtained for Precision, Recall, F1 measure and AUC-ROC, respectively.

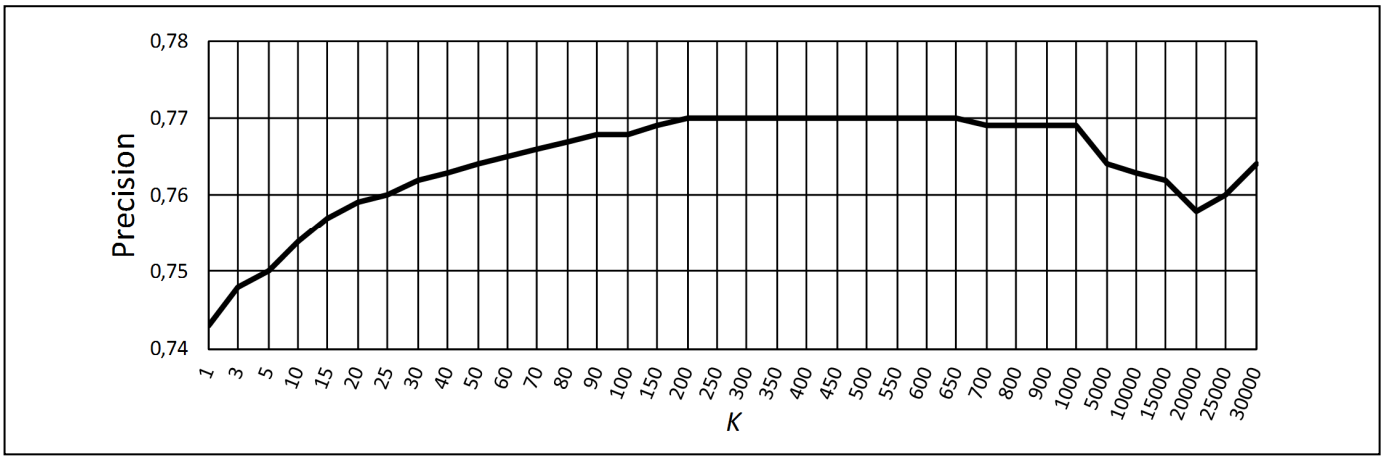


Fig. 1. Experimental results for Precision by using different values for *K*.

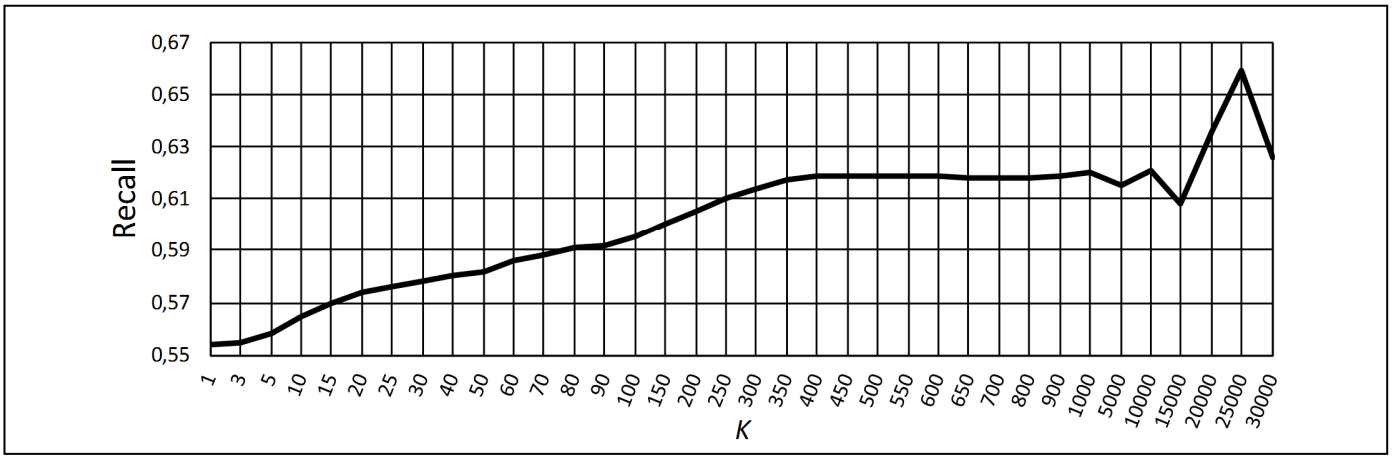


Fig. 2. Experimental results for Recall by using different values for *K*.

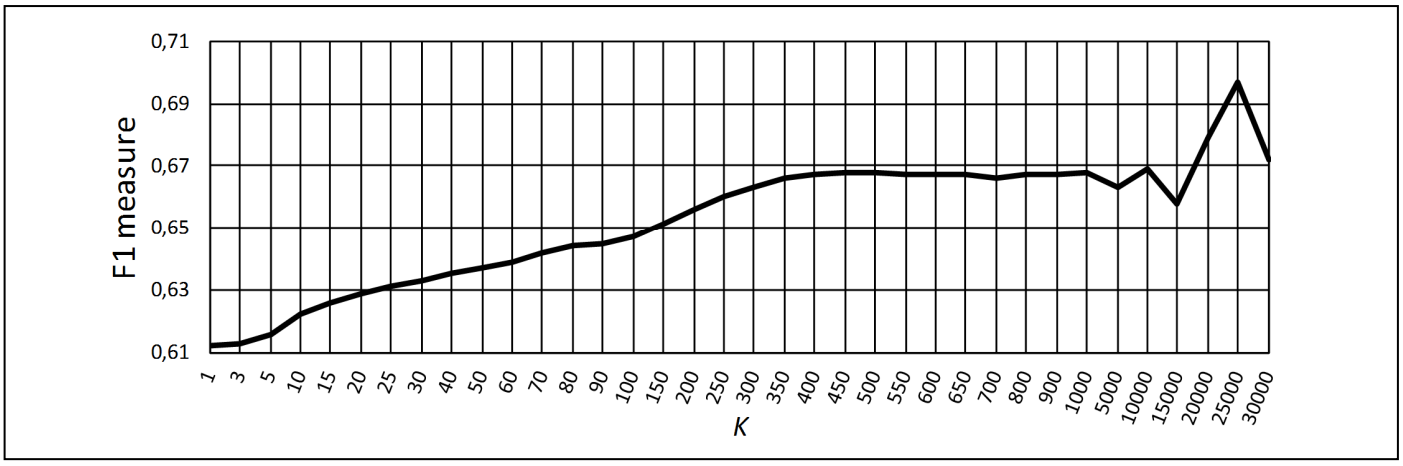


Fig. 3. Experimental results for F1 measure by using different values for *K*.

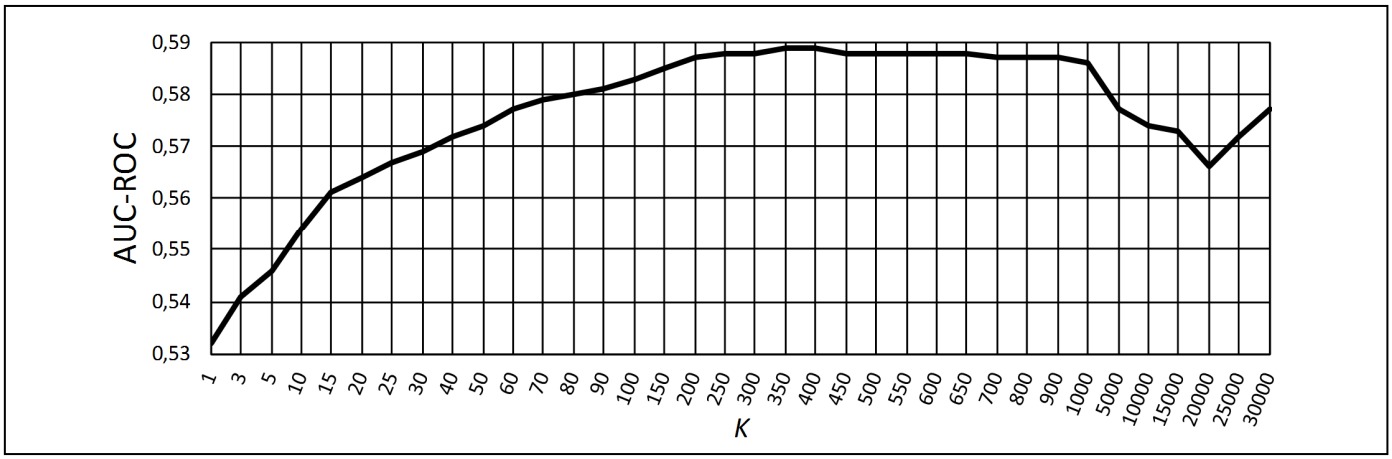


Fig. 4. Experimental results for AUC-ROC by using different values for *K*.

From the results presented on Fig. 1, we can see that as *K* rises from 1 up to 650, the Precision also rises, and then the Precision starts to drop as we increase the number of neighbors. Regarding Recall, from Fig. 2, generally as *K* increases, also the Recall increases. These two evaluation measures are combined in the F1 measure, which is a harmonic mean of the Precision and Recall. From Fig. 3, we can see that generally by using more nearest neighbors, better predictions are obtained. However, since the test dataset is not balanced, the most relevant evaluation measure for estimating the prediction power of the classification model is AUC-ROC. From Fig. 4, it can be seen that by increasing the number of nearest neighbors from 1 up to 400, the AUC-ROC is increased, while with higher values for *K*, AUC-ROC is getting worse. Overall, the experimental results demonstrate that by considering larger number of neighbors more accurate predictions are obtained, but it is best to set up *K* around 400.

IV. Conclusion and Future Work

In this research we presented method for identifying the residues of the protein structures that can be part of binding region where interactions occur. The method has two stages. In the first stage, several common characteristics of the amino acid residues are extracted for getting the descriptive attributes of the dataset. Then, in the second stage, we apply the fuzzy nearest neighbors classifier in order to determine if a given residue corresponds to residue that is compound of a binding region or not.

We made several experiment by using different number of nearest neighbors, and the results showed that it is best to use around 400 nearest neighbors in order to maximize the AUC-ROC evaluation measure. For $K = 400$, our method achieved the following results: Precision = 0.770, Recall = 0.619, F1 measure = 0.667 and AUC-ROC = 0.589.

In this research we used the Euclidean distance in order to measure the distance between two samples, so in future also some other distance measures could be used. Also, the

influence of the parameter *m* could be examined, which defines the degree of the distance between the test sample and its neighbor. In this work we used the ASA, DPX, CX and hydrophobicity as descriptive characteristics for forming the dataset, so further some other relevant characteristics could be incorporated in the dataset that could lead to better predictions. Besides the fuzzy nearest neighbor classifier that is used in this work, also in future we plan to build models by using some other classification methods in order to increase the prediction power of the models.

ACKNOWLEDGMENT

This work was partially financed by the Faculty of Computer Science and Engineering at the "Ss. Cyril and Methodius University in Skopje", Skopje, Macedonia.

REFERENCES

[1] A. E. Todd, C. A. Orengo, and J. M. Thornton, "Evolution of function in protein superfamilies, from a structural perspective," J. Mol. Biol., vol. 307, no. 4, pp. 1113−1143, 2001.

[2] A. R. Panchenko, F. Kondrashov, and S. Bryant, "Prediction of functional sites by analysis of sequence and structure conservation," Protein Science, vol. 13, no. 4, pp. 884−892, 2004.

[3] M. Kirac, G. Ozsoyoglul, and J. Yang, "Annotating proteins by mining protein interaction networks," Bioinformatics, vol. 22, no. 14, pp. e260−e270, 2006.

[4] N. Tuncbag, G. Kar, O. Keskin, A. Gursoy, and R. Nussinov, "A survey of available tools and web servers for analysis of protein-protein interactions and interfaces," Briefings in Bioinformatics, vol. 10, no. 3, pp. 217−232, 2009.

[5] A. Shrake and J. A. Rupley, "Environment and exposure to solvent of protein atoms," Lysozyme and insulin, J. Mol. Biol., vol. 79, no. 2, pp. 351−371, 1973.

[6] A. Pintar, O. Carugo, and S. Pongor, "DPX: for the analysis of the protein core," Bioinformatics, vol. 19, no. 2, pp. 313−314, 2003.

[7] A. Pintar, O. Carugo, and S. Pongor, "CX, an algorithm that identifies protruding atoms in proteins," Bioinformatics, vol. 18, no. 7, pp. 980−984, 2002.

[8] J. Kyte and R. F. Doolittle, "A simple method for displaying the hydropathic character of a protein," J. Mol. Biol., vol. 157, no. 1, pp. 105−132, 1982.

[9] G. Mirceva and A. Kulakov, "Improvement of protein binding sites prediction by selecting amino acid residues' features," J. Struct. Biol., vol. 189, no. 1, pp. 9-19, 2015.

[10] C. Z. Janikow, "Fuzzy decision trees: issues and methods," IEEE Transactions on Systems, Man, and Cybernetics, vol. 28, no. 1, pp. 1−14, 1998.

[11] Z. H. Huang, T. D. Gedeon, and M. Nikravesh, "Pattern trees induction: a new machine learning method," IEEE Transaction on Fuzzy Systems, vol. 16, no. 3, pp. 958−970, 2008.

[12] G. Mirceva and A. Kulakov, "Detecting protein binding sites by using the fuzzy nearest neighbours clasifier," 12th Conference for Informatics and Information Technology (CIIT 2015), Bitola, Macedonia, 2015.

[13] J. M. Keller, M. R. Gray, and J. R. Givens, "A fuzzy k-nearest neighbor algorithm," IEEE Transactions on Systems, Man, and Cybernetics, vol. 15, no. 4, pp. 580–585, 1985.

[14] C. Chothia, "The Nature of the Accessible and Buried Surfaces in Proteins," J. Mol. Biol., vol. 105, no. 1, pp. 1−12, 1976.

[15] G. D. Bader, I. Donaldson, C. Wolting, B. F. Ouellette, T. Pawson, and C. W. Hogue, "BIND: the Biomolecular Interaction Network Database," Nucleic Acids Res., vol. 29, no. 1, pp. 242−245, 2001.

[16] J. –M. Chandonia, G. Hon, N. S. Walker, L. L. Conte, P. Koehl, M. Levitt, and S. E. Brenner, "The ASTRAL Compendium in 2004," Nucleic Acids Res., vol. 32, pp. D189−D192, 2004.

Exploring context for development of context-aware smartphone applications

Suzana Kordumova Nikolova
Faculty of Computer Science and Engineering
"Ss. Cyril and Methodius" University
Skopje, Macedonia
suzana.kordumova@gmail.com

***Abstract* - With the evolution of smartphones and their wide acceptance, the smartphone applications are used in different places, different times, different environments, different culture of living, which in this paper we refer to as context. When humans interact with each other they adapt to the context in which the interaction is occurring. If in a quiet room, one will not speak with a loud voice. While running, one will not read a newspaper. Inspired by this, in this paper we propose to bring context awareness also to smartphone applications, which we refer to as context-aware smartphone applications. So far context has not been exploited to the best of its abilities in application development.**

We contribute the following: first, we define context; second, we better explain and list the relevant contextual information; third, we show ways of acquiring these information, and last, we define functions of how context can be used to improve the user experience with smartphone applications. With this paper we build a strong starting foundation for the development of reach context-aware applications.

Keywords: context, context-aware applications, smartphone applications

I. Introduction

People always use the context of the situation when they get things done. We use our understanding of the circumstances to define activities, to navigate the world around us, to organize information and to adopt to conditions. Using our senses, we can easily learn the context in which we are in. The modern smartphones have a lot of different sets of sensors, and together with internet connectivity, there is a rich possibility for them to also learn the context in which they are used. The smartphone applications that exploit context to adapt to timing, location, to available resources, and other information regarding the users' situation are called context-aware smartphone applications. Context-aware applications, unlike the regular applications, have the advantage of presenting relevant information and services, with respect to users' current context. Users in mobility today can access a very large amount of different information. Thus, the search for the relevant information and services for each specific situation is not always a trivial task and can be time consuming. Furthermore, in a mobile setting (e.g. walking, sightseeing, driving), users are often unable or reluctant to spend large amount of time locating services or information they need at that particular moment and place. Consequently, it is clear that the mobile user experience can be significantly improved by providing the user with relevant and personalized information and services, based on his/hers current context.

When taking into account all the advantages of context-aware smartphone applications mentioned above, we can conclude that the context-aware computing is a promising paradigm. However, if we analyze the applications on the stores for mobile devices, we can see that very rarely the smartphone applications take advantage of the context in which they are used. Most used contextual information is the location, but there are variety of other contextual information, that if used for adapting and personalizing the application can improve significantly the users' experience. Several efforts have been developed in order to support context-aware application development through context and contextual information definitions, software infrastructures, middlewares, etc. However, developing such applications is still an extremely complex task because of the many encounters that need to be overcome when building context-aware applications for smartphones. There are many challenges in defining the users' context, the specific contextual information that should be used, defining the appropriate format in which to present the information to the user and selecting the relevant information to be presented.

In order to use the context in the development od smartphone applications, the notion of context first needs to be well defined. There have been many attempts to define context, but we still think that those definitions do not cover every aspect od the context that is important in the human-computer interaction, so we propose our own context definition. We also define which contextual information are relevant and can be used in order to describe the rich context of the user. The acquiring of contextual information is a challenging task in the development of context-aware application, so additionally we discus how those contextual information can be acquired. At the end of the paper the functions od the context-aware applications are analyzed in more details. There are many ways of how context can be exploited when developing context-aware applications, and our proposals of these features

can help the developers when they build these types of applications.

II. Related work

Context-aware computing is widely accepted as a promising paradigm to implement seamless computing. Several research efforts have been made in order to enable such paradigm, producing a number of context definitions, prototypes, middleware and frameworks.

In Reference [1] Day at al. present the first accurate definition of context, identify categories of contextual information, and characterize context-aware application behavior. One of the most beneficial part of their work is the presentation of the Context Toolkit, that initializes their proposed conceptual framework and supports the rapid development of context-aware applications.

The notion of context and the contextual information have been analyzed by several researchers. Towards the goal of fully understanding what constitutes a context-aware application and what context is, the authors in [2] have surveyed existing work in context-aware computing, they provide an overview of the results form the survey, and define particular definitions and categories of context and context-aware.

The context can be separated in several categories as in [3]. According to their study, the context needs to be considered in terms of the devices relationship with the technical infrastructure, the application domain, the socio-technical system in which it is situated, the location of its use and the physical nature of the device.

Numerous studies as [4], [5], [6] prove that bringing context to recommender systems can improve the recommendations and give more precise personalized information and services to the user.

The challenges that the developers of context-aware applications have are common for all types of applications. Therefor in [7], [8], [9] the authors define general models and frameworks for rapid development of context-aware applications.

III. Context

The use of context, as elaborated earlier, is important in interactive smartphone applications. It is particularly important for applications where the user's context is changing rapidly, such as in most of the smartphone applications usage. In order to better understand how we can use context and facilitate the building of context-aware applications, we need to more fully understand what constitutes a context-aware application and what context is.

A. Previous definitions of context

There have been several attempts to define context. First definition of context was given by Schilit and Theimer [9]:

Context is the location, identities of people, nearby objects and the changes of those objects.

Similar definition is given by Brown at al. [10], where the context is defined with listing of information:

Context is the location, identities of people near the user, time of the day, season, temperature etc.

In [11] Dey in the definition of context lists these contextual information:

Context is the emotional state of the user, his focus, location and orientation, date and time, the objects and the people near the user.

This type of definitions, when the context is defined with listing of contextual information, are inadequate for general usage, and are also not flexible and extendable. If we want to find out if some information can be used as part of the context, if that information is not listed in the definition, it is difficult to state.

Some researchers define context with synonyms, that context is the surrounding environment, the situation in which the interaction is occurring, but these definitions are just synonyms to the word context, and are not precise enough. Sometimes the context is defined as the surrounding of the user, and sometimes as the surrounding of the application. In [12] Brown defines the context as:

The context includes the information for the situation in which the user is in, for which the computers are aware of.

This definition is not correct because the computers, and even the smartphones with all their sensors can still not be aware of the whole context around them.

From the definitions above we can conclude that context is a very general word with very liable definition. Closest to the correct definition of context is the definition given by Dey in [1]:

Context: any information that can be used to characterize the situation of entities (i.e., whether a person, place, or object) that are considered relevant to the interaction between a user and an application, including the user and the application themselves. Context is typically the location, identity, and state of people, groups, and computational and physical objects.

This definition encompasses the definitions given by the previous authors and it is quite general. However, this definition is not precise enough and the term "any information" may include some information that are not relevant for context-aware applications.

B. Our definition of context

The most important aspect when creating a definition for the notion of context, is for that definition to cover all relevant contextual information, even if they still can not be detected with the current devices. In the same time the definition should not cover the environment information that are not relevant for the interaction of the user with the application. The definition we think includes these two conditions is:

Context is a set of all information that describe the environment of the user and the device, in which the interaction is occurring, relevant to that interaction, including

the information that characterizes the user and the device itself.

In the next part we will list the relevant contextual information, and we can see that the definition covers all of them.

IV. Contextual information

Given the diversity of contextual information, it is useful to attempt to categorize them, to make it easier to comprehend in a systematic manner.

A. Classification of contextual information

1) Previous classifications

There have been several attempts to categorize contextual information. In [1] the authors introduce a simple classification of contextual information based in the entities whose context is assessed. They introduce four essential categories, or characteristics of context information:

- Identity
- Location
- Status (or activity)
- Time

On figure 1 the categories and their usage by a context-aware application is shown.

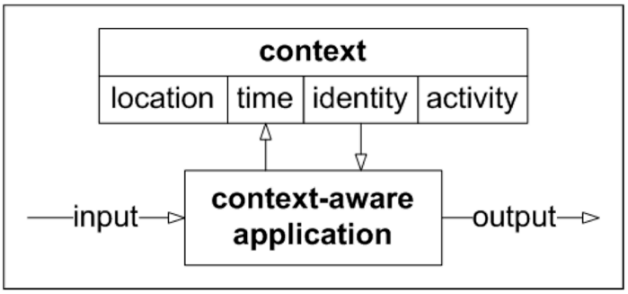


Fig. 1 Context categorization

Very similar classification is proposed by Ryan [14], where instead of the category "Status" the category "Environment" is defined. The authors claim that these contextual information can be easily acquired, and when combining some of them, automatic application actions can be triggered.

Context can be also categorized from a different perspective [13]. Schilit at al. point out three important aspects of context:

- Where are you?
- Who are you with?
- What resources are nearby?

They indicate that context is much more then just a location, because other information of interest are also mobile and changing. According to the authors this classification covers the important contextual information.

Another distinctive context classification is proposed in [15]:

- Infrastructural context
- Application context
- System context
- Location context
- Physical context

This classification takes in account that the mobile applications do not function on their own but are part of a larger system. So the system and the infrastructure of the system are also part of the context.

2) Our classification

The previous classifications of contextual information are a great starting point for thinking about the possible categories of these information. An important aspect for defining the classification, like it was for defining context, is that this classification should cover all possible contextual information. We base our classification on the definition of context, which we already mentioned covers all relevant contextual information. We propose three main categories of context:

- User
- The environment in which the interaction is occurring
- The device that runs the application

This classification is defined semantically, based of where does one contextual information belong. With defining categories of context, we propose layered structure of contextual information, which allows easier stating, usage and discovery of information.

B. List of relevant contextual information

A relevant contextual information is an information that can bring a value when developing a context-aware application, in a manner that based on this information the application can be adapted and personalized. Not all contextual information can be used in all smartphone application. Based on the domain of the application and the services it offers, different contextual information are relevant for different applications. In table 1 all possible contextual information are listed, relevant for different types of applications.

TABLE I. List of contextual information

Main category	***Subcategory***	***Contextual information***
User	Demographic caracteristics	Birthdate
		Age
		Sex
		Birth place
		Place of living
		Language
		Marital status
	User interests	Hobbies
		Music type
		Movie genres
		Friends list
		Favorite food
		Etc.
	Activity	Static
		Walking
		Running
		Automotive movement
		Moving direction
		Moving speed
	Condition	Heart rate
		Weight
		Fat index
		Height
		Inhalers usage
		Etc.
Interaction environment	Time	Time of the day
		Day of the week

Main category	***Subcategory***	***Contextual information***
		Date
		Season
	/	Location
		Noise level
Interaction environment	/	Level of lightning
		Social status (walk, party, meeting, etc.)
		Meteorological state
		Traffic
		Nearby people
		Nearby locations
Device	/	Input sources (keyboard, on touch, etc.)
		Processor
		Operating system
		Battery status
		Connectivity (internet access)
		Bandwidth
		Available sensors
		Communication technologies (NFC, Bluetooth , Wi-Fi, Cellular data, etc.)
		Device orientation

C. Acquiring contextual information

One of the greater challenges for the context-aware applications is to learn the context in which the interaction with the user is occurring. Context awareness represents generalized model of input including both explicit and implicit input, allowing almost any application to be considered more or less context aware, as it reacts to input. In this part we concentrate on the gathering and use of implicitly sensed input by the application. In an ideal setting context would be obtained automatically and there would be no need of manual acquisition. However, some contextual information cannot be sensed automatically and applications must rely on the user to manually provide them. Example for such information are the user condition information defined in table 1. Easing circumstance is the Health application in iOS devices, and the Google Fit application on Android devices, where the users enter their data only once on the device, and then all applications can use the same data. This way the extra interaction required by the user is lowered.

The implicit acquiring of information is possible through the different sets of sensors that the smartphones have in them. We can generally divide the sensors into two categories:

- Software sensors
- Hardware sensors

As software sensors are considered the other applications in the smartphone that allow access to their data. Example for a software sensor is the calendar application. By allowing access to its' records, valuable contextual information can be gathered for the current social situation of the user. Another example are the social network applications, from which important information can be gathered regarding the users' demographics, list of friends, interests, etc.

The hardware sensors that are build in the smartphones can implicitly gather contextual information that the applications may use. For example, the GPS sensor available in almost all smartphone devices, can very precisely determine the location of the user. The raw data that the sensors provide is not always semantically valuable information that the applications can immediately use. Even the location that the GPS sensor returns is just latitude and longitude numbers. If the application needs semantic information like the city and the state the user is in, further processing needs to be done. For some of the most used basic information like the location and the motion of the user, the native SDKs for developing smartphone applications already have APIs for returning valuable contextual information. For other raw sensor data additional processing needs to be done, which can be executed either on the device, or if more complicated, on a remote server.

V. CONTEXT-AWARE SMARTPHONE APPLICATIONS

In addition to being able to obtain contextual information, applications need to have some "intelligent" component that will analyze and use these information. These types of applications that use the context for adapting to it, or for filtered representation of services and/or information in respect to the users' context, are called context-aware applications. In [2] Dey at al. give more precise definition of what context-aware applications are:

A system is context-aware if it uses context to provide relevant information and/or services to the user, where relevancy depends on the user' s task.

They have chosen a more general definition, that requires a response to context, allowing the detection and interpretation of context to be performed by other computing entities.

A. Functions of context-aware application

There are several ways how a context-aware application can use the context. Smartphones have changed the way of functioning and organizing of millions of people. The revolution of smartphones has increased the possibility of communication and interaction, but in the same time brought some negative side effects, like information overload and constant interruptions. Using the context in the smartphone applications can help the users to overcome these issues. We propose two main functions that the context-aware applications can implement:

- Adapting the application to the context in which the interaction is occurring
- Show filtered relevant services and/or information in respect to the current context

The first function type refers to changing the application behavior or its' user interface in different context situations. For example, the font of the letters can increase if the user is moving while reading the information presented by the application. In a low light environment, the colors of the user interface can change so that the application is more easily usable. Another adaptable example is an application in which the sounds are muted if the user is in a meeting. All these examples and many more prove that the user experience can significantly improve if the application is adapting to the current context in which it is used.

The second type of functions that the context-aware applications can implement are also very beneficial for the users. For example, when a user is at work, to show only important messages, and leave the less important once for later. This way the constant interruptions that users have from their smartphones can decrease. Showing a reminders form a shopping list when the user is near the store is also smart context-aware filtering of information. Another example is the showing of the weather report for the current user location. In these types of functions we can also mention the context-aware recommender systems that use the context to recommend information or services to the user using his/hers current context.

VI. Conclusion

In handheld computing like with smartphones usage, a user's context is constantly changing. When using applications in these different environments, a user has much to benefit by the efficient use of implicitly acquired context. It allows an application's behavior to be optimized to the user's current setting. To encourage a more efficient use of context, we have presented definitions and categories of context and context-aware computing. These definitions and categorizations will help developers understand the limitations of context-aware computing, and help application creators select context to use, organize context in their applications, and decide what context-aware functions to implement. We established the basics of understanding context, for more efficient development of context-aware applications.

References

[1] A conceptual framework and Toolkit for supporting the rapid prototyping of Context-Aware Applications – Admin K. Dey, Gregory D. Abowd, Daniel Salber – Human-Computer Interaction, 2001, Volume 16, pp. 97-166

[2] Towards a Better Understanding of Context and Context-Awareness - Anind K. Dey and Gregory D. Abowd - HUC '99 Proceedings of the 1st international symposium on Handheld and Ubiquitous Computing, Pages 304-307

[3] [10] Exploiting Context in HCI Design for Mobile Systems - Tom Rodden, Keith Chervest, Nigel Davies - in Workshop on Human Computer Interaction with Mobile Devices, 1998

[4] Introducing Context into Recommender Systems - Wolfgang Woerndl, Johann Schlichter

[5] Using a Rich Context Model for a News Recommender System for Mobile Users - Alisa Sotsenko, Marc Jansen, Marcelo Milrad

[6] [7] Introducing Context into Recommender Systems - Wolfgang Woerndl, Johann Schlichter

[7] A model for proactivity in mobile, context-aware recommender systems - Wolfgang Woerndl, Johannes Huebner, Roland Bader, Daniel Gallego-Vico - RecSys '11 Proceedings of the fifth ACM conference on Recommender systems, Pages 273-276

[8] A Framework for Developing Mobile, Context-aware Applications - Gregory Biegel, Vinny Cahill - PERCOM '04 Proceedings of the Second IEEE International Conference on Pervasive Computing and Communications (PerCom'04), Page 361

[9] Disseminating Active Map Information to Mobile Hosts - Bill N. Schilit and Marvin M. Theimer. IEEE Network, 8(5), pages 22-32, September/October 1994, IEEE Computer Society.

[10] Brown, P.J., Bovey, J.D. Chen, X. Context-Aware Applications: From the Laboratory to the Marketplace. IEEE Personal Communications, 4(5) (1997) 58-64

[11] Dey, A.K. Context-Aware Computing: The CyberDesk Project. AAAI 1998 Spring Symposium on Intelligent Environments, Technical Report SS-98-02 (1998) 51-54

[12] Brown, P.J. The Stick-e Document: a Framework for Creating Context-Aware Applications. Electronic Publishing '96 (1996) 259-272

[13] Bill N. Schilit, Norman Adams, and Roy Want - Context-Aware Computing Applications

[14] Ryan, N., Pascoe, J., Morse, D. *Enhanced Reality Fieldwork: the Context- Aware Archaeological Assistant.* Gaffney,V., van Leusen, M., Exxon, S. (eds.) Computer Applications in Archaeology

[15] *Exploiting Context in HCI Design for Mobile Systems* - Tom Rodden, Keith Chervest, Nigel Davies - in Workshop on Human Computer Interaction with Mobile Devices, 1998

Competition – Based Learning Method in Teaching Algorithms and Complexity

Marija Mihova, Bojan Ilijoski
Faculty of Computer Science and Engineering
Ss. Cyril and Methodius University
Rugjer Boshkovikj 16, PO Box 393, 1000 Skopje, Macedonia
{marija.mihova, bojan.ilijoski}@finki.ukim.mk

***Abstract*— Students talented in algoithms must be identified by a range of measures that go beyond usual activities and traditional standardized tests. Such students prefer challenging problems that examine their intelegence and deep understanding of the material rather than easy standard problems that they are learning through the lessons. The course Algorithms and complexity at the Faculty of Computer Sciencies and Egineering in Skopje is one of the most popular courses among the students that are very good in programming and students who love to take part in programming competitions. Most of them are already familiar with a number of algorithms before they take the course. Therefore, in order to make the course more interesting for them and to increase their motivation, we've decided to do an evaluation of their knowledge by competitions.**

Keywords - Algorithms and complexity; talented and gifted students, competitions, challenging problems; additional assessment.

I. INTRODUCTION

Traditional standardized tests are not very appropriate for the evaluation of the knowledge of the students talented in solving algorithmic and mathematical problems since such students have lower stimulus for solving easy and standard problems. Sometimes, due to this reason, they achieve worse results than the other students. To inspire them to increase their knowledge, within the undergraduate course named "Algorithms and Complexity", we have introduced a *competition – based* learning method. The objective of this method we introduce in this paper is to stimulate and cultivate interest in solving hard problems, to increase the level of algorithmic understanding of the students, as well as to evaluate their problem solving and logical reasoning skills. Algorithm and Complexity course is the most appropriate for evaluating such characteristic of the students of informatics, since the topics covered in this course are most frequently topics of the tasks at the International Olympiads in Informatics, [1].

In this paper we will present our motivation to introduce this method, we will explain the concept and analyze the impact of this approach.

II. MOTIVATION

Informatics (Computer Science and Engineering) is one of the most popular areas to study in Macedonia and the Faculty Computer Science and Engineering (FCSE) in Skopje has the primacy among all of the other such faculties, [3]. Relatively high percent of the Macedonian students that are gifted for solving algorithmic and mathematical problems choose FCSE for their professional development. In fact, most of the members in the Macedonian teams for International Olympiads in Informatics and Mathematics are students at FCSE, and we can say for sure that if some Macedonian participant at the International Olympiad in Informatics do not continue their education at a high-ranked foreign university, they enroll at FCSE. FCSE is also popular among other competitors from national state competitions and Olympiads in mathematics and informatics. Almost all of these students enroll in the course Algorithms and Complexity which is elective course for the most popular program "Computer Science and Engineering" and obligatory course for the program "Academic Informatics Studies". Since the course is obligatory for one of the programs, there are students in the course with limited mathematical and algorithmic knowledge. This disproportion between students with limited knowledge in algorithm design and mathematics and those that are gifted for that, makes teaching and especially evaluation of the students in this course very difficult. Therefore the professors at FCSE always look for innovative solutions which will motivate extraordinary students[2].

Having in mind that extraordinary students deserve to be rewarded, and inspired to further develop their skills, we have decided to offer them some special and unusual way of examination – examination through problems that are typical for the competitions in algorithms, i.e. problems that are difficult for the average students.

The concept of this approach is to make some of the regular classes, classes for competitions. The students who have demonstrated good performance in these classes will receive a grade based on their performance, and will avoid the stress of the standard exam.

III. ALGORITHM AND COMPLEXITY CURICULUM

The areas that are covered with the Algorithm and Complexity course are the following: Dynamic Programming, Greedy Algorithms, Divide and Conquer Algorithms, Topological Sorting, Breath First Search, Depth First Search, Advanced search techniques, Advanced algorithms for connected undirected graphs, Advanced algorithms for connected directed graphs, Shortest paths in the graph.

Network flows, Computational Geometry, Iterative algorithms. Backtracking algorithms etc.

Algorithm and Complexity has 2 hours of lectures, 2 hours of theoretical exercises and 2 hours of laboratory exercises per week.

The standard testing includes two partial exams or one full exam which contains practical and theoretical part. The exams consist of three programming tasks and theoretical questions about those tasks. Students should solve two tasks and submit them to the online system for the practical part of the exam and answer all the questions on paper for the theoretical part. Most often the tasks are graded with 40, 50 and 50 for the practical part, so the students should pick two tasks for solving. Even if they solve more than two tasks, the maximum number of points they can get for this part is 100. Also the maximum number of points for the theoretical part is 100. Students can pass the exam if they have at least 40 points on the practical and 40 on the theoretical part and at least 100 points in total. The 5 passing grades (6 to 10) are distributed by the following principle: 6 more than 50 and less than 60 points, 7 between 60 and (less than) 70 points, 8 between 70 and (less then) 80 points, 9 between 80 and less then 90 points and 10 for 90 and more points.

IV. Gifted and Talented Students

The most current definition of gifted individuals comes from the Jacob K. Javits [5]. He defines gifted students as those "who give evidence of high performance capability in areas such as intellectual, creative, artistic, or leadership capacity, or in specific academic fields, and who require services or activities not ordinarily provided by the school in order to fully develop such capabilities". On the other hand Calvin Taylor divides these students into two groups: gifted students – those at the very top of any talent area, and talented students - those who are above average. He has suggested that nearly all students are talented in at least one way, but not gifted [6]. According to this interpretation we think that most of the students that enroll in the course Algorithm and Complexity, especially those that choose this subject as an elective one, are gifted students.

A lot of studies about gifted students show that these students cannot be regarded as a homogenous group since they can have multiple abilities [7, 8], and they vary in their emotional, social and physical development [9]. Therefore most educators and researchers believe that special programs are necessary for the gifted because of their unique learning needs [10, 11]. Gifted programs usually offer a challenging and fast-paced curriculum, which is quite different from what gifted students typically experience in the regular classroom.

In their research Osborn and Parnes define the Creative Problem Solving process [12], especially for talented students. The six stages of the model provide a structured procedure for identifying challenges, generating ideas and implementing innovative solutions. The process flows logically through the six steps of:

1. Objective (Mess) Finding - identifying the goal, challenge and future direction.
2. Fact Finding - collecting data about the problem, observing the problem as objectively as possible.
3. Problem Solving - examining the various parts of the problem to isolate the major part, stating the problem in an open-ended way.
4. Idea Finding - generating as many ideas as possible regarding the problem, brainstorming.
5. Solution Finding - choosing the solution that would be most appropriate, developing and selecting criteria to evaluate the alternative solutions.
6. Acceptance Finding - creating a plan of action.

Our approach of giving complex problems in competitions allows the students to develop all the skills defined by Osborn and Parnes.

V. The Concept of the Competition – Based Learning Method

In order to encourage students to build new, complex algorithmic concepts, in the competition classes they receive relatively difficult problems connected to the material. In this section we will give a detailed explanation about the problems; their difficulty and the connection with the studied topics.

As we have previously mentioned the students are evaluated and graded based on their results on the competitions. Besides the problem solving they should explain the complexity of the algorithm that they used for solving some task and explain in detail the solution. The students who demonstrate the best performance are graded with the highest grade.

The level of difficulty of the most of the problems is the same as the level of the problems on national and international competitions for high school students in mathematics and informatics. Some of the tasks are on the same level as the tasks from International Olympiad in Informatics. The students write a program in a programming language of their choice (out of Pascal, C, C++ and Java). These programs are evaluated using the official Macedonian system for competitions in informatics, MENDO, [13, 14].

As for solutions, we expect correct, comprehensive and clear proof of the given problem, as it is usual for the informatics Olympiads, ACM, TopCoder, CodeForces, CodeChef, etc. For the programs that are evaluated by MENDO we make different test cases to assess whether a student has considered all cases, and to measure the complexity of the solution's algorithm. This means that if the student has a more complex which means more time and/or memory consuming solution, that solution will give the right output only in some test cases (not in all) and the student will get only part of the points from that task (depending on how many test cases they have a right answer for). So, for solving the tasks correctly and to show their best performance on the competitions the students need to understand the problem first. This is not an easy task for most of the students and there are problems understandable only for the gifted ones. Then, the students need to learn the material necessary for solving the task, to investigate some solutions of similar problems and to identify the connection with their problem. In the next stage the students must examine the problems in order to identify all subtasks that need to be solved and after that, they must find ideas for the solution and to find

the most appropriate proof techniques to explain it. As a last step, the students need to write their proof, as clear as possible.

When solving the problems the students can employ and further develop their higher-order thinking skills. Evaluate few aspects of mathematical, algorithmic and logical thinking: do the students know algorithm for solving system of linear equations and do they know which algorithm has a better complexity; do the students know the conditions when a system of linear equations has no solution and when there are infinitely many solutions; will the students realize that all the prizes are positive real numbers; and which of them will find an algorithm to test whether a system with infinite number of solutions has at least one solution with certain constraints (positive prices).

VI. Results and Analysis

Here, we will analyze the results of this approach for the past five years of teaching this course. We have had more than 250 course students in those five years.

The presentation of the results includes several aspects:

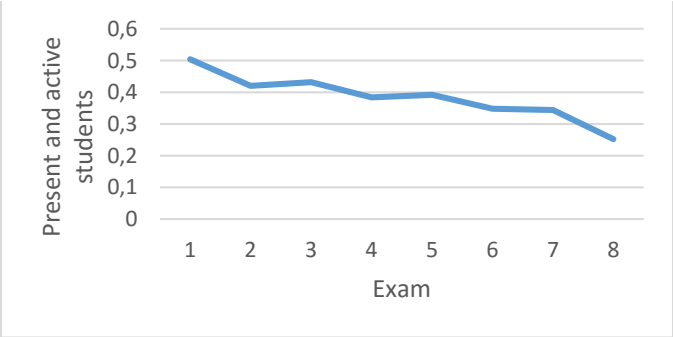
- The level of success of the students;
- The profile of the students that are successful in solving such problems;
- The effect of the challenges on the grade.

In the next table we present the average points and the standard deviation of the points of tasks grouped by areas covered in this course.

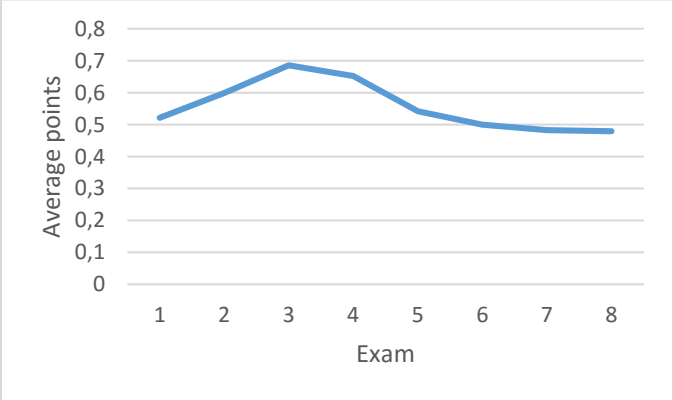
	AVG	SD	Average submitted task (>0 points) by competition
ad hoc	0.45	0.32	32.1
dynamic programming	0.27	0.30	24.3
greedy	0.47	0.33	35.1
bfs	0.44	0.40	21.8
dfs	0.43	0.34	21.4
mst	0.53	0.37	13.5
min path single source	0.44	0.34	13.2
min path all sources	0.48	0.27	10.6
geometry	0.33	0.40	9.6

From this table we can see which part of the material is the most difficult for the students. Also from the number of solved tasks we can see which one is most challenging for the students.

The next chart presents present and active students on the competitions (exams). As we can see the rate is almost constant with a small decrease in the second and the last competition.



The following chart shows the average points earned by the students through competitions (exams). The average points are between 0.47 and 0.69.



In the competitions almost 25% of all and 50% of active students passed the exam. The students who pass the subject through competitions often have the highest grades.

VII. Conclusion

In this paper we propose a new concept of teaching Algorithm design trough competitions and describ our experience in the organization of the course Algorithm and complexity in this way.

Acknowledgment

The research presented in this paper is partly supported by the Faculty of Computer Science and Engineering, at Ss. Cyril and Methodius University in Skopje.

References

[1] http://www.ioi2014.org

[2] M. Mihova, M. Jovanov and E. Stankov On The Role of Challenging Math Problems in the Discrete Mathematics Courses, Global Engineering Education Conference (EDUCON), 2015 IEEE

[3] The portal for the new students at FCSE, http://join.finki.ukim.mk/Dokumenti.aspx?id=419

[4] Computer Science Curricula 2013, Association for Computing Machinery (ACM), IEEE Computer Society, December 20, 2013

[5] National Association for Gifted Children "Jacob Javits Gifted and Talented Students Education Act". nagc.org. Retrieved on December 31, 2007.

[6] B. A. Romay, The talent-ed Dr.Taylor, Finding John Galt: People, Politics, and Practice in Gifted Education, Information Adge Publishing Inc. 83-98

[7] C. W. Taylor, Cultivating new talents: A way to reach the educationally deprived. The Journal of Creative Behavior, 2, 83-90, 1968Taylor, C. W. Cultivating simultaneous student growth in both multiple creative talents and knowledge. In J. S. Renzulli (Ed.), Systems and Models for Developing, 1986.

[8] C. W. Taylor, Cultivating simultaneous student growth in both multiple creative talents and knowledge. In J. S. Renzulli (Ed.), Systems and Models for Developing, 1986.

[9] L. Clark and J. Zubizarreta, "Inspiring Exemplary Teaching and Learning: Perspectives on Teaching Academically Talented College Students" (2008). NCHC Mongeraphs Series.Paper 2

[10] P. M. Olszewski-Kubilius, Special summer and Saturday programs for gifted students. In N. Colangelo & G. A. Davis (Eds.), Handbook of gifted education (3rded.; pp. 219–228). Boston: Allyn and Bacon, 2003

[11] B. Kerr, Nicolas Colangelo, James Maxey and Paula Christensen Characteristics of Academically Talented Minority Students, Journal of Counseling & Development, Volume 70, Issue 5, pages 606–609, May-June 1992

[12] S. Parnes, Source book for creative problem solving. Creative Educ Fndn, 1992.

[13] Mendo.mk

[14] M. Jovanov, B. Kostadinov, E. Stankov, M. Mihova, M. Gushev, State Competitions in Informatics and the Supporting Online Learning and Contest Management System with Collaboration and Personalization Features MENDO, OLYMPIADS IN INFORMATICS, 2013, Vol. 7, 42-54

Leveraging reactive programming in developing applications based on semantic web technologies

Vlatko Dimitrov
Pelagus IT
Skopje, Macedonia
Email: vlatko.dimitrov@gmail.com

Dimitar Trajanov
Faculty of Computer Science and Engineering
Ss.Cyril and Methodius University, Skopje, Macedonia
Email: dimitar.trajanov@finki.ukim.mk

***Abstract*—In this paper we elaborate on the reactive paradigm of designing software and examine how it can be applied in the software design process to adequately handle requests that come as an asynchronous stream. We explain various tactics to reduce the scope of mining big data for information viable to business interests, and we propose a software architecture that accepts linked data and responds reactively to developments it's been designed to monitor, that can be applied to new and current implementations of "smart city" software platforms.**

Keywords—reactive; reactive programming; semantic web; smart cities;

I. Introduction

Applications are constantly increasing in scope and complexity. The ever changing requirements of the business world have given rise to the adoption of multiple programming frameworks, targeted at improving the responsiveness and usability of modern applications. This constant rise in scope and complexity, has made it very hard for developers to maintain and to add new features, instead forcing them to dedicate time to invent ways to keep their applications dynamic to constant modification and refactoring. Multiple methods have been suggested, such as keeping business rules and logic in metadata, which can be changed at runtime, inversion of control and dependency injection, use of scaffolding and code generators.

Regardless of the methods selected, they all add an additional layer of complexity to the system. This behavior leads simple requests in the system to iterate through multiple levels of abstraction, performing many time-expensive and unnecessary checks at runtime. It is necessary to develop applications that are able to adapt and react to their environment, reducing scope and complexity as needed.

The semantic web has one of the most flexible and expressive methods to express data using RDF/XML. And while complicated data structures and the rules which govern them can be expressed using it, these data structures are often impossible to be leveraged in real world scenarios, due to the complicated rules governing their instantiation and traversal.

We propose a reactive approach towards building modern applications – using semantic web technologies to express the business rules in metadata, and using reactive technologies to update the data flow, guaranteeing optimal usage of system resources, as well as a simpler design, replacing complicated rule-checking and application layering with simple methods that react to the dynamic nature of the application, as similar methods have been successfully applied before [1].

II. Reactivity in Theory and in Practice

The trend of designing systems that are able to react and adapt to various changes in their environment is not a new concept, but rather something that emerges as a must - from real system requirements that need to capture, check and respond to direct stimuli, either from a data stream or a moving servo of a robot. The topic is not an unexplored avenue for researchers as well, as the concepts of reactive programming have a scientific approach as well, especially the asynchronous and adaptable approach towards calculating complex operations in real-time [2].

The Reactive Manifesto claims that these issues emerge in the development of various independent software systems and are solved similarly each time, making the systems more robust, loosely coupled and easily responsive to change [3]. This has led to the creation of multiple software development techniques, patterns and good practices which can be re-applied to new projects, so that all can benefit from this style, as a new software engineering paradigm.

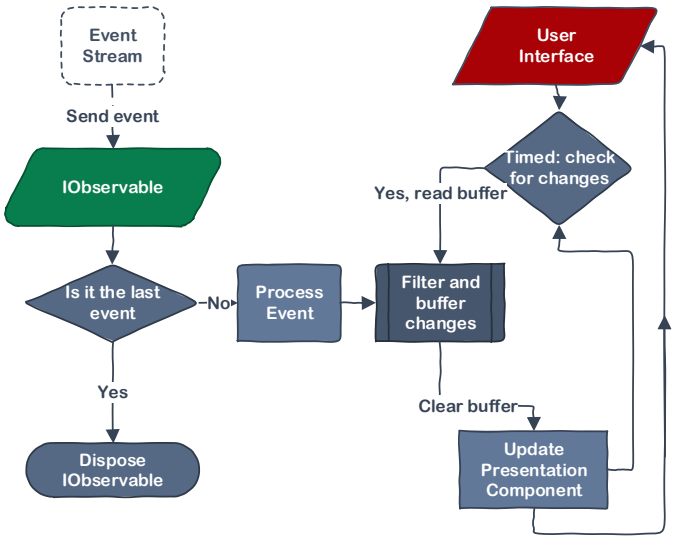
In simple terms, reactive programming can be summarized as "programming with asynchronous data streams" [2]. One of the central tenets of reactive programming is that all data can be represented via a data streams, and we are working with said streams in a functional way - responding to data emitted by them, merging streams, filtering and querying through them etc. In this way, reactive programming is an implementation of the Observer design pattern as suggested by the Gang of Four [4]. As the Observer pattern is composed of two main elements - a Subject class which has a list of objects it should tell when something has changed in its state, and the Observer class which is notified when a change has occurred.

A second important aspect of reactive programming is immutability of data. The data and event streams that it works with are not modified in any way, making operating with them memory and thread safe (as it's often required to operate without knowing the threads or even processes where the execution takes place).

Several programming frameworks (Facebook React, Meteor), have embraced the reactive paradigm, and multiple extensions [5] for various programming languages have been created to enable developers to create reactive applications. Additionally, the method of asynchronous data stream querying has been used in cloud-based web services, such as Azure Stream Analytics, allowing developers to query streams similarly to querying data in a database, in this case - using SQL. This trend is expected to continue in near future, as software companies are starting to see the benefit from using reactive frameworks in their applications, which will lead to better and well supported tools of this nature to be developed.

Reactivity can be used at both the user interface and the business logic level, and it can be used to filter events simplifying asynchronous updates to user interface elements, especially when they depend on external sources (ex. a Twitter stream), as well internally - when processing through a large amount of data depending on user interaction, it can be useful to "observe" elements such as the speed of the user's movement, filter through it, "react" at certain places of interest, thereby skipping a large amount of processing (show only select places), creating a seamless experience, as can be seen on Figure 1.

Figure 1: A visual representation of reactive filtering of events



At the business level, the format of the event changes, but not the need to filter, skip, and react only to the necessary and important elements of it. For example, in traditional business as a new product is launched on the market, response to it is closely monitored, rated and then returned back to the company in a structured format, enabling future iterations to be better and accomplish even more. With an reactive approach, reactions and user feedback can be implemented natively in the program, whether it is only to gather feedback on a greater scale, or to scale and update the product in question in real-time (allocating more servers for features in high demand, caching highly requested content etc.) to their environment, reducing scope and complexity as needed.

III. Leveraging Big Data and Real-Time Analytics

It has become a facet of modern software development to equip applications to quickly deliver high-quality and interesting content to their users as quickly as possible. But said high-quality content comes with high development cost and then it must be delivered to a subset of the users that has shown interest for such type of content. After that, the user's feedback must be garnered and their filter for content correctly adjusted for the future. Additionally, the general knowledge – rating of the content must be analyzed and propagated so that users with similar interests will be exposed to it according to them.

Creating high-quality content often comes at a significantly high prices, so companies must be aware at all times of what is "current" and requested by their customers. This leads to developing applications with complicated architectures that have to:

- Use intelligent ways of measuring user engagement
- Adapt their filters using said feedback for future use
- Order additional content based on said feedback and knowledge gathered

Some companies, such as Netflix, use machine learning to accomplish this, to see and find the correlations established between types of users and types of content, time watched a certain train of content, and reinforcing or weakening the connections between them.

We can look at it as a large graph with many possible ways of traversal, depending on the users and types of content. Where the reactive approach comes into play is observing how changes of the graph affect user engagement, like when a new episode or movie is added of a series that users like, or when others are removed (for example, if their relevant copyright has expired in the meantime). All of this is important for the general picture and health of the company, that needs to quickly gather feedback and make use of any such changes in their future business decisions.

Additionally, with a reactive approach, the volume of big data captured regarding data usage and user interaction can be reduced in real-time, as many users are doing similar things at the same time, adding a different level to data compression as data is meaningfully compressed as it is emitted, possibly saving up space and reducing garbage data in the long run.

A. *Reacting to small changes*

And not only user engagement, market changes as well can effectively be measured even on a smaller time-scale, using a combination of reactive programming and filtering event streams. If we look at the market exchange, changes in the market value of companies and goods are fired really fast by many sources simultaneously. And while changes in the market are reported daily in many newspapers across the planet, for workers at the stock exchange it is necessary to observe them really close, and sometimes seconds are all it takes to notice and adapt to changes in price and volume.

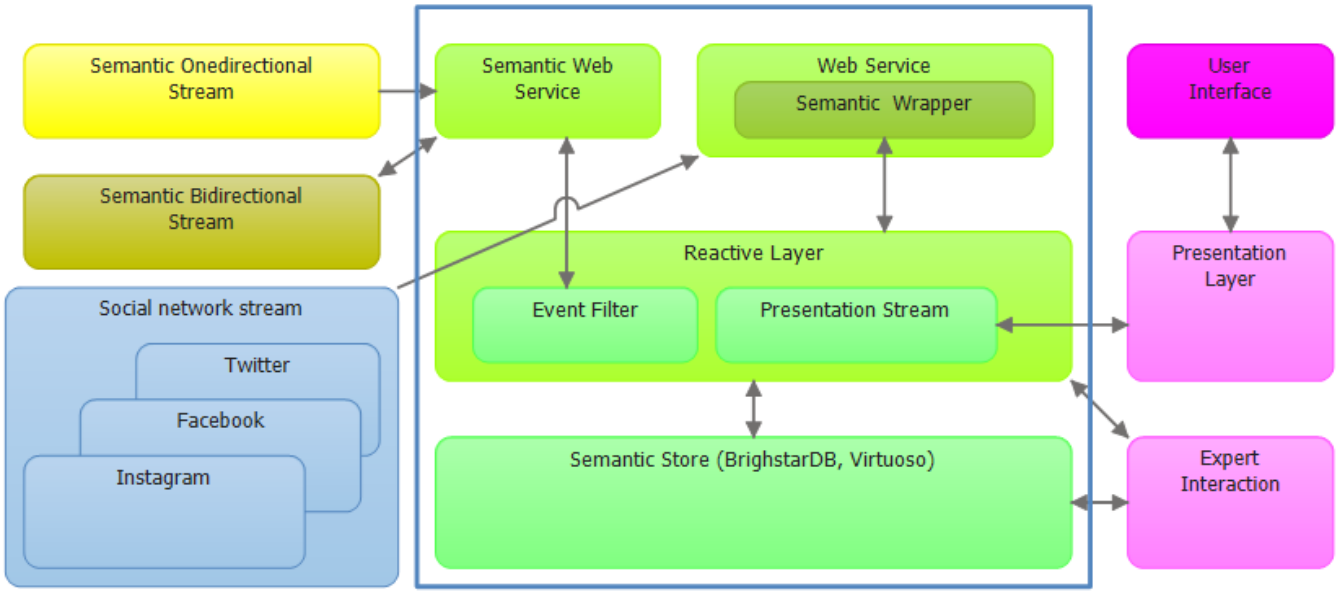


Figure 2: Architecture of a reactive semantic system

This is why reactive systems have found their place in stock exchange applications. Their model of processing and filtering multiple event streams into noticeable visual presentations is of great value to business clients.

IV. Reactivity and the Semantic Web

Semantic web researchers of have a specific interest in adopting a reactive approach towards collection and classification of data. As data and the rules that govern it become more complex, it becomes a challenge to make any significant change within a semantic system, due to the complications that necessarily arise with data cascading between various branches of the system. Several different systems to deal with this have been attempted and documented, like creating ontologies, rules and domain languages to capture and react to events [6-7], as well as and methods using complex programming languages such as Java and Scala [8].

There is no single way, but creating a specified ontology that explains event streams seems to be the general idea, but these rarely approach the problem of modifying an ontology (and any linked data instantiated with it) as circumstances of where and how it's used change.

This is in line with research going on about implementing reactive techniques for programming web applications [7] that need to be responsive, fast and to be able talk with each other and servers using simple communication protocols.

V. Case Study: Smart Cities

It has become a common occurrence in cities to adopt a scientific approach towards urban management and planning [10]. This is usually accomplished by mapping the area as a logical structure using a geographic information system, and then capturing live data generated by sensors of various types (heat, sound, video etc.) The data is then processed by combination of data mining and machine learning, the results gathered and presented into a meaningful format.

There exist several ongoing projects of a city level scale using real-time streaming of linked data [9,10,11] that can benefit from adopting a reactive approach especially when they try to capture a combined IoT element (messages from various devices) with a social aspect [9] as there are many highly complicated elements in each message that need to be correctly categorized. A question on how to generate proper ontologies that encompass the whole scope of what a "smart city" is supposed to be has also been raised [12], with multiple best practices identified and implemented.

An event driven paradigm has already been shown to assist in this process [10] with good results, however the same also raises the complexity of the decision logic, as various conditions and sequences of events must also be checked in runtime. A reactive approach can directly help with this, as it can be added independently of the event system, filtering and reacting to said conditions by default (significantly reducing complexity), asynchronously dealing with checking said conditions without blocking the main systems. and implemented.

An example is the project CityPulse [13], which is known for efficiently capturing and showing real-time IoT approach towards data collection from multiple European cities. This project can additionally benefit from adopting a reactive framework, as it already tries to implement similar features - smart adaptation and user centric decision support. Due to the nature of how the project is designed, these are now higher level features, requiring a lot of overhead to produce relevant results. As the reactive method can be implemented at several different levels, it can significantly reduce unnecessary operations, and make the results instantly visible and quantifiable to end users in a good presentation format.

A specific point of interest is capturing the social aspect of these live cities, as more and more events in the "smart city" receive a social network basis [9] - a parallel which includes pictures, check-ins, tweets from various sources, documenting

any and all occurrences of interest. These can be semantically captured and analyzed, processed with natural language processing and have knowledge about them extracted into a meaningful format.

All of these elements represent at a fundamental level asynchronous event streams and working upon them can be accomplished by the methods expanded upon before.If we're dealing with multiple events streams of similar nature, they can be combined and the data bandwidth reduced by a significant margin, saving up space and processing power as well.

A. An Intelligent city

We propose a unified approach towards solving the problem of a smart semantic city using reactive system engineering - a city that does not only "record" and classify data that originates from within it, but also can "react" to developments in it originating from the data being aggregated.

1) Semantic sensors

The suggested architecture of the system is divided into two parts - the smart aggregation of data, and the reactive method to store and act upon it - providing visible information to the end users. Data is generated by using sensors that emit linked data, either in RDF or JSON-LD format, that's already been linked in some way based on a global system ontology.

The data could be of various formats - a picture from a camera, a measurement from a toxicity or a temperature sensor, GPS coordinate of a location etc, it should be acceptable as long as it's properly embedded in a linked response format.

The data is then passed to a web service that is listening for it on a designated port. The service accepts the linked data, checks its properties and decides if it should "react" to the knowledge it portends.

2) Reactive storage and notifications

The backend incorporates reactivity in 2 places. Firstly, as data that comes can be represented as an asynchronous stream of linked data, it can read the events and merge or filter them, set flags etc. and when a specific condition is met, raise a notification, or generate a new event for a higher level listener. For example, if data is being read via multiple temperature sensors throughout the city, and the temperature readings are being raised in multiple areas simultaneously a heat wave warning could be issued to health departments (that could issue a general warning in general channels). If there's a heat increase in only a specified place, then a possible fire alert could be raised accordingly, and a professional may check (via a camera, or directly) if the warning is valid.

Additionally, the reactive layer can serve as a wrapper for non-linked data events generated towards the web service, for example a twitter stream. As updates of from it can be received in real-time, depending on business needs, they can be filtered or combined into a single semantic message that is correctly tagged as linked data. The web service can adequately process it, and raise further flags as needed.

At the end, the final "relevant" data is stored in a semantic store (Virtuoso, BrighstarDB) documenting events that happened and reactions that followed, along with a grade of how successful said reactions were measured by a system expert (that could be a human, or a different system). This could be further funneled into a machine learning algorithm that could learn important methods and heuristics of when and how events of importance occur, and reactions that could help resolving them. A final picture of the software architecture is represented in Figure 2.

VI. Conclusions

Within this paper we have explained the concept of reactive software programming in theory as well as shown its applications in practice. We have pointed out where and how said principles can be used in modern software development.

We have shown that the practice of developing reactive software can significantly assist in the development of modern applications, especially those which rely on analyzing the shifts in big data and orientation based on real-time data analytics from multiple event sources.

We have confirmed that the current technological stacks are fully compatible with applying said reactive philosophy on higher as well as lower layers to achieve optimum performance.Additionally, our analysis shows that modern applications can be better served important insights regarding their usage and important content by reactively analyzing real-time event streams with combinations of big-data mining.

We have suggested a modern software architecture that leverages semantic web technologies for communication and storing of data, that is capable of reacting in real-time to important events that it's designed to monitor and respond to, in a clear and concise manner.

References

[1] Engineer Bainomugisha, Andoni Lombide Carreton, Tom van Cutsem, Stijn Mostinckx, and Wolfgang de Meuter. 2013. A survey on reactive programming. ACM Comput. Surv. 45, 4, Article 52 (August 2013), 34 pages. DOI=http://dx.doi.org/10.1145/2501654.2501666

[2] The introduction to Reactive Programming you've been missing (https://gist.github.com/staltz/868e7e9bc2a7b8c1f754)

[3] The Reactive Manifesto (http://www.reactivemanifesto.org/)

[4] Erich Gamma, Richard Helm, Ralph Johnson, and John Vlissides. 1995. Design Patterns: Elements of Reusable Object-Oriented Software. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA.

[5] Rx (Reactive Extensions) https://rx.codeplex.com/

[6] José Júlio Alferes, Michael Eckert, and Wolfgang May. 2009. Evolution and reactivity in the semantic web. In *Semantic techniques for the web*, François Bry and Jan Małuszyński (Eds.). Lecture Notes In Computer Science, Vol. 5500. Springer-Verlag, Berlin, Heidelberg 161-200.

[7] Bob Reynders, Dominique Devriese, and Frank Piessens. 2014. Multi-Tier Functional Reactive Programming for the Web. In *Proceedings of the 2014 ACM International Symposium on New Ideas, New Paradigms, and Reflections on Programming & Software* (Onward! 2014). ACM, New York, NY, USA, 55-68. DOI=http://dx.doi.org/10.1145/2661136.2661140

[8] Calbimonte, Jean-Paul. "Scalable and Reactive Programming for Semantic Web Developers." Proc. of the ESWC Developers Workshop 2015, at ESWC 2015.. Vol. 1361. No. EPFL-CONF-212725. CEUR-WS, 2015.

[9] Athena Vakali, Leonidas Anthopoulos, and Srdjan Krco. 2014. Smart Cities Data Streams Integration: experimenting with Internet of Things and social data flows. In Proceedings of the 4th International Conference on Web Intelligence, Mining and Semantics (WIMS14) (WIMS '14).

ACM, New York, NY, USA, , Article 60 , 5 pages. DOI=http://dx.doi.org/10.1145/2611040.2611094

[10] Cretu, Liviu-Gabriel. "Smart cities design using event-driven paradigm and semantic web." *Informatica Economica* 16.4 (2012): 57.

[11] Sheth, Amit, Christopher Thomas, and Pankaj Mehra. "Continuous semantics to analyze real-time data." IEEE Internet Computing 14.6 (2010): 84.

[12] Sergio Consoli, Misael Mongiovic, Andrea G. Nuzzolese, Silvio Peroni, Valentina Presutti, Diego Reforgiato Recupero, and Daria Spampinato. 2015. A Smart City Data Model based on Semantics Best Practice and Principles. In Proceedings of the 24th International Conference on World Wide Web (WWW '15 Companion). ACM, New York, NY, USA, 1395-1400. DOI=http://dx.doi.org/10.1145/2740908.2742133

[13] Barnaghi, P., Tönjes, R., Höller, J., Hauswirth, M., Sheth, A., & Anantharam, P. (2014). Citypulse: Real-time iot stream processing and large-scale data analytics for smart city applications. In Europen Semantic Web Conference (ESWC).

Information Technology as Enabler for Bussiness Proccess Reengineering – A Successful Case Study

Dragan Jovanoski
Faculty of Information and Communication Technology
FON University Skopje, Republic of Macedonia
dragan.jovanoski@fon.mk

Toni Malinovski
Faculty of Information and Communication Technology
FON University Skopje, Republic of Macedonia
toni.malinovski@fon.mk

***Abstract* — Business process reengineering enables fundamental rethinking and radical redesign of the outdated rules and fundamental assumptions that underlie operations within the organization. The information technology provides new advantages to business operations and can be used as a tool for business process transformation that crosses several functional lines. However, despite the significant growth of the business process reengineering concept, such projects can still fail and don't achieve their intended results. Following the business process reengineering theoretical framework this study elaborates a reengineering project in the "Goce Delcev" University – Shtip in Macedonia, its implementation phases and set of measures, while evaluating benefits to students and to the organization itself. The results show that this technology enabled reengineering project provided cost-cutting effects, decreased time to process completion, while achieving better efficiency and increased student satisfaction.**

Keywords* — *information technology; bussiness process reengineering; reengineering project; customer satisfaction

I. INTRODUCTION

Different companies undertake process changing initiative to decrease the cost of producing a product or providing a service, to decrease the time for execution of specific processes, as well as increase efficiency and effectiveness of specific functions. Having in mind that a business process can be defined as specific ordering of work activities across time and space that produces specific customer-centric outputs, business process reengineering (BPR) aims to achieve dramatic improvements in the measures of performance, such as cost, quality, service and speed [1], [2]. The information technology (IT) has become fundamental enabler for BPR, while providing new sources of advantages for business operations [1], [3]. Hence, business can use IT to increase direct access to people and information without intermediaries, while cutting superfluous layers in the business process. IT mechanism can be used to reduce the organizational complexity and enable wider organizational change. Still, in order to have real improvements the change has to be performed by people who know the objective and know how to use the IT tools to achieve it [5].

The concept of reengineering was adopted by organization in early 1980's, but BPR projects can still fail and don't achieve their intended results [3], [4]. BPR can fail due to lack of strategy, management support, insufficient measured and planning activities, as well as improper comprehension of IT's potential role. Having in mind that BPR should not be treated as an automation process, BPR projects should use the technology in creative and innovative way with a customer-centric approach [3], [4], [6]. In order to have lasting effects organization must examine how strategy and reengineering can complement each other [7], with proper attention given to human dimension of organizing and running systems effectively.

In this article we explore the role of technology as basis for processes reshaping to meet organization's objectives, while evaluating implications for the involved parties, especially form a customer-centric point of view. Since the adaptation of proper methodology is essential for the success of radical process change when adopting BPR, we have used the theoretical framework to analyze a BPR project in University in Macedonia. In line with the organization's vision to increase simplicity and efficiency within students' administrational activities, the University has identified the process to be redesigned by an innovative solution via the support of technology. Hence, the change has enabled cost and time reduction, increased output quality and student satisfaction as an example for successful integration of BPR and technology.

This paper is organized as follows: section 2 provides related work, section 3 elaborates the case study for a BPR project, section 4 provides discussion and results, while section 5 concludes the paper.

II. RELATED WORK

Business processes often cross functional areas, which creates difficulty in the performance measures [5], [7]. But, we can improve the process and reduce the cost and cycle time, while increasing output quality and customer satisfaction. Hammer and Champy [2], as one of the first BPR theoretical researchers, have stated that the role of IT in a BPR project is to challenge the assumptions inherent in the work processes, long before the development of the latest IT solutions. They suggested a BPR methodology with six phases: (1) start of BPR and preparing a vision statement, (2) identification of business processes, (3) selection of business processes, (4) understanding and (5) redesign of the selected

business processes, and (6) implementation of the redesigned business processes. Even though the rest of BPR methodologies have different number of stages, they follow a similar path and focus on vision development, understanding of current process structure and processes to be redesigned, as well as implementation of the new process [5], [8], [10]. In addition, latest model modifications include evaluation of the new process [8], [9], especially form a customer-centric point of view.

BPR is often contrasted to Total Quality Management (TQM), which focuses on incremental improvements based on business process improvement [1], [5]. Even though they have similar aspects, BPR differs as a top-down approach with dramatic improvements, relatively short time frame for implementation and measurable progress and results [9], [11]. In addition, BPR is highly rated in terms of innovation and IT utilization, while TQM provides best practices for continuous improvement [11]. Still BPR can be used to completely redesign the process, while TQM can be used to continuously improve the same process in the years to come.

Davenport and James [1] have identified the IT as a key enabler in BPR, while specifying the critical IT capabilities and their organizational impacts. They have identified IT levers and listed transactional, geographical, analytical, informational, sequential and disintermediation capabilities of IT as important factors that can reshape the business processes. Therefore the IT capabilities should influence the process redesign and initiate new options, rather than just supporting them.

This literature review shows that the BPR initiative provides a range of activities that can dramatically improve the business processes, including a new structure, top-down approach with top-management support, customer-centric focus, and validated theoretical framework and best practices.

III. A Case Study for BPR Project

Based on the Hammer and Champy [2] BPR methodology and Davenport and James [1] assumptions for the IT levers, we have evaluated a BPR projects in the "Goce Delcev" University – Shtip in Macedonia, its implementation phases and set of measures, while analyzing the improvements for the redesigned business process form a customer-centric point of view.

Having in mind that the initial step in BPR is the vision statement [2], [5], as well as the need to obliterate the existing process and take radical steps towards new one, the "Goce Delcev" University – Shtip has questioned the business process that relates to payments for student services. This process, "as is", was inefficient due to several internal and external factors, such as dealing with financial institutions that have certain defined work rules when it comes to transfer of deposited funds in a strictly defined period of time, bottlenecks on support offices for students, etc. In practice if the payments were made until 2:00 PM in a local bank, the funds were transferred via the Treasury of the Republic of Macedonia to the University the same day. If the payments were made after 2:00 PM, the transfer of funds was processed the next working day. This process was ineffective in practice, since there were cases when the students couldn't perform their student activities associated with financial transactions, if the payment was made on the last date defined for certain student activity.

The University management has decided to change the payment process via a radical change, in order to dramatically improve students' service, reduce costs and increase satisfaction, as well as save precious time to the students and University administration. During BPR it is important to identify the business process and its internal tasks that have to be redesigned. Since one of the University strategic goals is student focus for the educational and administrational activities and increased students' satisfaction, the BPR of the payment process for services can make a change in different cross functional areas and reduce organizational complexity, with increased efficiency. Thus, the decision to redesign this process among others, justifies the need for change since the process is problematic, fragmented and inefficient, and the redesign will affect process time and cost, as well as will probably increase students' satisfaction, which will contribute to the University's strategic vision and goals.

A process map can graphically represent the activities within the process, which can be further decomposed into sub-processes, shown in different maps. The business process related to students' payment for services within the University is executed in 6 steps, graphically depicted as a flow of events in Figure 1.

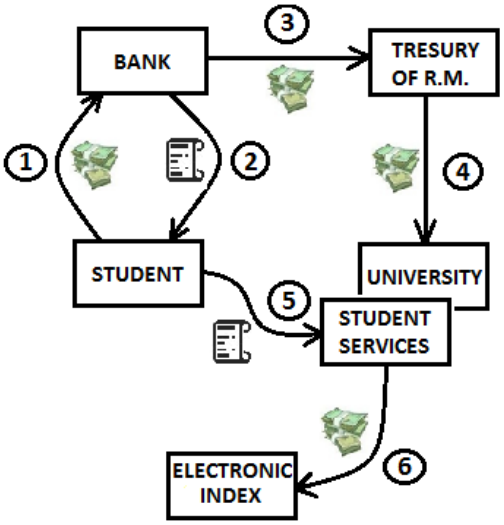


Fig. 1. Students' payment for services - flow of events

A payment within the process ("as is") is processed via the following order:

1. The first step shows the payment activity by students in one of the financial institutions i.e. banks.

2. The second step involves confirmation by the bank for the deposited funds, i.e. verification of the payment to the University.

3. During step three, the funds are transferred from the bank to the Treasury of the Republic of Macedonia.

4. The fourth step marks the funds transfer from the Treasury to the University.

5. The fifth step includes payment validation by the University administration. The students are obliged to appear in the student services department, so the administrator can verify the payment order.

6. As a final step, after the verification of the payment by the student services department, deposited funds are recorded in the electronic index of the students, so the students can complete their activities related with financial transactions.

The comprehension of a selected business process is a crucial step in the BPR project. Hence, the detailed understanding of the old process, will avoid possible mistakes while the new business processes is designed. In this study, the first step includes a students' activity for payment in a bank, which should be differently organized with the use of IT. Thus, the next step can be automated or fully obliterated. The transfer of funds between the banks, the Treasury of the Republic of Macedonia and the University include steps that are time dependant and involve several institutions, which often leads to significant delays. Hence the students weren't able to timely perform their academic activities related to the financial transactions, missed deadlines, inefficiencies during operations regarding identification services and registration of payment, etc. The validation of funds by the administration in the University requires an obligation by the students, which have to appear in the student service department and show the verified bank receipt. This step is inefficient and redundant, which also creates bottlenecks during other operations in the student services department. A typical example is the peak during exam enrollment sessions throughout the year, registration of semesters, as well as registration of new students. Finally, after the verification, the funds are recorded in the electronic students' index and the students can continue their activities.

Based on the BPR theoretical framework, as well as one or more knows principles of reengineering [1], [2], [5], the University has shown deep knowledge for each activity within the process, what can be dramatically changed and improved, and which steps can be completely avoid. In addition, it has identified the IT as the key enabler of BPR, with proper alignment between strategic goals, business activities and students. Figure 2 illustrates the connection between the existing process ("as is") and the redesigned process ("to be").

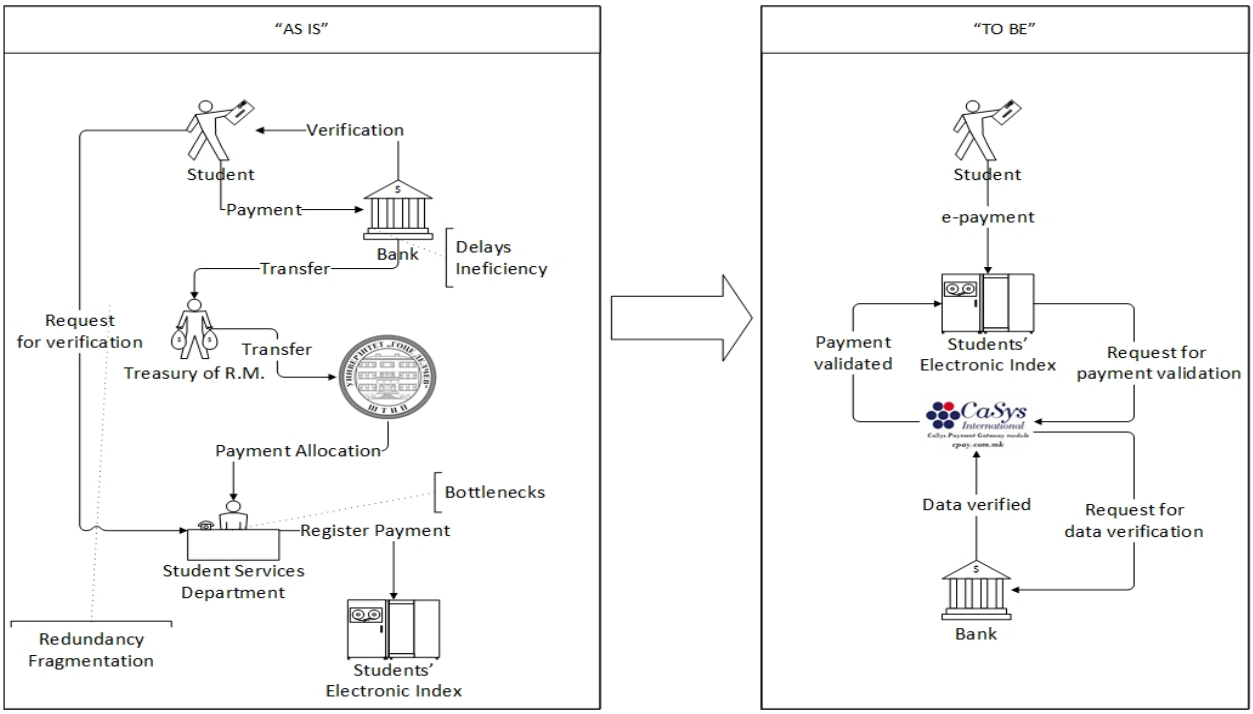


Fig. 2. Existing proccess for students' payments ("as is") and the redeisgned proccess ("to be")

The redesigned business process for students' payments for services in the University is based on the IT platform where the students' electronic indexes reside, via an additional module for electronic payment (e-payment). Hence, as a result of the BPR project, the redesigned business process has the following steps:

1. Students initiate payment via the integrated module for e-payments with its own payment card issued by a bank (Master, Visa, etc.).

2. The integrated module is connected to a cPay platform for transactions processing of CaSys International Company, which processes the financial transaction according to the payment information and communicates with the student's bank for transaction verification.

3. The bank returns to the cPay platform an appropriate response, regarding validity of the account and issued payment.

4. The cPay returns forwards this information to the e-payment module within the electronic index of the student.

5. If the answer is positive, the financial transaction is registered in the electronic index of the student. On the other hand, if the answer is negative, failure information is transmitted, with the more details regarding the failure cause.

By capturing information once and at the source [2], [5], integration between the electronic index and the module for e-payments is established, so the redundant steps of financial validations were removed. Thus, the time consuming activities were eliminated via proper alignment between the IT and the business process. In addition, while transferring most of the activities to the students, without unnecessary administration, the University has enabled those who use the output of the process to perform the process, as a key BPR principle [2]. In like manner the BPR project was organized around outcomes, not tasks and all dispersed resources were consolidated via a streamlined method of payment, which provides solid position for success [2], [4], [6].

IV. Discussion and Results

Having in mind that BPR projects can still fail and don't achieve their intended results [3], [4], an analysis of the results is needed after each BPR effort. Also, the successful IT investment has been increasingly considered by many researchers and practitioners as a vital component of successful BPR efforts [2], [3], [12]. In this case study, the student focus and increased satisfaction was the primary goal, as well as the need to decrease the cost and process completion time. Therefore, the process was redesigned via proper integration of the IT solution with the business process to achieve desired objectives with measurable outcomes.

This BPR project regarding the students' payments for services in the "Goce Delcev" University – Shtip in Macedonia was implemented in 2013/2014 (pilot/full production), allowing coexistence with the old process for comparison. Before BPR of the referred process, 100% of payments were perform in the premises of banks. As soon as the BPR project has finished, the situation started to change and students started to use the redesigned process.

Figure 3 shows the intensity of payments for the old process ("as is") and the newly established process ("to be") for the years 2013, 2014 and 2015.

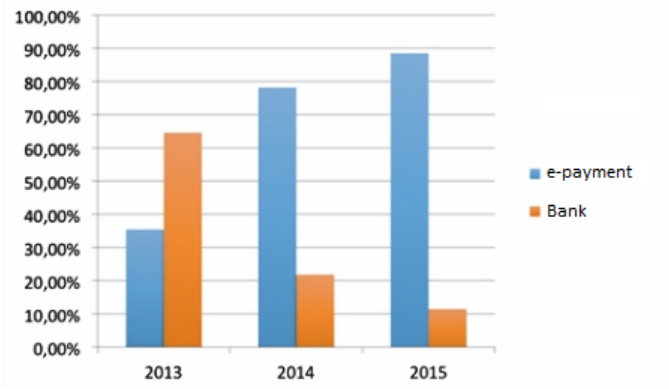


Fig. 3. Students' acceptance of the redesigned bussines proccess

This statistical information clearly shows the impact of the dramatic change and students' acceptance of the new IT enabled solution, after the BPR efforts. Hence, during 2014 and 2015 the trend for students' payments on the banks' premises declined dramatically. Even though the number of payments for both processes, old and new, is helpful, further analysis is need to understand students' experience and factors influencing their satisfaction. Therefore we have chosen several indicators, as process performance measures from students point of view, that formulate a complex construct, representing students' satisfaction after the BPR project. We have designed a survey with questions based on these indicators and surveyed students' opinion on each question and obtain overall picture for the achievement of desired outcomes. The students were able to grade each indicator on a seven-point Likert scale [13], where 1 = not satisfactory and 7 = excellent. We have performed a Cronbach's alpha test [14] to measure internal consistency of indicators within the students' satisfaction construct. Table I shows the chosen indicators for the students' satisfaction construct, as well as results from the students' responses to the survey.

TABLE I. Students' Satisfaction Indicators and Survay Results (n = 244)

Indicator	Description	Mean	Standard Deviation	Skew	Kurtosis	Cronbach's alpha
Sat_1	Students' opinion for the decreased time for proccess complition	5.75	1.816	-1.496	1.226	
Sat_2	Students' opinion for the cost-cut	5.82	1.894	-1.620	1.371	
Sat_3	Diffrence between students' expectations and reall practice during e-payments	5.59	1.829	-1.405	0.939	
Sat_4	Students' overal expirience for the newly establish proccess	5.41	1.858	-1.236	0.505	
Sat_5	Students' opinion for the administrative activities within the Universition in general	4.73	1.981	-0.631	-0.783	
Sat_6	Students' overall satisfaction from the University	4.39	2.006	-0.385	-1.067	0.906

We have received 244 responses to the survey, from 63.9% female and 36.1% male students. The descriptive statistics show that students' responses were constrictive in general and that data distribution is within the recommendations (skew < 3.0 and kurtosis < 8.0 [15]).

In addition, the Cronbach's alpha value is much higher than the 0.7 minimum [14], which shows strong internal consistency among indicators chosen to represent the students' satisfaction construct. The results show that students' opinion for the cost-cutting effect of the BPR project was highly rated, followed by the decreased time to complete the process. The BPR efforts increased the process efficiency, which was positively reflected on the students' satisfaction. The lowest performance indicators were the students' general option for the administrative activities within the University, as well as students' overall satisfaction.

Therefore we can conclude that an increase of students experience should continue to be one of the University primary strategic goals, since the BPR project itself produced good results reading different aspects of students' satisfaction from the e-payment redesigned process, with proper IT investment.

V. CONCLUSION

BPR focuses on the key business processes in one organization, while questioning the status quo situation, and defining certain objectives [5]. It is a strategy driven approach that evaluates, prioritizes objectives and identifies new products and process opportunities. IT has a distinctive role in shaping tomorrow's business operations and thus can be a key enabler for BPR [1]. Even though there is sound theoretical framework, a BPR projects can still fail if they are not conducted properly [3], [4], [6].

Following a BPR project for the implementation of e-payments services in the "Goce Delcev" University – Shtip in Macedonia, this study elaborates its implementation phases, as well as researches different variables that influence students' experience and satisfaction after the BPR efforts. It shows the benefits to students, indicated as cost-cutting effects, decreased time to process completion and positive experience during everyday practice for e-payments. The results can also be useful to the management structure to reflect on the University strategic goals, as well as to the BPR project team for quantification of the student impact.

In our future work, we plan to continue with the established methodology and research additional factors and their mutual relationships, which are important during BPR and influence the success of the whole project, as well as customer satisfaction. Hence, we hope to provide links between strategic goals, human attitude towards the change during BPR, technology, customer experience, etc.

REFERENCES

[1] T. H. Davenport and E.S. James, "The new industrial engineering: information technology and business process redesign," Sloan Management Review, vol. 31, no. 4, 1990

[2] M. Hammer and J. Champy, "Reengineering the Corporation: A Manifesto for Business Revolution," Nicholas Brealey, London, 1993

[3] M.N. Habib, "Understanding Critical Success and Failure Factors of Business Process Reengineering," International Review of Management and Business Research, vol. 2, no.1, pp. 1, 2013

[4] M. Al-Mashari, M. Zairi, "BPR implementation process: an analysis of key success and failure factors," Business process management journal, vol. 5, no. 1, pp. 87-112, 1999

[5] V. Sethi and W. R. King,"Organizational transformation through business process reengi-neering: applying the lessons learned," Pearson College Division, 1998

[6] P. Ravesteyn and R. Batenburg, "Surveying the critical success factors of BPM-systems implementation," Business Process Management Journal, vol. 16, no. 3, pp. 492-50, 2010

[7] A. Goksoy, B. Ozsoy, and O. Vayvay, "Business process reengineering: strategic tool for managing organizational change an application in a multinational company," International Journal of Business and Management, vol. 7, no. 2, pp. 89, 2012

[8] A. Chiarini, "Business process reengineering," From Total Quality Control to Lean Six Sigma, , Springer Milan, pp. 25-27, 2012

[9] V. Weerakkody, M. Janssen, and Y.K. Dwivedi, "Transformational change and business process reengineering (BPR): Lessons from the British and Dutch public sector," Government Information Quarterly, vol. 28, no. 3, pp. 320-328, 2011

[10] M. Subramanian, L. Whitman, and S.H. Cheraghi, "Business process reengineering: a consolidated methodology," Proceedings of the 4 th Annual International Conference on Industrial Engineering Theory, Applications, and Practice, 1999 US Department of the Interior-Enterprise Architecture, 2006

[11] Y. F. Jarrar and E.M. Aspinwall, "Integrating total quality management and business process re-engineering: is it enough?," Total Quality Management, vol. 10, no. (4-5), pp. 584-593, 1999

[12] A. Nowak, F. Leymann and R. Mietzner, "Towards green business process reengineering," Service-Oriented Computing, Springer Berlin Heidelberg, pp. 187-192, 2010

[13] R. Likert, "A technique for the measurement of attitudes," Archives of Psychology, New York: Columbia University Press, 1931

[14] L. J. Cronbach, "Coefficient alpha and the internal structure of tests," Psychometrika, vol. 16, no. 3, pp. 297-334, 1951

[15] P. J. Curran, S. G. West and J. F. Finch, "The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis," Psychological methods, vol. 1, no. 1, pp. 16-29, 1996

Naïve Bayes classification of nitrogen impact on lake ecosystem

Andreja Naumoski, Georgina Mirceva
Faculty of Computer Science and Engineering
Ss. Cyril and Methodius University in Skopje
Skopje, Republic of Macedonia
{andreja.naumoski, georgina.mirceva}@finki.ukim.mk

***Abstract*— The results from the analyses of the measurement data can be used to support the process of understanding the functionality of the ecosystem. Many natural processes and phenomena leave traces in the history of the ecosystem, and the only way to get that information is through analysis of this data. Thus, this paper aims to use machine learning methods to extract this information in order to assess the impact of nitrogen component. The data that are used for analyses is measured at several measuring points and contains both physico-chemical information and diatom abundance. Each diatom can be classified into one of trophic quality classes (TQCs) and therefore, the Naïve Bayes method is used to build one model for several classes. The obtained results are presented on a geographical map using a GIS in order to determine the important places where nitrogen has the greatest impact on the ecosystem biota.**

Keywords*— *Naïve Bayes, nitrogen component, classification models, trophic quality models.

I. Introduction

The analysis of the ecological measured data and the process of extraction that environmental information can be done in many ways. The data can be a subject of different type of analysis, like: clustering, correlation analysis, analysis of variance, entropy or classification. These types of analysis open a way to new knowledge discovery about the lake ecosystem and deepens our understanding about the ecosystem itself. Based on these facts, the data can be transformed and analysed to provide a better understanding of internal processes affected by changing environmental factors. Since the diatoms are ideal bio-indicators, and preserve important information of the abiotic impact, their classification can help in the process of discovering the link between certain processes and abiotic factors or physic-chemical parameters. Based on the results of this analyse, we can conduct classification of a given lake ecosystem. To start this process, we need to define classes where certain diatom will be classified in certain trophic quality class (TQC). To be compatible with the known literature, the classification system based on nitrogen parameter will be taken into account. TQC of water quality can be obtained for a specific set of physico-chemical parameters responsible for the process of eutrophication of the lake ecosystem. The process of eutrophication is a very important process that impacts on the life cycle of many organisms. Several physical-chemical parameters can serve as indicators of such processes: the total concentration of phosphorus, total nitrogen concentration and Secchi disk [1]. The total concentration of phosphorus and Secchi disk were studied in previous papers [1]. Based on these studies, in which the influence of these two parameters was analysed, the classification of the total concentration of nitrogen was not analysed. This can contribute significantly to expand the knowledge about properties of bio indicating diatoms. In order to get a better interpretation of the models, the measurement data is discretized and labelled. As previously research paper [2] concluded, the discretization of the input data is performed yet for another reason. There is evidence that the Naive Bayes method produces low values for probabilities, or performs poorly in classification process for a given class, if the input measurement data have a value close to zero [3, 4]. Additionally, some researchers suggest other techniques to avoid this drawback [4].

In the literature we can find other research works that are focused on finding the influence of physico-chemical parameters on the diatoms and vice versa, using classical statistical approach, such as canonical correlation analysis (CCA), detrended correspondence analysis (DCA) and principal component analysis (PCA). These techniques are the most used modelling techniques for discovery of the diatom indicating properties [5]. They are also useful to provide information on the measured data, but they are limited in terms of representativeness and they contain a certain degree of domain expert subjectivity. These methods build models with different types of input conditions, but they do not provide detailed information on the basis of new data and the accuracy of prediction. Additionally, the biological expert should have an understanding of graphs, where transform axes are plotted and given interpretation of distance and form clusters of points on the image. This is one of the main reasons why this study extends these methods, to methods that provide tools with better classifying accuracy and more interpretable graphical tools. One such type of methods are the decision trees [6]. These methods do not only aim to improve the interpretability, but also they have relatively high classification accuracy. Several attempts to model the relationship between the diatoms as indicators and the abiotic environmental factors are made in [6, 7, 8, 9]. The models obtained with these decision trees techniques, are verified with the known knowledge found in the ecological literature. The next step towards improvement of these methods, is made by introducing a new class of multi-

objective decision trees, in order to understand the dynamics of the whole range of physico-chemical parameters that are important for the ecosystem [9]. Although these methods contributed in increasing the accuracy and comprehensiveness of the models; they are not robust to the dynamic nature of environmental measurements. This important property has to be somehow imported into the classification process and even more the probability of finding the certain diatom in certain trophic quality class.

So, methods like Naïve Bayes could help in understanding the environmental information. Studies conducted in previous years, indicate that the performance of Naïve Bayes and decision trees (C4.5 [10]) are equally good for number of real data domains [11, 12, 13]. In many studies, the researchers conclude that the good performance of the Naïve Bayes method, though sometimes surprisingly, not due to the basic requirement of this method regarding the attribute independence property. More evidence for the use of Naïve Bayes technique on the diatom indicating features comes from Domingos and Pazzani [14] study, in which they showed that Naive Bayes method due to its good performance comes from the zero-one loss function. Another important feature of Naive Bayes is that this algorithm shows better results for classification instead of regression [15]. These issues needs to be further investigated, because in the lake ecosystem one diatom can be indicator not only for one, but for several parameters and some of the organisms are competing among themselves [1]. However, from an environmental perspective, it is important to assess the degree of belonging of a given diatom in a specified environment. In this paper we assume that the impact of physico-chemical parameters on different diatoms is independent, which will satisfy the basic condition under which this method operates.

The rest of the paper is organized as follows: Section II provides the definitions of the method itself. In Section III we present the input dataset and the trophic quality classes. Furthermore, Section IV gives the experimental models followed by the presentation of the graphical maps and the verification of the model results. Finally, Section V concludes the paper and presents directions for future research.

II. NAÏVE BAYES METHOD

First, the Naïve Bayes classification algorithm constructs a model based on a set of training examples with specific label. The standard definition of classification for the given example E is represented by a tuple of attribute values (x_1, x_2,.., x_n), where x_i is the value of attribute X_i. From another point, the classification variable C is represented with the value of c. The input dataset y is processed the algorithm that assigns a function and then this function appoint a class label to the given example. According to the Naïve Bayes classifier, the probability that a given example $E = (x_1, x_2,.., x_n)$ belongs to class C is:

$$p(c \mid E) = \frac{p(E \mid c)p(c)}{p(E)} \tag{1}$$

In this stage the algorithm assumes that all attributes are independent by a given value of the class variable; that is,

$$p(E \mid c) = p(x_1, x_2, ..., x_n \mid c) = \prod_{i=1}^{n} p(x_i \mid c) \tag{2}$$

The resulting classifier is then:

$$f_{NB}(E) = p(C) \prod_{i=1}^{n} p(x_i \mid C) \tag{3}$$

The function $f_{NB}(E)$ is called a Naive Bayesian classifier, or simply Naive Bayes. In order to estimate the probability that one diatom belongs into one TQC, we will use the standardized normal distribution or Gaussian distribution, expressed as:

$$f(x \mid \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{(x-\mu)^2}{2\sigma^2}} \tag{4}$$

The implementation of the Naïve Bayes classifier in WEKA software is used to build the diatom models [16]. The implemented algorithm provides estimation of both μ (mean value) and σ (standard deviation) for each variable and diatom trophic class, and then calculates the probability that a given diatom belongs in a given trophic class. The x value is inputted as discrete class terms, because of the ecological (uncertainty) nature of the diatom dataset, and better performance is reported in [3, 4]. The values for the discrete classes are given below, together with the labels that provide better interpretability. Just a short note, Diatoms Term 0 (DTerm 0) lays in a range of 0 to 3.25, and they are labelled as bad indicators. The normal distribution takes as input the value of a discrete class of the TQC in Table I.

TABLE I. THE MAXIMAL VALUE FOR EACH NITROGEN BASED TQC DTERM

Diatoms	**DTerm 1 – DT1**	**DTerm 2 – DT2**	**DTerm 3 – DT3**	**DTerm 4 – DT4**
Label	Weak	Good	Very Good	Excellent
APED	3.25	6.5	9.75	13
CJUR	21.5	43	64.5	86
COCE	20.25	40.5	60.75	81
CPLA	10	20	30	40
CSCU	10.25	20.5	30.75	41
DMAU	3	6	9	12
NPRE	4.75	9.5	14.25	19
NROT	6	12	18	24
NSROT	7.75	15.5	23.25	31
STPNN	5.25	10.5	15.75	21

III. DATA DESCRIPTION

The dataset used for the experiments contains 11 input parameters, from which 10 relate to the 10 most abundant diatoms collected by the biological expert [17] during the TRABOREMA PROJECT [18], and the 11-th parameter is the nitrogen parameter that refers to the class of the trophic state index.

TABLE II. TROPHIC QUALITY CLASSES FOR THE TOTAL NITROGEN PARAMETER

Parameter	Name of the Trophic Quality Class	Range
Total Nitrogen	*Oligotrophic*	0-90 mg/l
	Mesotrophic	90-100 mg/l
	Eutrophic	100-120 mg/l
	Hypereutrophic	120-140+ mg/l

The TQC is defined according to the classification system proposed in [19], which is experimentally obtained for lakes in Florida and is consistent with the Carlson classification system for trophic indices. The definition of such a system is given in Table II. For experimental evaluation, we designed several types of experiments to test the obtained models and later to validate the known ecological knowledge in the diatom literature. The results of the experiments estimate the highest probability of belonging of a given diatom in a given class of trophic index. Once the algorithm processes the data, the resulting model shows the affiliation of each diatom for each class of trophic indices separately. For calculation, the normal distribution probability distribution is used.

IV. EXPERIMENTAL RESULTS

A. Interpretation of the Classification Models

The algorithm defines certain range of the discretized class terms, so estimating them is important in the process of interpretation. The output of the algorithm depicts the probability estimate of the indicating properties of a certain diatom.

TABLE III. PROBABILITY THAT A GIVEN DIATOM BELONGS TO A CERTAIN CLASS ACCORDING NAÏVE BAYES

Diatoms	Bad	Weak	Good	Very Good	Excellent
Class APED	*oligotrophic* 99.73%	*hypertrophic* 15.89%	*mesotrophic* 8.05%	*mesotrophic* 2.78%	*mesotrophic* 0.44%
Class CJUR	*mesotrophic* 35.28%	*eutrophic* 2.45%	*eutrophic* 0.00%	*eutrophic* 0.00%	*oligotrophic* 0.00%
Class COCE	*hypertrophic* 1.51%	*hypertrophic* 2.41%	*mesotrophic* 2.07%	*oligotrophic* 1.06%	*oligotrophic* 0.47%
Class CPLA	*eutrophic* 99.73%	*eutrophic* 4.82%	*eutrophic* 0.04%	*eutrophic* 0.00%	*eutrophic* 0.00%
Class CSCU	*oligotrophic* 16.13%	mesotrophic 7.74%	*eutrophic* 2.88%	*eutrophic* 0.40%	*eutrophic* 0.02%
Class DMAU	*oligotrophic* 99.73%	*eutrophic* 13.87%	*mesotrophic* 7.24%	*mesotrophic* 1.69%	*mesotrophic* 0.16%
Class NPRE	*mesotrophic* 77.76%	*hypertrophic* 12.35%	*hypertrophic* 3.29%	*hypertrophic* 0.25%	*hypertrophic* 0.01%
Class NROT	*oligotrophic* 99.73%	*eutrophic* 12.44%	*eutrophic* 0.89%	*eutrophic* 0.01%	*eutrophic* 0.00%
Class NSROT	*oligotrophic* 99.73%	*mesotrophic* 11.13%	*mesotrophic* 1.28%	*mesotrophic* 0.01%	*mesotrophic* 0.00%
Class STPNN	*oligotrophic* 99.73%	*eutrophic* 9.00%	*eutrophic* 0.30%	*eutrophic* 0.00%	*eutrophic* 0.00%

The results of the classification model are presented in Table III. All the diatoms are interpreted in similar way, so we give only several examples for interpretation of the models. As we can see from the table, APED diatom is bad indicator of *oligotrophic* waters, but good to excellent of *mesotrophic* waters with probability less than 8%. Additionally, the classification model identifies the APED diatom as weak indicator of *hypertrophic* diatom with probability of almost 16%. Other diatoms obtained similar results. The COCE diatom is bad and weak indicator of *hypertrophic* waters, while good to excellent indicator of *oligotrophic* waters. The CPLA diatom, according the model is an *eutrophic* diatom with high probability to be weak indicator of *eutrophic* waters and low probability to be excellent indicator of these waters. Totally opposite is the NROT diatom, which is a bad indicator of *oligotrophic* waters, but weak to excellent indicator of *eutrophic* waters with probability less than 12.44%. And if we see the STPNN diatom, according the model he is a bad indicator of *oligotrophic* waters, but weak to excellent with more than 9% for *eutrophic* waters etc. It is important to note that the bad and the weak indicating properties of certain diatom is not due to using inappropriate method for classification, but more due to the quality and quantity of the measured data. This was concluded for this diatom dataset in

experiments with previous methods [5, 7, 8, 9]. Also, is important to note that the classification model classifies some of the diatoms as bad indicators, because mainly of the data contained values of diatoms abundance near zero. The design of the experiments must be reconsidered or to be defined to reduce this error of closing the values of the probability near to zero in order to get more accurate models. In this direction some of the results obtained from the model may or not fit in the known diatom literature.

B. Geographical Representation of the Models

Previously obtained classification model with Naive Bayes technique can be represented as a spatial model in GIS. Two models are presented in GIS to examine spatial information in terms of the impact of physico-chemical parameters on the diatoms.

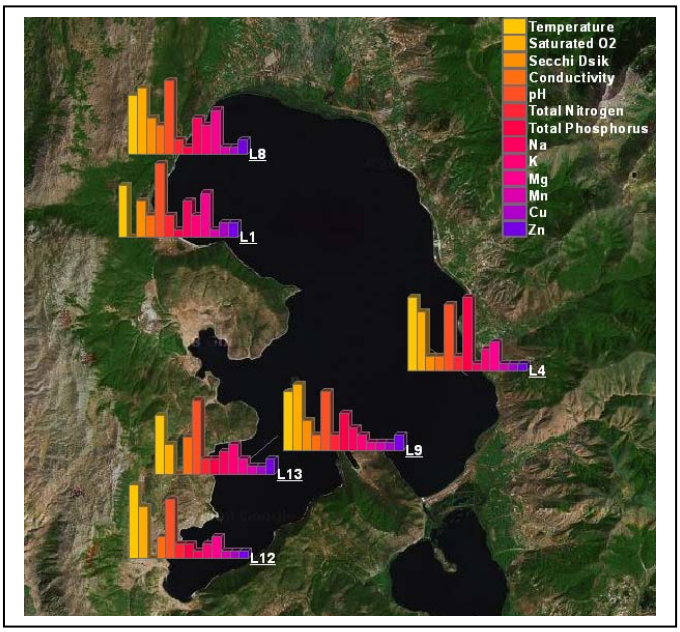


Fig. 1. GIS fuzzy model for the COCE diatom been a *mesotrophic* indicator.

As stated by the model presented on Fig. 1, the COCE diatom as *mesotrophic* indicator classified by the model, can be relatively good indicator for low level concentration of metals in all locations. There is an exception for location L_{13}, where high concentration of Total Phosphorus can be indicated. The pH value is relatively high of 7.6, except for the location L_9. According to the model, the eutrophication parameters stay relatively low along all measurement stations. In this direction for each class of the trophic quality eutrophication index, a geographical model regarding the physico-chemical parameters can be made. Another diatom considered by the model in Table 1 is the STPNN diatom. The GIS model is presented on Fig. 2. According to the model given in Table III, the STPNN diatom is an *oligotrophic* indicator. Based on the map given in Fig. 2, the model shows that the Saturated Oxygen concentration for L_4 and L_8 are very low, compared with the rest of the measurement stations, also the concentration of metals is low for L_3 and L_5 etc. In this way, for all physico-chemical parameters the threshold levels can be shown for each measurement station.

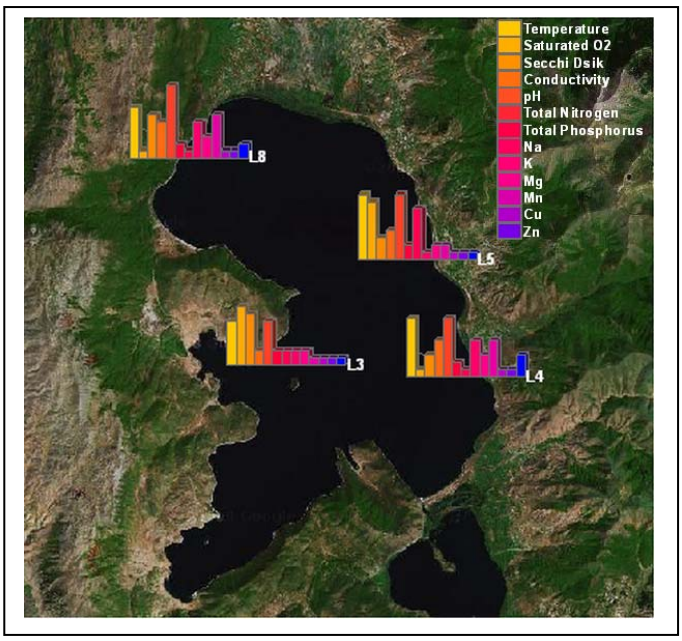


Fig. 2. GIS fuzzy model for the STPNN diatom been an *oligotrophic* indicator.

C. Verification of the Diatom Models

In this section, the results from the models are verified with the known ecological indicating references from the literature. The ecological references for the 10 most abundant diatoms are taken from the diatom ecology publication [20], that is used in several recently published papers [6, 7, 8, 9]. Some of the ecological references of the 10 most abundant diatoms in Lake Prespa, the CJUR and NPRE diatoms are newly described organisms with no records for their ecological preferences. Therefore, some of the results from the classification models are the first known ecological reference for certain trophic quality class based on nitrogen.

If we compare what is known for the APED diatom in the relevant literature [20], according to the models this diatom is *mesotrophic* and bad indicator of *oligotrophic*, while in the know literature is *eutrophic* class. If we look at the CSCU diatom, from the known literature [20] and the models, we can agree that this diatom is *eutrophic* indicator but with small probability according to the model. On another hand, the COCE diatom according to the model is bad indicator for *hypertrophic* and then ends as *oligotrophic*, which definitely needs further investigation. Moreover, the know literature for COCE reveal that this diatom is *eutrophic* [20]. Based on the model results, the STPNN diatom based on the nitrogen classification is *eutrophic*, while it is a bad indicator of *oligotrophic* waters. The other ecological references for the rest of the diatoms are new and they have to be further investigated before any solid conclusion is made.

V. Conclusion

This paper presents the Naïve Bayes method for classification used to obtain knowledge for environmental measurement analysis based on a specific physico-chemical

parameter. The data contain information about the abundance of the diatoms and total nitrogen concentration. The obtained model is presented in a form of table and easily can be converted into rules and presented on a geographical map as it is done later. Because the indicating properties of the diatoms do not depend on the geographical location according to [22], the method and the model output can be used to reference other diatoms in other ecosystems. The estimated probability that is produced by the models can be used for advancing the algorithm and combining other techniques to increase the accuracy of the models. After this, the models are verified with the known ecological knowledge. The knowledge obtained from the models, that is not found in the relevant literature must be further investigated before any strong conclusion is made. The probability that is derived from the models can be used to express the membership of a certain diatom in a given class. For each of the defined TQC, the method has found a relationship between the diatoms and a given class with certain probability. Important factor of the models is the interpretability, where the results from the models were interpreted, represented in a form of rules and then inserted in decision model. Based on the results in this paper, we can conclude that the Naïve Bayes method can be a good algorithm for diatom classification.

The research study conducted for Lake Prespa showed that studies like this, which combines the ecological knowledge for the processes in the lake ecosystem based on machine learning methods, are necessary to provide understanding of the physical, chemical and the biological processes and their relationship with the living organisms. Encouraged by the results from the models, further research regarding the Naive Bayes method can be focused on different probabilistic functions instead of Gaussian probability distribution to better describe the large number of low abundance data or near zero data. Other methods could be also combined and used for diatom classification.

ACKNOWLEDGMENT

This work was partially financed by the Faculty of Computer Science and Engineering at the Ss. Cyril and Methodius University in Skopje, Republic of Macedonia.

REFERENCES

[1] S. R. Carpenter, D. L. Christensen, J. J. Cole, “Biological control of eutrophication in lakes,” *Environmental Science and Technology*, vol. 29, pp. 784–786, 1995.

[2] B. P. Bennett, “Assessing the calibration of Naïve Bayes’ posterior estimates, “ in *Technical Report* No. CMUCS00-155, 2000.

[3] S. Monti, G.F. Cooper, “A Bayesian network classifier that combines a finite mixture model and a Naïve Bayes model,” in *Proceedings of the 15th Conference on Uncertainty in Artificial Intelligence*, 1996 Morgan Kaufmann, pp. 447-456.

[4] P. Langley, W. Iba, K. Thompson, “An analysis of Bayesian classifiers”, Proceedings of the Tenth National Conference on Artificial Intelligence, 1992, San Jose, CA, AAAI Press, pp. 399-406.

[5] E. F. Stoermer and J. P. Smol, *The diatoms: Applications for the Environmental and Earth Sciences, 2nd Edition*. Cambridge, UK, Cambridge University Press, ISBN-13: 978-0521509961, pp. 42-46, 2010.

[6] A. Naumoski, D. Kocev, N. Atanasova, K. Mitreski, S. Krtić, S. Džeroski, “Predicting chemical parameters of water quality form diatoms abundance in lake Prespa and its tributaries,” in *Proceddings of the 4th International ICSC Symposium on Information Technologies in Environmental Engineering - ITEE 2009*. Springer Berlin Heidelberg press, Thessaloniki, Greece, 2009, pp. 264-277.

[7] A. Naumoski, K. Mitreski, "Pattern tree spatial models for ecological classification," in *Proceedings of the 9th International Conference for Informatics and Information Technology (CIIT 2012)*. Bitola, Macedonia, 2012.

[8] A. Naumoski, S. Krstic, K Mitreski, "Novel Inverse Sigmoid Fuzzy Approach for Water Quality Diatom Classification," in *Proceedings of ICT Innovations 2011, AISC 150, L.* Kocarev (Eds.), Springer-Verlag Berlin Heidelberg, 2012, pp. 207-217.

[9] D. Kocev, A. Naumoski, K. Mitreski, S. Krstić, S. Džeroski, “Learning habitat models for the diatom community in Lake Prespa,” *Ecological Modelling*, vol. 221, no. 2. pp. 330-337, 2010.

[10] J. Quinlan, “C4.5: Programs for Machine Learning,” Morgan Kaufmann: San Mateo, CA, 1993.

[11] P. Langley, W. Iba, K. Thomas, “An analysis of Bayesian classifiers,” in *Proceedings of the Tenth National Conference of Artificial Intelligence*. AAAI Press, 1992, pp. 223-228.

[12] I. Kononenko, “Comparison of inductive and naïve Bayesian learning approaches to automatic knowledge acquisition,” In *Wielinga, B., ed., Current Trends in Knowledge Acquisition.* IOS Press, 1990.

[13] M. J. Pazzani, “Search for dependencies in Bayesian classifiers,” in Fisher, D., and Lenz, H. J., ed. *Learning from Data: Artificial Intelligence and* Statistics. Springer Verlag, 1996.

[14] P. Domingos, M. Pazzani, “Beyond independence: Conditions for the optimality of the simple Bayesian classifier,” *Machine Learning*, vol. 29, pp. 103-130, 1997.

[15] J. Friedman, “On bias, variance, 0/1-loss, and the curse of dimensionality,” *Data Mining and Knowledge Discovery*, vol. 1, 1996.

[16] WEKA 3.6.2 - Machine Learning Toolkit - http://www.cs.waikato.ac.nz/ml/weka/, accessed on 13 January 2013.

[17] S. Krstič, “Description of sampling sites, “FP6-project TRABOREMA: Deliverable 2.2, 2005.

[18] TRABOREMA Project WP3, *EC FP6-INCO project no. INCO-CT-2004-509177*, 2005-2007.

[19] R. E. Carlson, J. Simpson, "A Coordinator's Guide to Volunteer Lake Monitoring Methods," *North American Lake Management Society*, vol. 96, 1996.

[20]. H. Van Dam, A. Martens, J. Sinkeldam, “A coded checklist and ecological indicator values of freshwater diatoms from the Netherlands,” *Netherlands Journal of Aquatic Ecology*, vol. 28, no. 1, 1994, pp. 117-133.

Consequences of Cyber Threats in the Digital Age

Zdravko Veljjanovski, Adrijan Božinovski, Ivana Stojanovska
School of Computer Science and Information Technologies
University American College
Skopje, Macedonia
zdravko.veljjanovski@gmail.com; bozinovski@uacs.edu.mk; ivana.stojanovska@uacs.edu.mk

Abstract**—**In the digital age of today, people are becoming excessively dependent on the Internet, which has surpassed all known information-offering sources, news and data, due to it being easily accessible. The Internet services offer a lot of benefits but they also pose challenges regarding user security. The study contained in this paper describes the most common and well-known cyber threats such as malicious software and threats in the daily use of IT technologies. The study also deals with dangers imposed by potential sites such as free Wi-Fi hotspots and public USB charging stations, where personal information theft could happen. Sharing personal information on social networks or logging into fake web pages can also be abused in social engineering which is a non-technical type of intrusion. Particular attention is given to the threats in e-payment and ransomware. Lastly, this study also elaborates the methods for alleviation of cyber threats.

Keywords— Cyber threats; cyber attacks; malicious software; detection; prevention; social engineering; phishing; ransomware.

I. INTRODUCTION

Cybersecurity as a field of science plays an important role in the understanding, development and practice of cyber threats protection. Cybersecurity consists of many categories covering the technologies and practices that are mostly used for the protection of communication networks, IT equipment and various types of data from being misused [1]. Resources of IT systems are protected by several types of security mechanisms that work towards timely containment of any type of malicious software that could lead to malfunction. The focus of most of the security mechanisms is to protect the operating system, while great attention is also given to data protection [2].

With the growing popularity of Internet services, it appears that the IT communication devices, i.e. phones and tablets, have become the device of choice if one wants to enjoy the benefits and advantages of the Internet [3]. Most web sites and services are developed by large corporations striving to increase the use of the Internet by creating services and websites in order to facilitate everyday activities. Over the last year, information technologies have become more tightly intertwined in everyday life. The availability of many services on mobile devices makes everyone rely on the new roles of the cloud services for storing personal and business data. This may be considered as an advantage for increasing the productivity, but at the same time, it makes everyone more vulnerable [3].

The end users are not sufficiently familiar with the protection against cyber threats present on the Internet and this results in theft of personal information; this is the most common and ultimate goal of infection with malware. End users do not see the full infrastructure required for providing a complete service and the security on the Internet is now at the forefront of computer network related issues [4].

The reality is that anyone who uses IT equipment and can connect to the Internet is likely to get infected by some form of a computer virus, which is a quickly growing problem. Even companies that are working on protection against viruses have got their systems infected during the tests [5].

Having in mind the possible cyber threats in this digital age it is of enormous importance to know the consequences and the methods for protection – the subject of this paper structured as follows: in Section II there is an overview of the development of Internet services, covering how the mobile Internet changed the old ways of connecting to the Internet. Section III describes the most common malwares and the protection against them. Section IV deals with the dangers of WI-FI hotspots and USB public charging stations and recommends certain protection while using them. Section V is about social engineering while Section VI is about e-banking and how it changed the old ways of using banking services. Section VII describes the level of damage ransonware can cause and Section VIII is the conclusion.

II. TRENDS OF INTERNET SERVICES USE

In 1995 the number of Internet users was less than 1%. From 1999 to 2013 the number increased tenfold and in 2005 it had reached the first million. In 2010 this number reached two billion. The highest growth was recorded in 2014: three billion users, figure 1 [6].

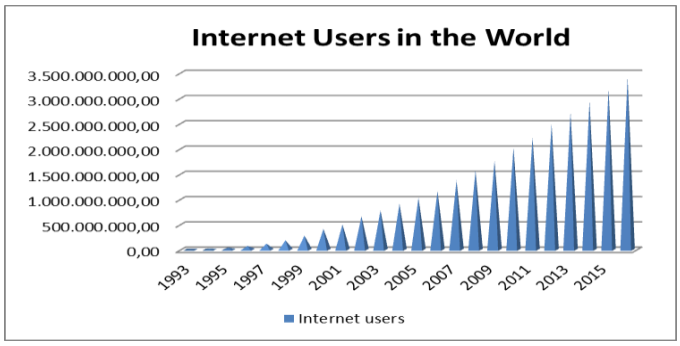


Figure 1. Internet Users in the World (recreated by the authors)

Mobile technology was introduced in 1980 and quickly became a substitute for the traditional telephone service worldwide. The introduction of Internet access with traditional 2G mobile technology in mid-2000 and the improvement to the current 3G and 4G technology offers faster access and thus higher speed of downloading data. The number of users is constantly growing due to the easy Internet access, which has wider use because it is available almost everywhere in the world. This is expected to have a major role in bringing the next billion people online. Mobile phone services are the reason that more than 90% of the global population uses mobile phones. However, the opportunity to have full-time access to the Internet on mobile phones does not make other IT equipment for Internet access redundant; it merely means that it is used for more demanding tasks, figure 2 [7].

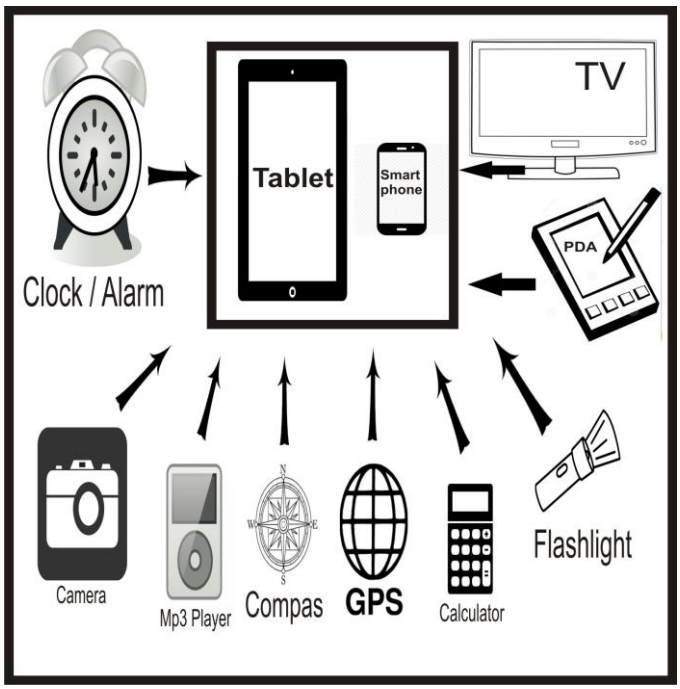


Figure 2. The sublimation of multiple devices caused by the technology development

As the number of Internet users has been growing, a new type of crime in the form of cyber crime has emerged, continuously finding new ways for abuse of Internet services and IT technologies. The development of cyber threats grows with a frightening pace on a daily basis [8].

III. MALICIOUS SOFTWARE (MALWARE)

This section explains what actually is the most common malware and how to protect from it.

1. Malicious code or rogue programs or malware (short for malicious software) is a software or a program that is intended first to infect the system then do the actions that will have unexpected or undesirable effects [9].
2. The computer virus is embedded in some legitimate source software that aims to reproduce by infecting other programs and thereby cause harm [2]. The reason this was given the name “virus” is because it behaves like a biological virus. There are two types of viruses: the first one is executed once and the other one is performed several times and then it is active in the background [9].
3. The worm is a malicious type of process that uses mechanisms to duplicate or make copies of itself. The worm multiplies by finding weaknesses in computer networks, but can multiply between systems thus doing great damage [2].
4. Software that poses as legitimate, while having embedded code for abuse, is called a Trojan horse. Also, spyware is a type of Trojan horse [2]. The name comes from a reference to the Trojan War [9].

The damage caused varies depending on the type of malware. In order to prevent damage it is necessary to undertake the following actions:

- The newest operating system as well as any program with the latest patches / updates shall be used.
- Emails from an unknown sender and especially those with attachments, shall not be opened.
- When searching the internet one shall be careful with the pages they open and the ads that show on the websites are not supposed to be open. Before installing anything, the content must be carefully checked.
- It is recommended to install anti-virus software preferably with a firewall that shall be set to scan and update on a daily basis. It can, in most cases, remove and prevent viruses, worms, Trojans, and (depending on software) some spyware [10].

IV. DANGERS IN PUBLIC WI-FI HOTSPOTS AND USB CHARGING STATIONS

People are using WiFi hotspots, mostly because they are available at little or no cost. It is a well known fact that WiFi hotspots are available in many public places, and they allow users to stay connected wherever they are, figure 3 [11].

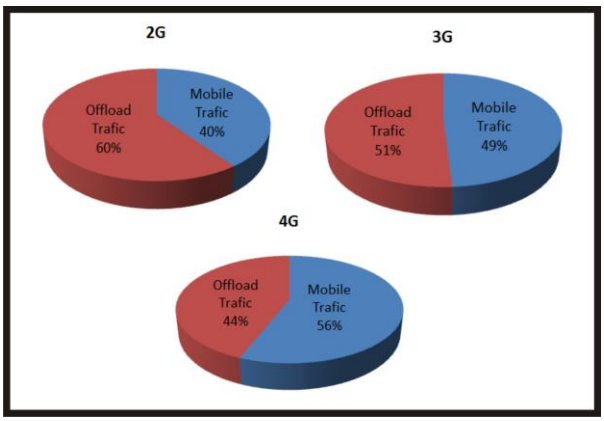


Figure 3. Mobile and Offload Traffic from Mobile-Connected Devices (recreated by the authors)

However, most WiFi hotspots are not safe enough because they are not encrypted and anyone with readily available software can see the Internet activity of users. Before one connects to a public WiFi hotspot it is necessary to ensure it is what it should be. When using public WiFi hotspots it is possible to be intercepted by announcements and therefore it is not recommended to use e-banking and other sensitive web services. However, most WiFi hotspots are not safe enough because they are not encrypted and anyone with readily available software can see the Internet activity of users. Antivirus solutions do not offer complete protection, because in order to be protected when connecting to a WiFi hotspot the users need to use a paid VPN service that will protect them from the possible consequences, figure 4 [11].

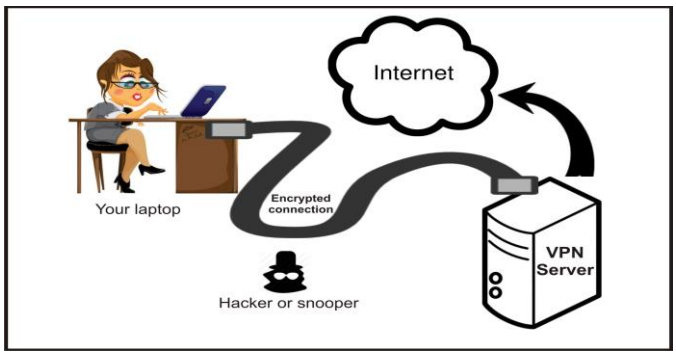


Figure 4. How a Personal VPN Works

Nowadays everyone relies on the smart devices regarding everyday tasks. This is due to the fact that the smart devices facilitate the everyday chores of the people and this leads to inevitable battery drainage and the need for charging is even greater. In order to satisfy this need, public charging stations for USB devices have been installed These stations offer free charging, but at the same time they increase the danger of personal data theft and malware infection. It is recommended that an USB cable be used with a charge option only and the device be shut down, in order to avoid these consequences. But there are also other methods as well, like charging the devices by an external battery and subsidiary or second battery which are far more secure [12].

V. SOCIAL ENGINEERING

Social engineering is when a user attempts to take an action that is seemingly safe, but it is exposed to possible abuse. There are many ways to lead users to be abused as shown in figure 5 and there are a few as well, that can be taken in order to protect against them:

- The best-known way is by Phishing. There are two most popular ways: through emails and by creating fake Internet sites. Protection regarding this is to read the email messages carefully and not to give feedback and not to open attachments unless the sender is known. Web sites before signing especially e-banking to see if the link corresponding to the real and login of any other use where username and password. It can also help install an add-in web browser [13].
- Unlike phishing emails, pretexting is aimed to create a false sense of confidence among the users. This is done by creating a credible story that leaves little room for doubt. For protection it is proposed to examine in detail what is required to be done and whether the one who sends the message really needs those data [13].
- Baiting is a direct attack which includes direct access methods: USB flash drives and DVDs that contain malicious files embedded with malicious code. In order to enhance the protection one must consider well before acepting and using something [13].

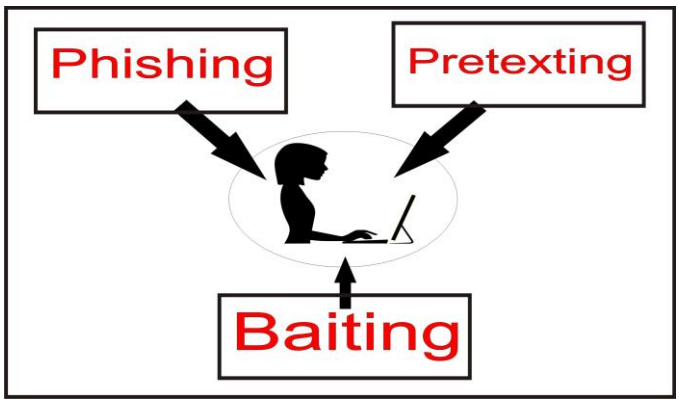


Figure 5. Types of Social engineering

VI. E-BANKING

Electronic banking is a trend aimed at changing the current way of performing banking-related activities. This method offers enormous benefits to users thus reducing waiting time at the counter compared to performing transactions online. But this brings new challenges for consumers and banks in designing a system for secure banking [14].

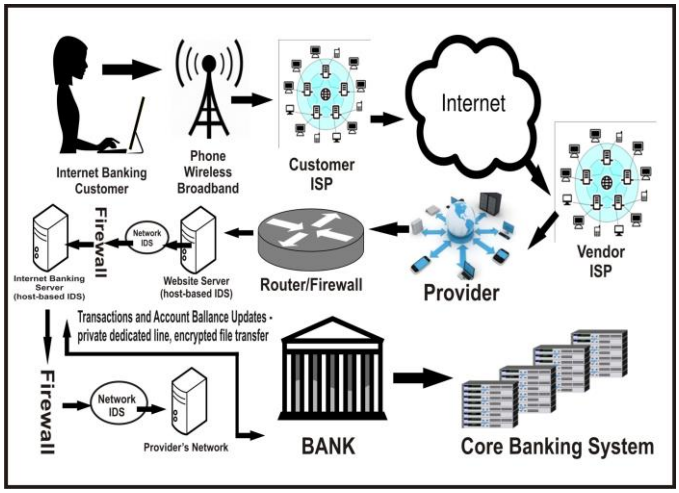


Figure 6. Diagram that illustrates the transaction flow

Figure 6 shows the flow of transaction from connecting to e-bank service to information.

- It starts with a connection to the Internet (via mobile Internet, wireless or broadband) and login to e-Bank service in order to execute transactions.
- An announcement is made on user authentication.
- After a successful login, Web Start out with various kinds of e-banking transactions are performed through a firewall to a web service banking server.
- Before each transaction a running check is performed, the default action depending on what is chosen to be executed.
- To perform any type of transaction, it is necessary to inject code from an OTP token or confirm using a digital certificate.
- As a bank doing security checks in order to protect consumers and, they should also be careful before they sign in to the E-Bank.
- It is always recommended to use only a secure network and in any case not a free WI-FI network.
- It is recommended not to use e-Bank via IT equipment used by many users because of the great possibility of abuse.
- User log in, OTP token and digital certificates should not be given to anyone under any circumstances [15].

As mobile technology and software solutions become more popular, so does mobile banking. This solution offers the same advantages as the computer. In order to be used safely it is necessary to take the following actions:

- To ensure the mobile device with a PIN or pattern.
- It is recommended to use secure communication networks, for example: local or network protected by a password, but not the free WI-FI networks.
- The users must ensure themselves that the mobile device does not have malicious software that can steal user login [16].

VII. RANSOMWARE

With the rapid growth of software technologies more and more people around the world are subjected to massive scale fraud and this trend reached its highest growth in 2015. [18]. The use of IT equipment to perform daily tasks and request information online has never been greater, and this leads to a growing number of cyber criminals who exploit sufficiently knowledgeable users about current cyber threats with an increased number of malicious software. The greatest threat today is ransomware - one of the numerous dangers of the Internet, which is currently among the most dangerous [17].

To date there are two types of ransomware, figure 7. The first one encrypts personal data and files and the other is designed to lock the computer to prevent its use [17].

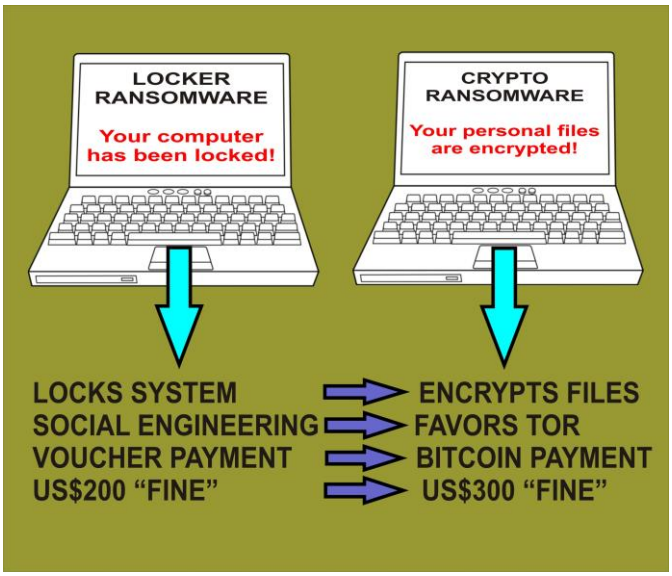


Figure. 7. Two main types of ransomware

In recent years the development of ransomware reached its growth in the form of money payment threats / extortion. Figure 8 shows the growth in the period between 2005 to 2015. Each year there is a new version of misleading apps, fake AV, locker ransomware and crypto ransomware [17].

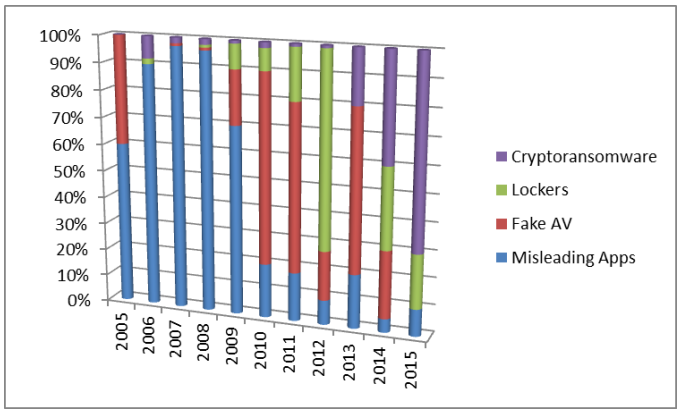


Figure 8. Growth of ransomware in the period between 2005 to 2015 (recreated by the authors)

Maybe ransomware is the most effective against home users who are not apt enough in working with computers or are not familiar with ransomware and the damage it can inflict. The lack of information and not knowing someone who can help bring a sense of isolation and helplessness of the user thus further increasing the pressure to pay. Often the systems store sensitive information, files and documents that are personally valuable to their users, such as college projects, photos and videos. Despite the fact that the data is of value to its users, it is unlikely that they have an effective back-up strategy in place to successfully recover from critical events such as fire or theft, or even an attack from ransomware. Even if some users had some form of backup strategy, there is a great possibility for infections to occur, which means that their data are to be destroyed [17].

An attack on the business or the public sector may be very harmful because they keep vital information about their operations. The worst thing that could happen is to have the data servers damaged, in which case the damage would be much greater. It is therefore recommended to be more cautious before taking any action or to consult technical support [17].

This type of malware never ceases to grow and it has only one purpose: to infect more IT devices, not just PCs, laptops, tablets and mobile phones i.e. all devices having any operation system, figure 8 [17].

VIII. CONCLUSION

Internet services and IT devices are being intensely and rapidly developed. There is great difficulty in keeping track of the progress which is the biggest reason for the lack of knowledge about the dangers behind cyber threats. Everyday, more and more new types of malware emerge and their goal is not only to do harm, but also to make profit. In the near future cyber threats will be present with even greater intensity since the number of Internet users grows by the day – the biggest reason for this threat.

REFERENCES:

[1] Josiah Dykstra.," *Essential Cybersecurity Science: Build, Test, and Evaluate Secure Systems"*, O'Reilly, 2015.

[2] Abraham Silberschatz, Greg Gagne, and Peter Baer Galvin.,*"Operating System Concepts 9th Edition"*, Wiley, 2012

[3] Wenke Lee, Bo Rotoloni, *"EMERGING CYBER THREATS REPORT 2015"*, Geogia Tech Cyber Security Summit, 2014

[4] Satish Kumar Thalod andRam Niwas "SECURITY MODEL FOR COMPUTER NETWORK BASED ON CLUSTER COMPUTING", *[International Journal of Engineering and Computer Science ISSN:2319-7242 Volume 2 Issue 6 June, 2013 P. 1920-1927] available from* http://www.ijecs.in/issue/v2-i6/31%20ijecs.pdf *[last accessed on April 27, 2015]*

[5] Peter Szor , *"The Art of Computer Virus Research and Defense"*, Addison Wesley Professional, 2005

[6] "Number of Internet users", available from http://www.internetlivestats.com/internet-users/ [last accessed on March 27, 2016]

[7] "Internet Society Global Internet Report 2015 Mobile Evolution And Development Of The Internet", available from https://web.archive.org/web/20150709014701/http://www.internetsociety.org/globalinternetreport/assets/download/IS_web.pdf [last accessed on February 20, 2016]

[8] James Lyne, "Security Threat Trends 2015", available from https://www.sophos.com/en-us/threat-center/medialibrary/PDFs/other/sophos-trends-and-predictions-2015.pdf?cmp=70130000001xKqzAAE [last accessed on February 20, 2016]

[9] Charles P. Pfleeger, Jonathan Margulies, Shari Lawrence Pfleeger, *"Security in Computing, 5th Edition"*, Prentice Hall, 2015

[10] *"How to Protect Against Malicious Software"*, available from http://www.seas.ucla.edu/security/malware.html [last accessed on February 17, 2016]

[11] "The Hidden Dangerous of Public WiFi", available from http://www.privatewifi.com/wp-content/uploads/2015/01/PWF_whitepaper_v6.pdf [last accessed on March 28, 2016]

[12] "Charging Your Smartphone…What Can Possibly Go Wrong", available from https://blog.kaspersky.com/charging-your-smartphone/1793/ [last accessed on February 20, 2016]

[13] Christopher Hadnagy, Paul Wilson, "*Social Engineering: The Art of Human Hacking"*, Wiley, 2010

[14] "Challenges of the "E-Banking Revolution", available from http://www.imf.org/external/pubs/ft/fandd/2002/09/nsouli.htm [last accessed on February 15, 2015]

[15] "E-Banking Components", availble from http://ithandbook.ffiec.gov/it-booklets/e-banking/introduction/e-banking-components.aspx [last accessed on September 5, 2015]

[16] "M-banking Safety Recommendations for clients", available from http://www.stb.com.mk/content/Pdf/M-banking%20Safety%20Recommendations%20for%20clients%20-%202014.pdf [last accessed on July 12, 2014]

[17] Kevin Savage, Peter Coogan, and Hon Lau, *"The evolution of ransomware"*, available from http://www.symantec.com/content/en/us/enterprise/media/security_response/whitepapers/the-evolution-of-ransomware.pdf [last accessed on February 20, 2016]

Hotel and ATM spatial distribution in Skopje using ArcGIS

Katarina Krstevska, Simona Stojanoska, Kire Timov, Andreja Naumoski, Kosta Mitreski
Faculty of Computer Science and Engineering
Ss. Cyril and Methodius University in Skopje
Skopje, Republic of Macedonia
{andreja.naumoski, kosta.mitreski}@finki.ukim.mk

***Abstract* – Integration of several everyday services is important step towards building modern informatics society. In this direction, tool like ArcGIS software serves as a good platform on which these services can thrive. Therefore, in this paper we present two case studies related to two such interconnected services in ArcGIS. The two cases are built on information gathered in the city of Skopje. The first service is related to the hotel spatial distribution in the Skopje centre and its surroundings. It offers a geographical information concerning the hotel start rating, address, telephone, e-mail or webpage. All this information is useful for any tourist or guest to have. The hotels are classified as: hotel, hostel and apartments. There is also a bookmark option for easy navigation on the map. The second case study is related to the location of the ATMs for a particular bank in Macedonia. However, this can be extended with other ATMs from different banks. This is a mobile application based on Android platform that shows the location and it has the ability to search the nearest ATM for a given bank. Our future work, will be focused on integration of more services and providing a mobile and web service for all these services for everyday users.**

Keywords— Hotel services, ATM services, Skopje, ArcGIS, spatial distribution

I. INTRODUCTION

Geographic information system (GIS) is a software of various tools that are used to manage spatial relations across space and integrate different types of information. This software package provides practical tools for visualizing and analysing data like hotel information or bank information, revealing trends, dependencies and inter-relationships. The produced information can be used to learn about a geographical area, to manage a hotel location information, location of everyday services like ATMs, choose an ideal site for ATM deployment or choose a delivery route for faster and best available service. ArcGIS can acquire, store, manage, and geographically integrate large amounts of information from different sources of data and other programs from different sectors. Therefore, in this direction, tourist services such as hotels can be mapped in a relation with the surrounding environment and provide important analysis in a relationship with monuments, public events, historical building etc. In this way integrated information can provide more advanced analysis or heat maps of best possible location for such hotels. The same methodology can be applied in other sectors, such as financial services from the bank with the geographical location of ATMs. This is even more powerful when GIS is integrated with mobile platforms that serve as an initial ground to provide mobility of information for the user.

GIS has become active platform on which users can represent various information regarding the public and private sector services in a given spatial area. The research in the relevant literature shows numerous and important case studies about the applicability of GIS regarding ATMs public services [1, 2, 3, 4]. Different strategies for optimal deployment of ATMs are discussed and analysed in several papers [3, 4]. In these research papers, different analyses are conducted through the ATM location in different countries and areas. On another hand, the hotel analysis for different application is no different and have implication in various economic and social disciplines. For example, in [5] authors conducted analysis for hotel room estimates and price reduction, while in [6] the authors inspect the influence of hotel distribution in retail tourism. Also, researchers take different spatial data to analyse the hotel spatial distribution [7]. There also web based decision support systems based on the geographic analysis to represent not only the data, but to further improve the functionality of such help tool for decision makers [8]. This type of applications is wide and broad, and can be very informative for the general public.

The aim of this paper is to present two case studies of integrated service for everyday users. The first service is useful not only for the citizens that live in Skopje, or Republic of Macedonia, but also for tourists that are visiting the city. The map contains information regarding the hotels in Skopje and the neighbouring area. Each hotel is georeferenced together with the information regarding his address, number of stars, classification and other types of information. Moreover, each hotel is bookmarked for faster search results. The list of hotels is obtained using search engine, together with the information of the longitude and latitude of the hotel. The longitude and latitude information are placed in georeferenced database and it is used to build GIS layers, which is a major step towards building a decision making system for general tourist information. The second case study is related to the mapping of the ATMs of particular bank in Skopje and also other cities in Republic of Macedonia. The data used in this project were collected from the web site of that particular bank. Another important step of the integration is that this ArcGIS map is integrated with mobile version, which uses http protocol and is presented on google maps to represent the data.

The structure of the paper is organized as follows: Section 2 provides information about the ArcMap module in GIS. In Section 3 we present the case study of the hotel spatial distribution, while Section 4 shows the procedure and the results from the ATM mobile application. Section 5 concludes the paper and gives further directions on how we plan to develop our ArcGIS decision support system in the future.

II. ARCMAP

ArcMap is one of the several important modules of the ArcGIS Desktop package. ArcMap is module where user can present and explore georeferenced data. On the map, the user can place symbols and can use the software to create a map for publication. This part of the ArcGIS is also used to create and change the data. The user builds the map in ArcMap by representing the geographic information as a combination of layers and user inputted elements of the map. After building the map, the user can perform many common GIS tasks. The GIS layers are actually a mechanism that is used to represent geographic data in ArcMap. Each layer references a data set and specifies how that data set is represented using symbols and text labels. When the layer is created and added to the map it specifies the data set and a set of map symbols. Layers are one of the primary ways to work with geographical data in ArcGIS.

In this paper the ArcGIS module is used to show the spatial distribution of the objects of interest. We used real data of the longitude and latitude of the specific objects. This data for the hotels were obtained in various ways; through the internet using method of aggregation or by searching the provider of internet (particular bank of given institutions), while the data for the ATMs were obtained from a particular bank web site. Since the software has various ways of showing the object, using different symbols or colours, we have used the most intuitive symbols in order to get more informative look of the map.

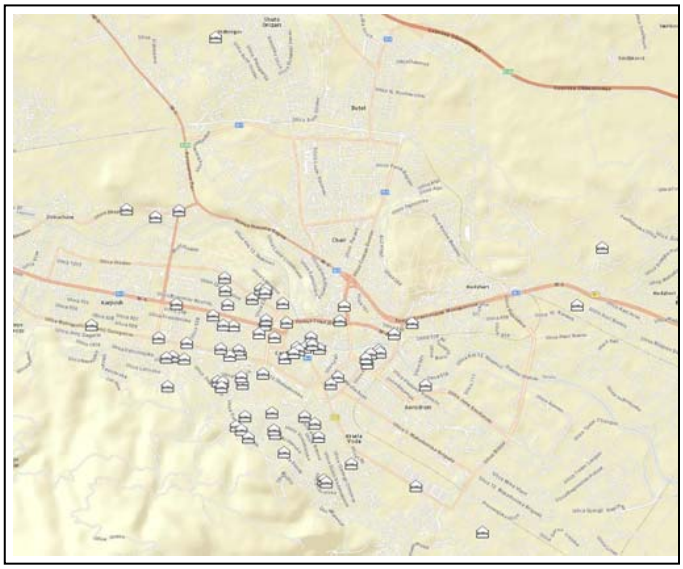


Fig. 1. ArcMap representation of the hotels' spatial distribution in Skopje.

Additionally, for each object we provide as much information as we can find, and the user can obtain this information simply by selecting the object. On Fig. 1, we present the spatial distribution of the hotels in Skopje. Hotels are not only located near the city centre, but also there are hotels that are located around the city centre and periphery. The presentation of the above features combines a multi-layered map in ArcMap. Moreover, the ArcGIS can display different layers that can be combined with each other to provide a clear representation of what is of interest to the user.

III. HOTEL SPATIAL DISTRIBUTION

In order to build the hotels' spatial distribution map in Skopje we have taken a geographical map of the capital city from online map ArcGIS repository.

The next step is creation of the georeferenced database, which is located in the ArcCatalog module of the ArcGIS software. Then we create a Personal Geodatabase, and for each hotel a particular Feature class. With this step activated, the process of entering the data begins. For each hotel we enter the name of the hotel in both Latin and Cyrillic alphabet, star rating of the hotel, the address of the hotel, telephone, e-mail for contact, webpage, picture of the hotel or possible pictures from the rooms and the type of the hotel. On Fig. 2 we show one example of the many hotels represented on the geographical map.

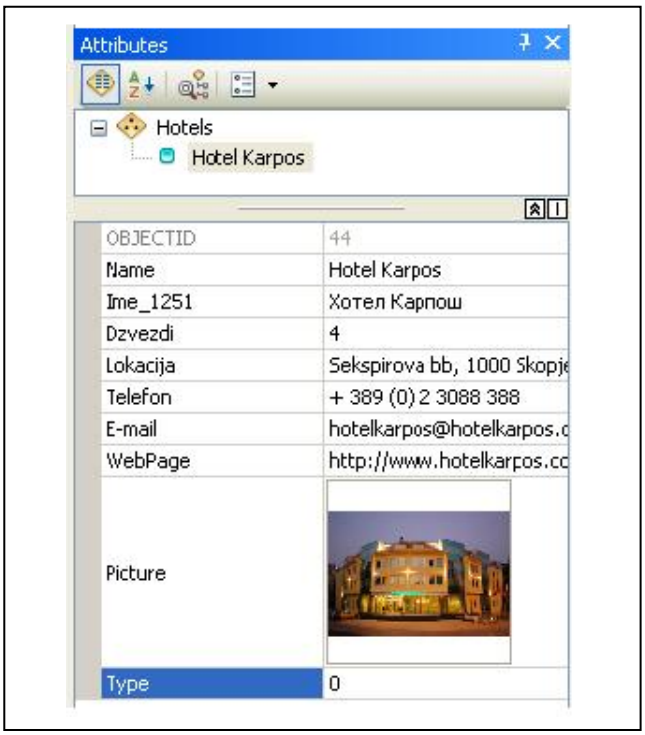


Fig. 2. Example of one hotel attribute.

Type 0 means that this is a hotel, type 1 means it is a hostel, and type 2 means that the building is apartment. Additionally, we selected the symbol of each hotel on the map. That is done in the menu Layers properties and then choosing Symbology

tab. Here we choose the symbol and change the colour or the size. Also, different symbols for hotel can be selected using the Symbol Selector tool, or provide other symbols through file.

Fig. 3. Adding hotel bookmarks on the map.

After we finish the entering this information, we start the digitalization process. For this purpose, we activated the Editor Toolbar and Start Editing. After this we can start entering the hotels on the map. If we want to edit the attributes of certain hotel, just right mouse clicks and select the attribute menu that will open a database like table where the user can enter different information. When we finish entering the hotels on the map, we produce a geographical map that look like the map presented on Fig. 3. As we mentioned earlier, for better and reliable search we use bookmarks for each hotel, in this way the user can have easy access to the hotel information.

IV. ATM Location Mobile Application

The main purpose of the ATM mobile application is geographical coding of the distribution of the ATMs of a particular bank in Macedonia using ArcGIS and Aptana.

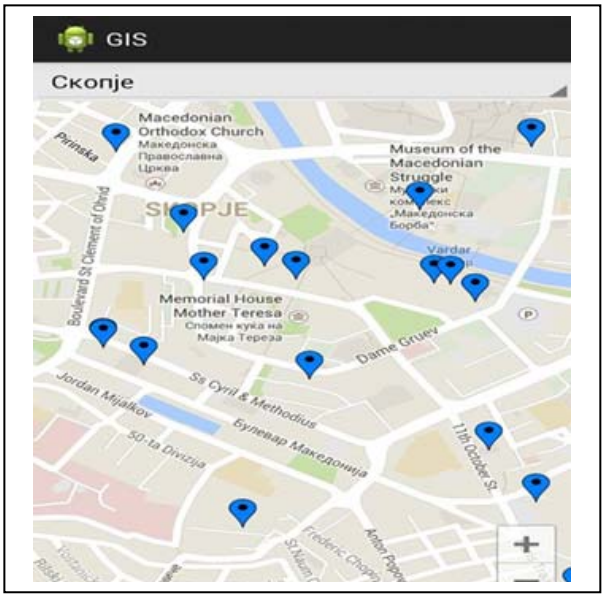


Fig. 4. Mobile spatial location of ATMs in Skopje.

After we have entered the data using the ArcCatalog module in ArcGIS software and generated georeferenced database, we continue by presenting the ATMs location on the mobile application.

The mobile application offers, for now, two options; searching the ATMs by city in Republic of Macedonia and searching the ATMs by location near me. The option to search ATMs near me, uses the Internet GPS location or this can be done by using the built-in GPS device. It depends from the user preferences and the availability of the GPS service. The example of one such search is given on Fig. 5.

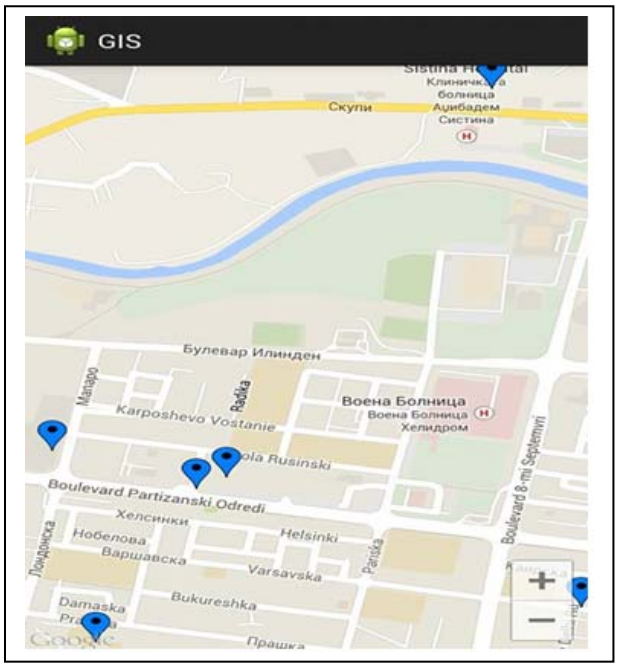


Fig. 5. Searching ATMs by Location near me.

The application has a limit radius of 3 kilometres, in which if there is no ATM, no object will be shown on the map. Otherwise, if we select: by searching the ATMs by city, the application asks you which city do you want to search. After that, the particular city is shown and the location of the ATMs are presented on the map.

V. Conclusion

In this paper we presented two examples of what can be interconnected applications in ArcGIS. First, we use the ArcGIS map to show the existing hotels in Skopje and then a particular example of mobile application to further improve the visual representation and mobile availability of the available data. The entire process, starting with the data collection through the process of making the maps and later through data geo-processing with mobile application is presented in this paper. The GIS application provided four groups of aptitudes to work with geo-referenced; data input, management, manipulation and data analysis. Once the data in imported in ArcGIS, particularly the data related to the ATMs locations, then it can be presented in various ways. In this paper we presented two ways to show the results from ArcGIS; on map

and on mobile device. The presented maps in this paper included several important information for both hotel and ATMs locations. These maps can be used as decision support tool in every aspect of the informative tourist system as an analytical tool and based on the map provided in this paper. This process can further improve the understanding of the hotel/tourist process, and to explore the relationships between cities, rural areas, population and the reservation in those hotels.

In conclusion, our findings demonstrate the obvious connection between the hotels and the financial service that are interconnected and they can be used to further improve the tourist industry not just in the main capital – Skopje, but also in other cities in Republic of Macedonia. Nevertheless, as the government rapidly invest in the tourist industry, this should eliminate any gaps between these industries and provide a bridge between other industries using the ArcGIS software. In this direction the maps presented in this paper can be a useful tool in the process of carefully planning the development. Therefore, this study highlights the ongoing need for monitoring the distribution of the hotels or ATMs or any other important institutions, to ensure they match the evolving socio-economic development.

Further improvement of this system can be done by implementing methods and techniques to analyse other type of data. Such systems could use data from the social network sites [8] and then develop system that interconnects the existing knowledge together with the services that this industry provides. An interconnected system that will be updated on daily level and stream the map information to every day users is our major future focus. This couldn't be done better than using the latest smartphone technology. In this direction our major future research is directed in developing interconnected smartphone applications.

ACKNOWLEDGMENT

This paper work was partially supported by the Faculty of Computer Science and Engineering, Ss. Cyril and Methodius University in Skopje, Republic of Macedonia.

VI. REFERENCES

[1] N. Al-Hanbali,"Building a Geospatial database and GIS data - Model integration for Banking: ATM site location," In Commission IV Joint Workshop: Data Integration and Digital Mapping Challenges in Geospatial Analysis, Integration and Visulization II, Stuttgart, Germany, September 8-9, 2003.

[2] B. Farhan, "A GIS-Based Approach for Evaluating ATM Partnership Opportunities," In Second National GIS Symposium, 2007.

[3] A. Alhaffa, and A. Wael, "A Market_Based study of optimal ATM's Deployment Strategy," international journal of machine learning and computing, Aprill (2011).

[4] S. Awaghade, D. Priyank, and R. Pinak, "Site Selection and closest faculity analysis for automated teller machine (ATM) centers: Case Study for Aundh (Pune), India," International Journal od advancmen in remote sensing, GIS, and Geography 2(1), pp. 19-29, 2014.

[5] S. Kisilevich, D. Keim, L. Rokach. "A GIS-based decision support system for hotel room rate estimation and temporal price prediction: The hotel brokers' context," Decision Support Systems 54, no. 2, pp. 1119-1133, 2014.

[6] J.C. Rachel, "Geographic information systems (GIS) applications in retail tourism and teaching curriculum," Journal of Retailing and Consumer Services 14, no. 4, 289-295, 2007.

[7] O.G. Omogunloye, O.O. Ayeni, "Analysis of hotels in lagos State with respect to other spatial data", Research Journal in Engineering and Applied Sciences 1(6), pp. 393-403, 2012.

[8] P. Fu and J. Sun, "GIS in the Web Era," in *In Web GIS: Principles and Applications*, Redlands, CA, USA, ESRI Press, 2011, pp. 1-24.

Multilingual dictionary of Slavic somatic phrasemes

Masha Bekjкovikj
Faculty of Computer Science and Engineering
University St Cyril and Methodius
Skopje, Macedonia
bekjkovikj.masha@
students.finki.ukim.mk

Elena Markova
Faculty of Computer Science and Engineering
University St Cyril and Methodius
Skopje, Macedonia
markova.elena@
students.finki.ukim.mk

Katerina Zdravkova
Faculty of Computer Science and Engineering
University St Cyril and Methodius
Skopje, Macedonia
katerina.zdravkova@
finki.ukim.mk

***Abstract*—Phrasemes are multi-word expressions and utterances with a distinctive syntax and semantics. Due to the ethnological diversity, they are language specific and in many occasions, the exact translation of a phraseme existing in one source language doesn't occur in other target languages, no matter their linguistic and cultural similarity. The presentation of source phrasemes, their meaning and examples in a form of an interactive and searchable multilingual dictionary is a good starting point for further linguistic research. This paper presents the creation of such a dictionary, its basic functionalities, the process of its development, and finally illustrates its performance according to assigned user roles. The paper ends with the further stages of its development.**

Keywords—somatic phrasemes, morphological search, stemming, search

I. Introduction

Phrasemes are linguistic signs consisting of several lexemes. They are represented as ordered triples: signified (the meaning of the phraseme), signifier (the phonetic form), and syntactic (the set of data referring to its co-occurrence with other signs) [1]. Similarly to multi-word expressions and multi-morphemic utterances, phrasemes are groups of lexemes with a meaning distinctive from the meaning of individual lexemes they are obtained from [2]. Therefore, apart from their purely syntactic description, phrasemes have a significant semantic distinction, making them a very fruitful field for linguistic research. The term phraseme usually encompasses the pragmatic and the semantic phrasemes, which are classified into pragmatemes, idioms, collocations and quasi-idioms [1]. Whenever among its constituent words, some body parts of internal organs are included, such linguistic signs become somatic [3]. They can be: somonymic (denoting parts of the human body), osteonymic (skeletal system), angionymic (circulatory system), scplanchonymic (internal organs), sensonymic (senses), and general body lexis (the body) [3].

Due to the ethnological diversity, phrasemes are language specific and in many occasions, the exact translation of a phraseme existing in one source language doesn't occur in other target languages, no matter their linguistic and cultural similarity. They have been extensively explored at bilingual [4], and multilingual level (https://syd.korpus.cz/). The best tool for aligning parallel corpora is undoubtedly GATES [5].

In spite of the variety of aligning tools and environments intended to present the multilingual corpora, none has tackled the aligning of phrasemes, mainly because the matching is affected by the cultural differences of the paralleled languages. In order to produce a correct idiomatic translation, it is necessary to delve deeper into culture, customs and beliefs of the region this expressions originate from. The linguistic and comparative analysis of the phrasemes is the main goal of a recent University project submitted by the Chair of Slavistics within the Faculty of Philology in Skopje. The project intends to contrast and present the behavior of the Slavic somatic phrasemes.

One of its crucial parts of the project is the visual presentation of parallel aligned corpora in a form of an interactive multilingual dictionary. It is based on the collected Macedonian source phrasemes, their meaning, which are enriched with the typical examples extracted from the literature. Each Macedonian phraseme is then associated with the corresponding target translations into Russian, Polish and Czech.

In this paper we present the development and the current stage of the multilingual phraseme dictionary. The structure of the paper is the following: in the Section II, use case analysis of the system is done, presenting the roles of the users and the major interactions they are authorized to perform. Section III presents the creation the dictionary. Search modules, which were one of our major obstacles are described in Section IV. Section V is a small user manual illustrated with the distinctive functionalities. Finally, the conclusions and further development of the system is presented.

II. Use case analysis

The multilingual dictionary is intended to be used by the linguist specialists, as well by users who are interested to research the phrasemes, their meaning and examples. Since most of them have only the basic computing skills, we decided to make the whole system Web based, exceptionally intuitive and user-friendly. In order to be able to create such system, we decided to build it using the open source content management system Drupal [6], which offered us the opportunity to create a powerful, modular, effective and reliable application. The application is hosted on a virtual machine and is running on Acquia Dev Desktop [7].

The system itself comprises four types of users:

- System administrators, authorized to: create the whole content in all the languages existing in the parallel corpora; approve the modifications initiated by language editors and registered users; enable the extension of the multilingual dictionary with new languages; assign the roles of the authorized users, which belong to next two types. This task will be maintained by the creators of the system.
- Language editors, which have an access to all the contents they are responsible to edit, including the right to add, modify and remove the existing phrasemes, their semantics, as well as their representative examples. This task is assigned to language specialists. In order to eliminate the risk of unintentional damage of the contents, modifications by the registered users will be temporarily stored, and become effective only after the approval by a language editor.
- Registered users capable of profound search of the phrasemes and their meaning. They can access the contents, search it using the keywords existing in the source language, extract the target phrasemes and download the final results. Registered users are also able to add or modify existing contents, but these actions are to be approved by the system administrator or language editors.
- Ordinary users have an authorization to search through the whole contents using the keywords existing in the source language and see the target translations.

The following two sections explain the implementation of the system, with particular emphasis to the realization of the search option, which was our main challenge.

III. Implementation of the System

We used a virtual machine with a 2.27 GHz QEMU Virtual CPU version processor, with installed memory (RAM) of 8 GB and 64-bit Windows Operating System. On this virtual machine we installed Acquia Dev Desktop to run our Drupal application on. Acquia Dev Desktop includes all the necessary elements for the needs of our application: Dev Desktop App - for managing the Drupal application; Apache web server [8]; Percona MySQL database server [9]; PHP - programming language that powers Drupal [10]; and finally, phpMyAdmin - for MySQL management and querying [11].

All these elements were properly installed and configured during the installation of Acquia Dev Desktop. On top of this combination of developing environments, we installed the Drupal application. Whilst setting and configuring the Drupal installation, all the parameters were set for the database name, users, passwords, the site name etc. For this application, we used Drupal version 7.42 although there is a newer 8.0.5 version. The reason we decided to use the older version is because all the modules that we wanted and needed to use are not yet implemented or they exist only in the beta version for Drupal 8.

We decided to start developing this application by first concentrating on the design. We chose a theme which is free and available for downloading, as a base for the design, that we thought was appropriate for this type of application, but modified it according to our needs.

Knowing that we will use multiple languages for the translations of the phrasemes, we defined a taxonomy – vocabulary called *Language* which contains the different languages as terms. This vocabulary can be edited and enriched with new languages at any time without altering any other part of the application.

To create the multilingual phraseme lists. We needed to define a custom content type with custom defined properties. We named this content type *Phraseme* and it is consisted of the following properties: a title which represents the meaning of the phraseme, the original Macedonian phraseme with an appropriate example of usage and a field collection, implemented using the Field Collection module, of all translations. Each translation contains the language of the corresponding phraseme (represented as a vocabulary term which we previously discussed), the corresponding phraseme in that language and an example of its usage in the given language.

To functionally test the accuracy and representation of the content type, we added a few entries of type *Phraseme*. Then we created a tabular view of all *Phraseme* entries. In this view we implemented sorting by column as an option of the table. Also, the five newest phrasemes are displayed as a block on every page. This functionality is implemented so that regular users can easily see if there are new phrasemes added. After we were ensured that this content type and the vocabulary are working as expected, we implemented the Search Option, explained in Section IV.

Defining the types of users and their roles and permissions was implemented last due to the fact that we wanted to be assured that every other functionality of the application works correctly. We defined four types of users as explained in Section II. The system administrator is assigned when installing the Drupal application. This type of user has full access to every aspect of the application, its structure, appearance, people, modules, configuration, content and reports. This user has access to all administer settings. The system administrator can add every type of content available or create new content types, while the other users can only modify or add entries of the content type *Phraseme* and the predefined content type Article (in case they needed to add any news or articles on the home page). Example of this action is shown on Figure 1, which is presented on the next page.

Drupal includes a so called Revisioning module, which is responsible for the configuration of workflows to create, moderate and publish content revisions. By using it we enabled every modification and addition of new content to be sent as a revision to the system administrator or language editor and be published after their approval. When making a revision there is an option to add a revision log message. This way it is easier to trace the changes and explain the motivation to the system administrators and language editors for making the change.

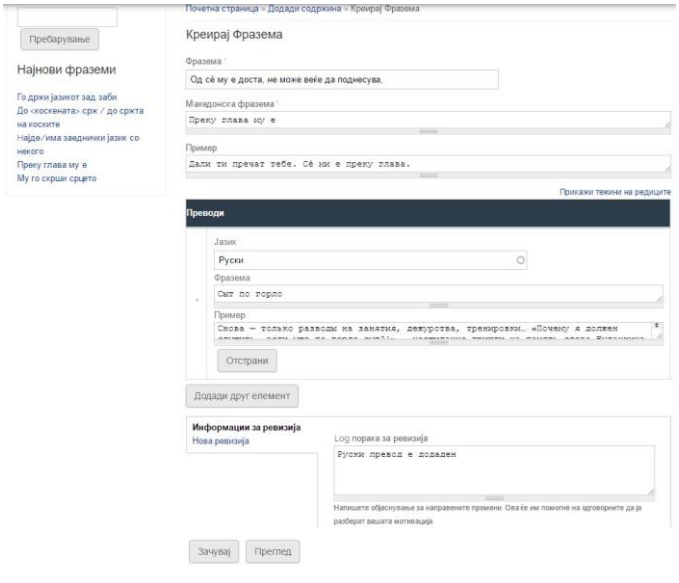


Fig. 1. Example of the action: Adding a new phraseme

The system administrator and the language editors have access to a page that displays revision summary for all content. Example of this page is shown on Figure 2. This way there is a better preview of all the revisions – archived, pending and currently published.

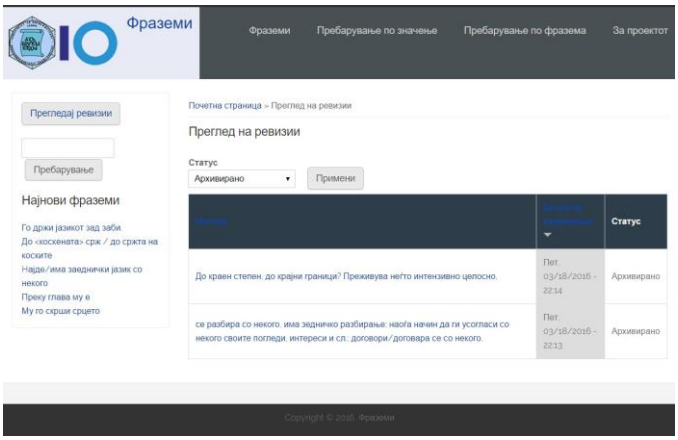


Fig. 2. Example of the Revision summary page.

Language editors have the permission to extend the corpus with new phrasemes, their explanation and the typical examples explaining the phraseme in Macedonian, together with the translations in all the languages existing in the system. Furthermore, they can add a completely new language and add the corresponding contents. That language becomes available for the phrasemes that were previously entered in the corpus.

To prevent the accidental removal of existing content, the language editor and registered user do not have the option to delete content, rather only to "unpublish" it. The system administrator has access to all content that has been unpublished and has the permission to delete or republish it. Unpublished content is then stored in the database if at any point this data is considered to be restored or published.

We used the Pathauto module to add URL aliases for the application. For example, for phraseme the URL alias is frazema/[node:nid]; for article it is article/[node:title]; for vocabulary terms it is [term:vocabulary]/[term:name] and for users it is users/[user:name].

At this stage of the project, the whole application is based on the dictionary where the Macedonian phrasemes are the core content, and consequently, the major users will be Macedonian native speakers. Therefore, we thought it is more natural and appropriate to have the whole navigation presented in Macedonian.

To enable this navigation, we localized the application by translating the complete Content Management System, its commands, actions, functionalities, buttons etc. using the Translate Interface, in Macedonian so the language editors, registered users and ordinary users would easily navigate on the application.

IV. Search Option

The search option, which is the main functionality for all types of users includes the possibility to perform the search using the exact match of the lemmas existing in the source phrasemes, their derivations, but also the parts of the words. This functionalities already exist for the English search. In order to enable it for the Macedonian, we used the electronic lexicon with more than 60K lemmas and their word forms [13].

We decided to install and configure the search server on the same virtual machine as the Drupal application because this way the application and the search server communicate locally thus providing better, faster and safer search on the site content. The search server we decided to use is Apache Solr Server 5 (http://lucene.apache.org/solr/). Drupal enables easy connection and communication with this server via modules like Apache Solr Search and Search API.

The Apache Solr Server is a standalone full-text search server powered by the Lucene search library at its core, which is a full-featured text search engine library written entirely in Java. By creating near real-time indexes for the content it is intended to search, Solr is able to achieve fast search responses. This means that Solr does not search in the content directly but in the indexes. Its comprehensive administration interface allows easy access and control over our Solr instance. On the other hand, Apache Server is more commonly used on a Unix-like systems, it was a challenge to configure and run it on a Windows Virtual Machine, also taking in account the fact that on top of an Apache Server configuration we had to configure Solr as well. We had to add Apache Ant in order to compile and run the Solr Server.

Apache Solr Server and Lucene are mainly built and have complete functionality for the English Language. It also has support for 32 languages, including English, but it does not provide support for search content in Macedonian. Due to the fact that both Apache Solr Server and Lucene library are open source, we were able to modify and configure the search to our needs and properly and correctly work for Macedonian language.

In our application, there are three types of searches implemented. They are briefly explained and illustrated below.

The first search is available as a block for every page and uses the Apache Solr Search module. It indexes and searches all the content of the application. This search is really useful if users want to find something fast or are not sure what exacatly are they searching for. Example of this search is shown on Figure 3. The other two searches are implemented with the Search API module because this module enabled us to create custom indexes which are used in custom views.

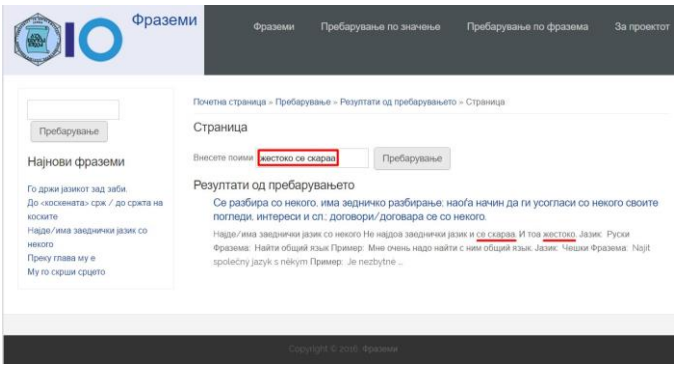


Fig. 3. Example of the search for all content.

The second search is for searching by the meaning of phrasemes therefore it indexes only the field meaning of the content type Phraseme. This will help users to explicitly search the meanings of phrasemes and will give them a clear view of the original phrasemes, their meaning and an example of usage. Example of this search is shown on Figure 4.

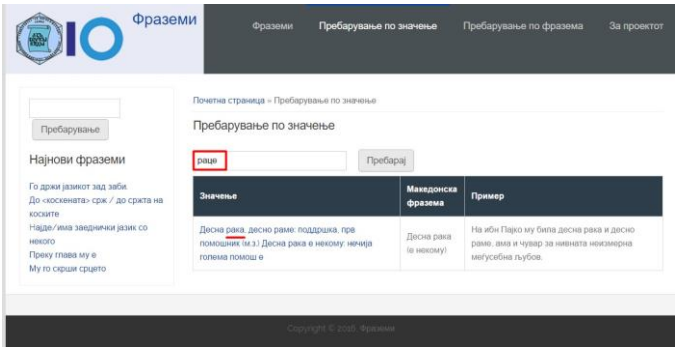


Fig. 4. Example of search by meaning.

The third search is intended for searching by the original Macedonian phraseme. Therefore it indexes only the field Macedonian phraseme of the content type *Phraseme*. This way users can search for concrete Macedonian phrasemes and similarly as the second search, it provides a tabular view of the original phrasemes, their meaning and an example of usage. Example of this search is shown on Figure 5.

The functionalities of the ordinary users and the registered ones are restricted to preview and search options only. They can go through all the existing contents, as presented on the Figures 3, 4 and 5. While the registered users have the permission to download and print the contents, these activities are disabled to ordinary users.

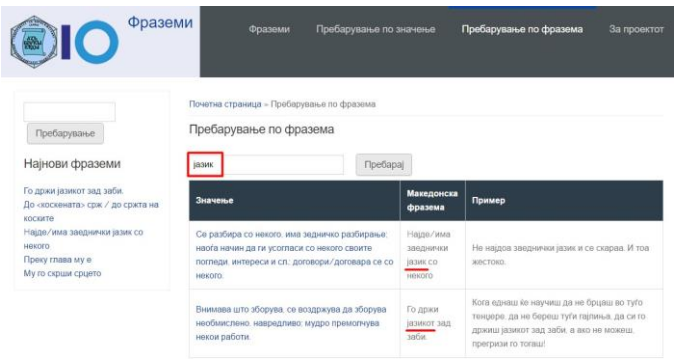


Fig 5. Example of search by Macedonian phrasemes.

V. Multilingual Dictionary at Work

The application is already stable, and it has been approved by the colleagues from the Faculty of Philology. After they test it more exhaustively, it will be available for the wider audience.

As visible from the figures presented in the previous two sections, the application has a very intuitive interface, so the users who are not very familiar with the technology can use it without any assistance.

The actions of the registered and ordinary users have been explained and illustrated in details so far. They can enjoy the wealth of the phrasemes, idioms, and phrases in many languages.

The language editors are currently exhaustively exploring the application. They have provided us with a very small pilot corpus intended to be used for the creation purposes. They have already started with the manual enlargement of the corpus with many new entries, and they also try to extend the multilingual aspect, wherever possible. Although their feedback is very favorable, we are very excited to see whether all of them will be able to use it.

We are steadily performing the system administrator's tasks by carefully observing their work and approving the modifications they produce. So far, the use of the application is very smooth.

VI. Conclusions and Further Extensions

Multilingual dictionary of Slavic somatic phrasemes is the first project that presents the collected Macedonian idiomatic expressions and compares them with the translation equivalents existing in Russian, Polish and Czech. It was tested and approved by the colleagues from the Faculty of Philology, who found the system intuitive and easy to use.

The system was created with an open source content management system, thus the contributed files, as well as our derived work are licensed under the GNU General Public License, version 2 [14] or later and copyrighted under the Creative Commons Attribution-ShareAlike 2.0 Generic [15]. However, the content of the multilingual dictionary is a property of the team from the Faculty of Philology and can't be downloaded and reproduced without their permission.

In the further stage of the project, search will be extended to enable the possibility for suggestion of the lemmas and their word forms. Currently, the whole navigation in the system, as well as the search option are restricted to Macedonian language only.

We have already enabled the multilingual navigation in all other languages. The enlargement of the search option will depend on the profound linguistic support of language editors.

You are all welcome to explore the application too, and to send us your professional opinion. Thank you in advance.

REFERENCES

[1] I. Mel'čuk. "Collocations and lexical functions." In A. P. Cowie (ed.), "Phraseology: Theory, analysis, and applications. Clarendon Press, 1998, pp. 23-54.

[2] B. Salehi, N. Mathur, P. Cook and T. Baldwin. "The impact of multiword expression compositionality on machine translation evaluation." In Proceedings of NAACL-HLT, 2015, pp. 54-59.

[3] M. Němcová, "Comparative analysis of English and French body Idioms." PhD diss., Masarykova univerzita, Pedagogická fakulta, 2013.

[4] C. Bannard, and C. Callison-Burch. "Paraphrasing with bilingual parallel corpora." In Proceedings of the 43rd Annual meeting on association for computational linguistics, association for computational linguistics, 2005, pp. 597-604.

[5] H. Cunningham, D. Maynard, K. Bontcheva and V. Tablan, A framework and graphical development environment for robust NLP tools and applications, 2002, pp. 168-175.

[6] J. Arumugam, R. Parthasarathy and L. Yacub, "Drupal as a content management system in libraries: A study". In Indian journal of science, 21(72), 2015, pp.184-188.

[7] Acquia Dev Desktop, available from https://www.acquia.com/products-services/dev-desktop

[8] Apache web server, , available from https://httpd.apache.org/

[9] Percona MySQL database server, available from https://www.percona.com/software/mysql-database/percona-server

[10] PHP - programming language that powers Drupal, available from http://php.net/

[11] phpMyAdmin - for MySQL management and querying, available from https://www.phpmyadmin.net/

[12] Metro Zymphonies Theme, available from https://www.drupal.org/project/metro_ zymphonies _theme

[13] A. Petrovski, "Morphological computer lexicon – contribution to Macedonian language resources" (in Macedonian), PhD thesis, 2008, University St. Cyril and Methodius, Faculty of Natural Sciences and Mathematics

[14] GNU General Public Licence, available from http://www.gnu.org/licenses/old-licenses/gpl-2.0.en.html

[15] Creative Commons, available from http://creativecommons.org/licenses/by-sa/2

Polar Hough Detector in the Presence of Randomly Arriving Impulse Interference

Ivan Garvanov
University of Library Study and Information Technologies
Sofia, Bulgaria
igarvanov@yahoo.com

Stoyan Vladimirov
University of Library Study and Information Technologies
Sofia, Bulgaria
stoyanvladimirov@yahoo.com

Magdalena Garvanova
University of Library Study and Information Technologies
Sofia, Bulgaria
m.garvanova@unibit.com

***Abstract*—In the paper we studied average decision threshold (ADT) of Hough detector with CFAR BI processor in the presence of randomly arriving impulse interference (RAII). The randomly arriving impulse interference is mathematically described as Binominal pulse sequences with Raleigh amplitude distribution. This model of impulse noise is used for numerical analysis of Cell Averaging Constant False Alarm Rate (CA CFAR) pulse detector and CFAR BI (binary integration) pulse train detector with Polar Hough transform. The effectiveness of research CFAR detectors is expressed in terms of the detection probability and the average decision threshold. The ADT of the researched detectors are obtained as values of SNR when detection probability is equal to 0.5.**

The experimental results are obtained by numerical analysis in MATLAB environment.

Keywords—Polar Hough Detector, Average Decision Threshold, Randomly Arriving Impulse Interference

I. INTRODUCTION

The Cell-Averaging Constant False Alarm Rate detector with binary integration (CA CFAR BI detector) used for target detection is a two-step thresholding technique. A primary threshold is a set for single pulse detection. The number of decisions, where the first threshold is exceeded, is counted. If this number exceeds the second threshold, target detection is declared. According to Finn and Johnson [1], the primary threshold in the CFAR BI detector is adaptive and formed by averaging the outputs from the reference window surrounding the test resolution cell. The presence of randomly arriving impulse interference in the observation space is cause for drastic degradation of radar target detection.

In radar technology and similar fields, track-before-detect (TBD) is a concept according to which a signal is tracked before declaring it a target. In this approach, the sensor data about a tentative target are integrated over time and may yield detection in cases when signals from any particular time instance are too weak against clutter (low signal-to-noise ratio) to register a detected target.

In this work, we propose to use an additional scan-to-scan Polar Hough Transform (PHT) for improving of detection in conditions of randomly arriving impulse interference where both parameters, the average power and the average repetition frequency, are very large (binominal flow from RAII). The applying of Standard Hough Transform for target trajectory detection in conditions of Poisson flow from RAII has been proposed in [2, 3]. Several Hough detector structures, employing different CFAR processors for signal detection in the (range-time) space are studied and compared in [2-6].

The analytical expressions for the probability of detection and the probability of false alarm of CA CFAR and CA CFAR BI detectors in the presence of Binominal distribution impulse interference are presented in [7]. The Binominal model is more general than Poisson distribution model. Therefore, all mathematical formulas for evaluation of both probability measures, the probability of detection and the probability of false alarm, should be derived for Binominal distribution pulse jamming situation for a CFAR detector.

In this paper, we study the performance of CA CFAR Hough and CA CFAR BI Hough detectors in the presence of Binominal distribution impulse interference. The ADT is obtained by detection probability for a highly fluctuating Swerling II target.

The obtained results may be successfully applied for target detection by using pulse train signals in both, existing radar and communication networks. The research work is performed in MATLAB computational environment.

II. STRUCTURE OF A CFAR BI HOUGH DETECTOR

The structure of a Hough detector is shown on Figure 1. The signal processing includes CFAR BI detection, plot extraction in (r-a) space, Polar Hough Transform and Binary integration [6]. The range-azimuth observation space is formed after N_{SC} radar scans. After CFAR detection, using the PHT, all points of the polar data space, where targets are detected, are mapped into curves in the polar Hough parameter space. In the Hough parameter space, after binary integration

of data, both target and linear trajectory are detected if the result of binary integration exceeds the detection threshold T_H.

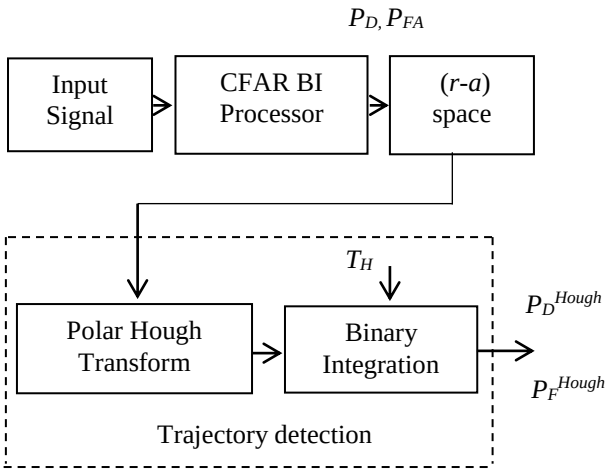


Fig. 1. Structure of a CFAR BI Hough detector

III. SIGNAL MODEL

Let us assume that L pulses hit the target, which is modeled according to the Swerling case II. The received signal power is sampled in range by using (N+1) resolution cells resulting in a data matrix with (N+1) rows and L columns. Each row of the data matrix is of signal values obtained for L pulse repetition intervals in one range resolution cell. The sampling rate in range is such that the samples in each column are statistically independent. Let us also assume that the first "N/2" and the last "N/2"rows of the data matrix are used as reference cells in order to estimate the noise level in the test resolution cells of the data matrix. The test resolution cells are the "N/2+1"row of the data matrix. The distribution law of samples in the data matrix depends on the impulse noise model.

The Binomial model describes a situation when the impulse noise is derived from two independent and identical impulse-noise sources, each of which generates a random impulse sequence with the same power intensity and the same average repetition frequency [7]. The probability of occurrence (e) of a random pulse generated by each impulse-noise source in each range resolution cell can be expressed as $e=F_j \cdot t_c$, where F_j is the average pulse repetition frequency of and t_c is the transmitted pulse duration. This means that the elements of the reference window are drawn from three classes. The first class represents the receiver noise only with probability $(1-e)^2$. The second one represents a situation when the signal samples are corrupted by a random impulse generated by one or the other impulse-noise source. This situation occurs with probability $2e(1-e)$. The third class represents a situation when the signal samples are corrupted by a total random pulse that is a sum of pulses generated by the two impulse-noise sources. This situation occurs with probability e^2. According to the theorem of total probability, the elements of the reference window are independent random variables distributed with the following probability density function (PDF):

$$f(x_i) = \frac{(1-e)^2}{\lambda_0}\exp\left(\frac{-x_i}{\lambda_0}\right) + \frac{2e(1-e)}{\lambda_0(1+r_j)}\exp\left(\frac{-x_i}{\lambda_0(1+r_j)}\right) + \frac{e^2}{\lambda_0(1+2r_j)}\exp\left(\frac{-x_i}{\lambda_0(1+2r_j)}\right) \quad (1)$$

where i=0÷N, λ_0 is the average power of the receiver noise, r_j is the average per pulse interference-to-noise ratio (INR) at the receiver input, and N is the number of samples in the reference window.

In the presence of a wanted signal in the test resolution cell the signal samples are independent random variables distributed with the following PDF [7]:

$$f(x_{0l}) = \frac{(1-e)^2}{\lambda_0(1+s)}\exp\left(\frac{-x_{0l}}{\lambda_0(1+s)}\right) + \frac{2e(1-e)}{\lambda_0(1+r_j+s)}\exp\left(\frac{-x_{0l}}{\lambda_0(1+r_j+s)}\right) + \frac{e^2}{\lambda_0(1+2r_j+s)}\exp\left(\frac{-x_{0l}}{\lambda_0(1+2r_j+s)}\right) \quad (2)$$

where s is the average per pulse signal-to-noise ratio (SNR).

IV. CFAR BI PROCESSOR ANALYSIS

The performance of the CFAR BI processor includes two steps – Cell Averaging CFAR pulse detection and a binary integration. Probability of detection and the false alarm probability of CA CFAR and CA CFAR BI detectors in condition of Binominal impulse interference are obtained in [7]. The probability of target detection of a CFAR BI detector with the binary rule M-out-of-L is evaluated by

$$P_D = \sum_{l=M}^{L} C_L^l (P_d)^l (1-P_d)^{L-l} \quad (3)$$

where

$$P_d = \sum_{i=0}^{N} C_N^i e^{2i} \sum_{j=0}^{N-i} C_{N-i}^j (2e(1-e))^j (1-e)^{2(N-i-j)} \{R_1 + R_2 + R_3\} \quad (4)$$

$$R_1 = \frac{(1-e)^2}{\left(1+\frac{T(1+2r_j)}{1+s}\right)^i \left(1+\frac{T(1+r_j)}{1+s}\right)^j \left(1+\frac{T}{1+s}\right)^{N-i-j}}$$

$$R_2 = \frac{2e(1-e)}{\left(1+\frac{T(1+2r_j)}{1+r_j+s}\right)^i \left(1+\frac{T(1+r_j)}{1+r_j+s}\right)^j \left(1+\frac{T}{1+r_j+s}\right)^{N-i-j}}$$

$$R_3 = \frac{e^2}{\left(1+\frac{T(1+2r_j)}{1+2r_j+s}\right)^i \left(1+\frac{T(1+r_j)}{1+2r_j+s}\right)^j \left(1+\frac{T}{1+2r_j+s}\right)^{N-i-j}}$$

where T is a predetermined scale factor that provide a constant false alarm rate (P_{fa}), P_d is the probability of pulse detection of a CA CFAR detector in the presence of binominal distribution impulse interference. The probability of false alarm of CA CFAR and CA CFAR BI detectors are evaluated by (4) and (3), setting $s=0$.

When the probability for the appearance of impulse interference is small (up to 0.1), then the flow is Poisson distributed and the probability characteristics of a CA CFAR processor are obtained as in [8]. The characteristics of the studied detectors for Binominal distribution impulse interference are more general and include the probability characteristics of these detectors in the presence of Poisson distribution impulse interference.

V. POLAR HOUGH DETECTOR ANALYSIS

In real radar applications, the estimated target coordinates are given in the polar coordinate system (distance and azimuth). The input parameters for the polar Hough transform are the output parameters of the search radar. For that reason, the detection algorithms using the polar Hough transform for trajectory and target detection. Another important advantage is the signal processing stability when the target changes its speed and moves at different azimuths. According to [6], the Polar Hough transform is defined by two polar coordinates, distance and azimuth (r-a). In such a way, the Polar Hough transform represents each point of a straight line in the form:

$$\rho = r\cos(a-\theta), \quad 0 < (a-\theta) \le \pi \tag{5}$$

where θ is the angle and ρ is the smallest distance to the origin of polar coordinate system.

After Ns radar scans a polar coordinate data map (r-a) is formed containing one trajectory, and it is presented on fig. 2.

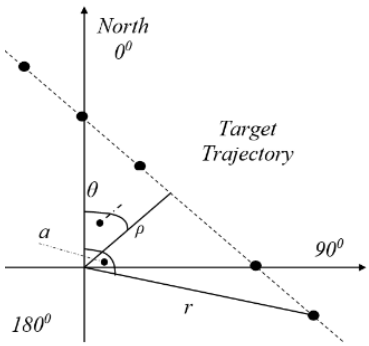


Fig. 2. Cartesian polar coordinate system.

The points' coordinates in (r-a) space form the polar parameter space. A single $(\rho-\theta)$ point in the parameter space corresponds to a single straight line in the (r-a) data space with that ρ and θ values. The result of transformation is a sinusoid with unit magnitudes (fig. 3).

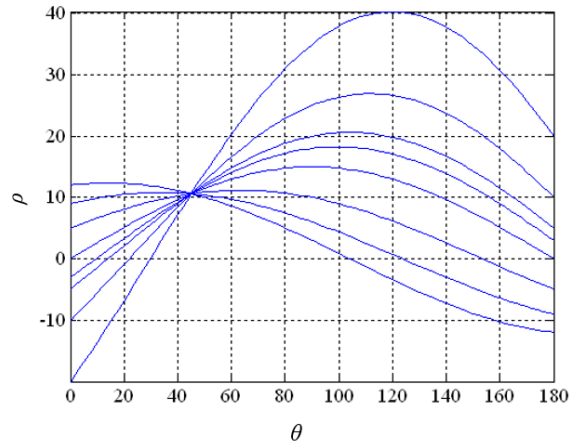


Fig. 3. Hough parameter space

Each point in the polar Hough parameter space corresponds to one line in the polar data space with parameters ρ and θ. A single ρ and θ point in the parameter space corresponds to a single straight line in the (r-a) data space with these ρ and θ values. Each cell from the polar parameter space is intersected by a limited set of sinusoids obtained by PHT. Each sinusoid corresponds to a set of possible lines through the point. If a line exists in the polar data space, by means of PHT it is represented as a point of intersection of sinusoids defined by PHT. The polar data space is built from range-azimuth cells, containing the coordinates of targets after NS scans. The parameters ρ and θ have the linear trajectory in the polar Hough parameter space and can be transformed back to the polar data space showing the current distance to the target. If the number of binary integrations (BI) of data in the polar Hough parameter space (of intersections in any of the cells in the parameter space) exceeds the detection threshold TH, both target and linear trajectory detection are indicated (fig. 4).

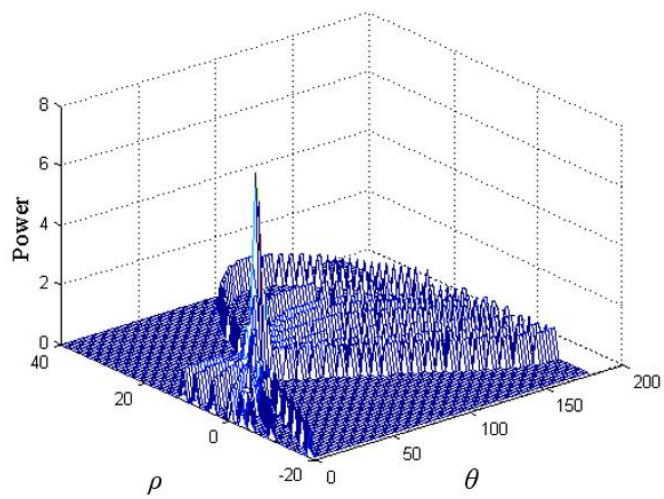


Fig. 4. Binary integration of data in Hough space

Target and linear trajectory detection are carried out for all cells from the polar Hough parameter space. The cumulative probability of target detection in Hough parameter space P_D^{Hough} cannot be written in the form of a simple Bernoulli sum. As a target moves with respect to the radar, the SNR of the received signal changes, depending on the distance to the target. The detection probability of a pulse $P_d^{(*)}$ changes as well. Then the probability P_D^{Hough} can by calculated by Brunner's method [8]. For N_s scans we have

$$P_D^{Hough} = \sum_{i=T_H}^{N_S} P_d^{(*)}(i, N_S) \tag{6}$$

where T_H is the detection threshold in Hough parameter space and $P_d^{(*)}$ is the detection probability of CA CFAR by P_d or CA CFAR BI by P_D detectors.

We study the performance of a CA CFAR detector and a CA CFAR BI detector with polar Hough transform in the presence of randomly arriving impulse interference with binomial distributed flow. The detection probabilities of the considered detectors are obtained by numerical analysis. The average decision thresholds of the detectors are obtained as values of SNR when detection probability is equal to 0.5.

VI. Numerical Results

The results are obtained for the following input parameters: average power of the receiver noise λ_0=1; average interference-to-noise ratio (INR) r_j=30dB; probability for the appearance of impulse interference with average length in the range cells from 0.1 to 0.9; number of reference cells N=16; number of test cells L=16; CFAR binary rules M/L=10/16 and 16/16, number of scans N_s=20, Hough detection threshold T_H=7 and 13, probability of false alarm P_{FA}=10^{-4}. The ADT of a CA CFAR and CA CFAR Hough processors are shown on Fig. 5.

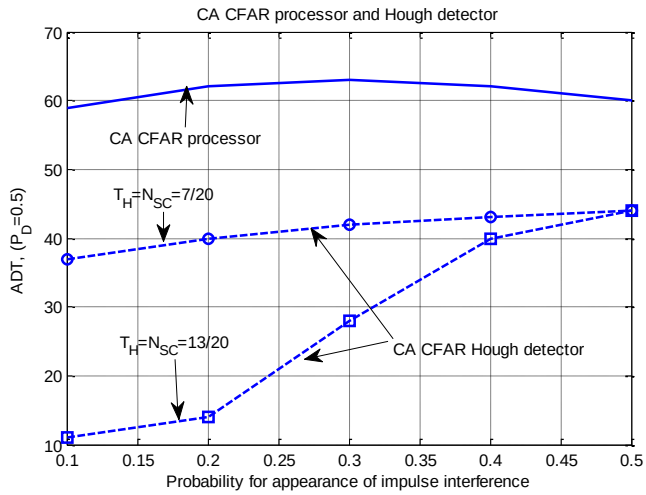


Fig. 5. ADT of a CA CFAR and of CA CFAR Hough detectors for TH=7 and 13

The ADT are received from (4 and 6) using the signal-to-noise ratio (SNR) required for the adjustment of the detection probability P_D^{Hough} = 0.5. The CA CFAR Hough processors with binary rule TH-out-of-Ns=13/20 is better in cases of lower values of the probability for the appearance of impulse interference, up to 0.5.

The ADT for CFAR BI and CFAR BI Hough processors are shown on Fig. 6. The ADT are received from (3 and 6). The CFAR BI Hough processor with the binary rule M/L=10/16 is better in cases of lower values of the probability for the appearance of impulse interference, up to 0.4. For higher values of the probability for the appearance of impulse interference, above 0.4, the using of the binary rule M/L=16/16 results in lower losses.

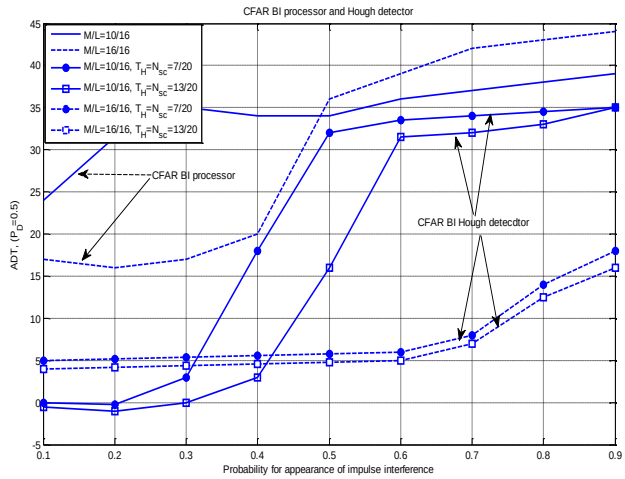


Fig. 6. ADT of a CA CFAR BI detector and of CA CFAR BI Hough detector for TH=7 and 13, M/L = 10/16 (solid line) and 16/16 (dashed line)

When the binary integration is applied in consequence, the benefit for the CFAR BI Hough detector with the binary rule M/L=16/16 and threshold T_H=7 is 25 [dB] and 30 [dB] for threshold T_H=13, in both cases e=0.9.

VII. Conclusions

The performance of CA CFAR and CA CFAR BI detectors with Polar Hough transform in the presence of randomly arriving impulse interference with Binomial distributed flow are studied in this paper.

The results show that the CA CFAR BI Hough detector, in which the binary decision rule is 10/16, is the most effective in case when the probability of occurrence of a random pulse is lower than 0.4. In a contrary situation, when the probability of occurrence of a random pulse is higher than 0.4, the CA CFAR BI Hough detector with binary decision rule is 16/16 is more appropriate.

Acknowledgment

This work was supported by Bulgarian Science Foundation, the projects DFNI-T 02/3/2014.

REFERENCES

[1] Finn, H., R. Johnson, "Adaptive detection mode with threshold control as a function of spatially sampled clutter estimation", RCA Rev, 29,3, 1968, pp. 414-461, 1968.

[2] Behar V., C. Kabakchiev, "Excision CFAR Binary Integration Processors", Compt. Rend. Acad. Bulg. Sci., vol. 49, № 11/12, pp. 45-48, 1996.

[3] Behar V., B. Vassileva, C. Kabakchiev, "Adaptive Hough Detector with Binary Integration in Pulse Jamming", ECCTD'97, Budapest, pp. 885–889, 1997.

[4] Garvanov I., C. Kabakchiev, H. Rohling, "Detection Acceleration in Hough Parameter Space by K-stage Detector", 6th International Conference "Numerical Methods and Applications" NM&A-2006, Springer-Verlag Berlin Heidelberg, Lecture Notes in Computer Science, LNCS 4310, (ISBN 978-3-540-70940-4), pp. 558-565, 2007.

[5] Garvanov I., C. Kabakchiev, "Radar Detection and Track in Presence of Impulse Interference by using the Polar Hough Transform", Journal of the European Microwave Association, Vol. 3, March 2007, (ISBN 88-8492-324-7), pp. 170-175, 2007.

[6] Garvanov I., C. Kabakchiev, "Radar Detection and Track Determination with a Transform Analogous to the Hough Transform", Proc. of International Radar Symposium – IRS 2006, (ISBN: 83-7207-621-9, IEEE Catalog Number: 06EX1284), Krakow, Poland, 24-26 May, 2006, pp. 121-124, 2006.

[7] Garvanov I., "CFAR BI Detector in Binomial Distribution Pulse Jamming", Comptes Rendus de l'Academie Bulgare des Sciences, (ISSN: 0861-1459), vol. 56, №10, pp. 37-44, 2003.

[8] Garvanov I., "Detection of moving targets and their trajectories", Sofia, "Za bukvite – O pismeneh", ISBN 978-954-2946-90-8, 2013.

[9] Carlson B., Evans E. and S. Wilson - "Search Radar Detection and Track with the Hough Transform", IEEE Trans., vol. AES - 30.1.1994, Part I, pp. 102-108; Part II, pp. 109-115; Part III, pp. 116-124, 1994.

Learning the Lorenz attractor with artificial neural networks and cooperative ensembles

Miroslav Mirchev
Faculty of computer science and engineering
Ss. Cyril and Methodius University in Skopje
Rugjer Boshkovikj 16, 1000 Skopje, R. Macedonia
Email: miroslav.mirchev@finki.ukim.mk

***Abstract*—Learning the short-term and long-term characteristics of chaotic systems is a widely studied topic. Some recent methods to study these characteristics employed a cooperative ensemble of imperfect models to merge the strengths of the different models. Certain types of neural networks have been known for their applicability in time-series predictions, such as multilayer recurrent and time-delayed neural networks. Therefore, in this work I compare the cooperative ensemble with several different architectures of artificial neural networks with recurrent connections using the Lorenz system exhibiting chaotic behavior.**

I. Introduction

Predicting the future behavior of the dynamical systems in the nature is a long studied topic. Some natural processes are not difficult to be predicted well with a certain accuracy, and some even have periodic or quasi-periodic behavior. On the other hand, there are many dynamical systems in the nature that exhibit a chaotic behavior, in a sense that a very small difference in the initial state cause a rapid divergence in the future system's behavior. Many analysis and methods exist for modeling and prediction of chaotic systems, such as [1] and [2], as well as state and parameter estimation [3].

Neural networks have been widely employed for chaotic time series prediction [4], [5], [6], [7]. Other methods are also known to be useful for time-series forecasting, such as Kalman filtering [8].

In this work we consider the application of several known artificial neural network architectures in representing the well known Lorenz oscillator. Moreover, we make some comparison with the representation of the same oscillator using the cooperative ensembles of imperfect models presented in [9] and [10].

The paper is organized in the following way. In Section II we present the Lorenz oscillator, the cooperative ensemble and the artificial neural networks. The results are given in Section III, while Section IV concludes this paper.

II. Models and Methods

A. The Lorenz oscillator

The Lorenz oscillator [11] exhibits a chaotic behavior that reminds to the atmosphere and is given as

$$\begin{aligned} \dot{x} &= \sigma(y-x) \\ \dot{y} &= x(\rho - z) - y \\ \dot{z} &= xy - \beta z, \end{aligned} \tag{1}$$

where the typical parameter values for obtaining chaotic behavior are σ=10, ρ=28 and β=8/3. I use these values to generated data that I refer to as "truth". The Lorenz oscillator can also exhibit other types of behavior for certain parameter values, but these are not part in the focus of this study.

B. The cooperative ensemble (supermodel)

In the modeling with a cooperative ensemble, it is assumed that several imperfect models already exist. I perturb the parameters in the Lorenz oscillator according to Table I, as used in [9] and [10], and use them as three imperfect models.

TABLE I: Perturbations in the parameter values in the Lorenz oscillators.

	σ	ρ	β
Truth	10	28	8/3
Model 1	13.25	19	3.5
Model 2	7	18	3.7
Model 3	6.5	38	1.7

The cooperative ensemble, also called a supermodel, is then built by coupling every variable in each imperfect model to its corresponding variables in the other models. Hence, for the μ-th model we have

$$\begin{aligned} \dot{x}_\mu &= \sigma_\mu(y_\mu - x_\mu) + \textstyle\sum_{\nu\neq\mu} C^x_{\mu\nu}(x_\nu - x_\mu) \\ \dot{y}_\mu &= x_\mu(\rho_\mu - z_\mu) - y_\mu + \textstyle\sum_{\nu\neq\mu} C^y_{\mu\nu}(y_\nu - y_\mu) \\ \dot{z}_\mu &= x_\mu y_\mu - \beta_\mu z_\mu + \textstyle\sum_{\nu\neq\mu} C^z_{\mu\nu}(z_\nu - z_\mu). \end{aligned} \tag{2}$$

The coupling coefficients are determined by fitting the ensemble to some previously observed training data. I generate the training data using the "truth" model and then minimize the following cost function, motivated from [9] and used in [10], to get some reasonably good coupling coefficients

$$E(\mathbf{C}) = \sum_{i=1}^{K} \int_{t_i}^{t_i+\Delta T} (\mathbf{x}_T(t) - \mathbf{x}_e(\mathbf{C},t))^2 w(t) d, \tag{3}$$

where the ensemble is denoted as $\mathbf{x}_e(\mathbf{C},t)$ and the truth state vector as $\mathbf{x}_T(t) = [x(t)\ y(t)\ z(t)]$. The difference between the ensemble and the truth is calculated over K non-overlapping trajectories of length ΔT. The differences along the trajectories are differently weighted so that those closer to the beginning have higher importance, except for the last

TABLE II: Characteristics of the multilayer artificial neural networks.

	NN I	NN II	NN III
Hidden units	10	5	5
Weights	163	68	63
Input delays	1	1	0
Output delays	2	1	1

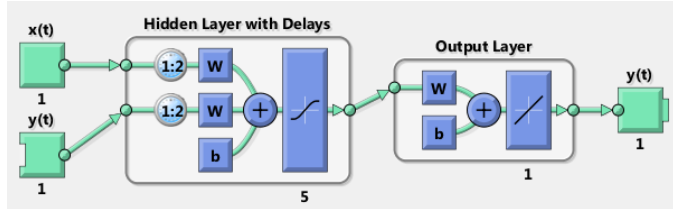


Fig. 1: The architecture of NNII.

point that has the highest importance of all, according to the following weight function

$$w(t) = \begin{cases} \alpha^j, & t \in [t_i,\ t_i + \Delta T) \\ 1, & t = t_i + \Delta T, \end{cases} \tag{4}$$

where j is the current integration step form the beginning of the ith the trajectory and $\alpha \in (0, 0.5)$ The interested reader can see a deeper interpretation of the weight function in [10]. For each trajectory the state variables of the imperfect models are initialized from the "truth" values. This cost function can be minimized using different optimization methods, two are shown in [10], and here I employ the BFGS method.

C. The neural networks

I consider three different artificial neural networks (NN) and some of the characteristics of the neural networks are given in Table II. Each of the networks has three inputs and three outputs that correspond to the three state variables of the Lorenz system. In the hidden layer I use five or ten units with sigmoid activation function, while in the output layer there are three units with linear activation function, one for each state variable. In the first two NN there is a recurrent feedback from the output to the input of the hidden layer, while in the third NN the recurrent connection goes from the output of the hidden layer, back to the input of the same layer. These recurrent feedbacks have proved to be necessary for modeling nonlinear time series and attempts to fit the training data with networks which lack the recurrent connection were unsuccessful, so I focus only on networks with recurrent feedback. Furthermore, the input of the first two NN is time delayed so as input to the network at each iteration we present the values at two consecutive time steps. Without the time-delayed input these networks fail to represent the process. In NN I the feedback is also time-delayed, which makes the network more complex, but at the same time the network can learn the modeled process better. For all neural networks considered here the back-propagation algorithm is used for training of the network. It can be noticed that NN I has considerably higher complexity than the other NN, which means that it takes more time for its training. NN II and NN III have similar complexity, with equal number of units and similar number of weights that are learned during the training. The architecture of NNII is shown in Figure 1.

III. Results

The obtained models need to be evaluated over test data, which are a continuation of the training data. In order to evaluate the models I use the same measures defined in the previous section to measure how good they represent the modeled process (truth). The first aspect was how well the models predict the truth, which can be evaluated simply by calculating the difference between the truth and the super-model. While, the second aspect is whether the models learn the general behavior of the modeled process over a long time scale. Therefore, in the evaluation I use summary statistic, such as mean values, standard deviations and covariance given in Table III. In addition, the table also includes the final training mean-squared error for each of the neural networks. From Figure 2a and Figure 2b it can be seen that the both SM and NNI learn the general behavior very good and move along the same trajectory even after 5000TU. The summary statistic given in 4 confirms that the truth is represented well by both SM and NNI, but with NNI slightly better results are obtained. The short term prediction capability is shown on Figure 2c and Figure 2d it can be noticed that NN is significantly more successful and keeps the prediction error around zero for 5TU, and closely follows the truth for 7TU, while SM suffers a desynchronization burst at 2TU and follows the truth for 5TU.

On Figure 3a and Figure 3b it is shown that also NN II and NN III learn the general behavior even though they have considerably lower complexity than NNI. The summary statistic given in Table III also confirms that the truth is represented well by both NN II and NN III. The summary statistic obtained with the cooperative ensemble (supermodel) can be found in [10]. However, generally with NN I slightly better results were obtained. The short term prediction capability of NN II and NN III is similar and it is shown on Figure 2c and Figure 2d. In fact the predictions of NN II and NN III are more close to each other than to the truth.

IV. Conclusion

I have considered several artificial neural networks that can be used to learn the complex chaotic behavior of the Lorenz oscillator. The neural networks that were successful were further tested and compared, and it was shown that with a reasonable amount of complexity the general behavior of the Lorenz oscillator can be learned well. The neural networks lacking recurrent connections or time-delay failed to represent well the modeled process. Compared to the cooperative ensemble (supermodeling) approach neural networks are more time consuming and require larger amount of training data, although they do not require any prior knowledge of the modeled process nor assume that any models already exist.

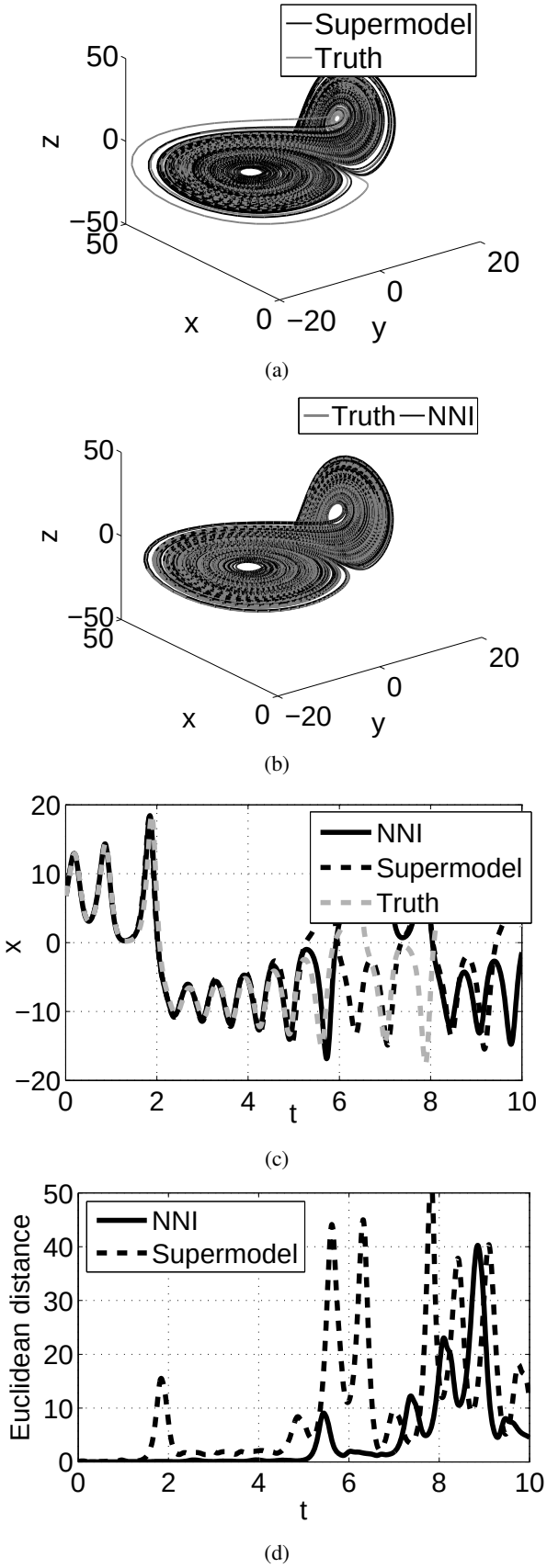


Fig. 2: Comparison of the supermodel (SM) with parameter differences and NN I trained on 40TU with the truth: (a, b) evaluation of SM and NNI over the last 100 TU of test data; (c, d) the x variables and the Euclidean distance during the last 2TU of training data and the first 10 TU of test data.

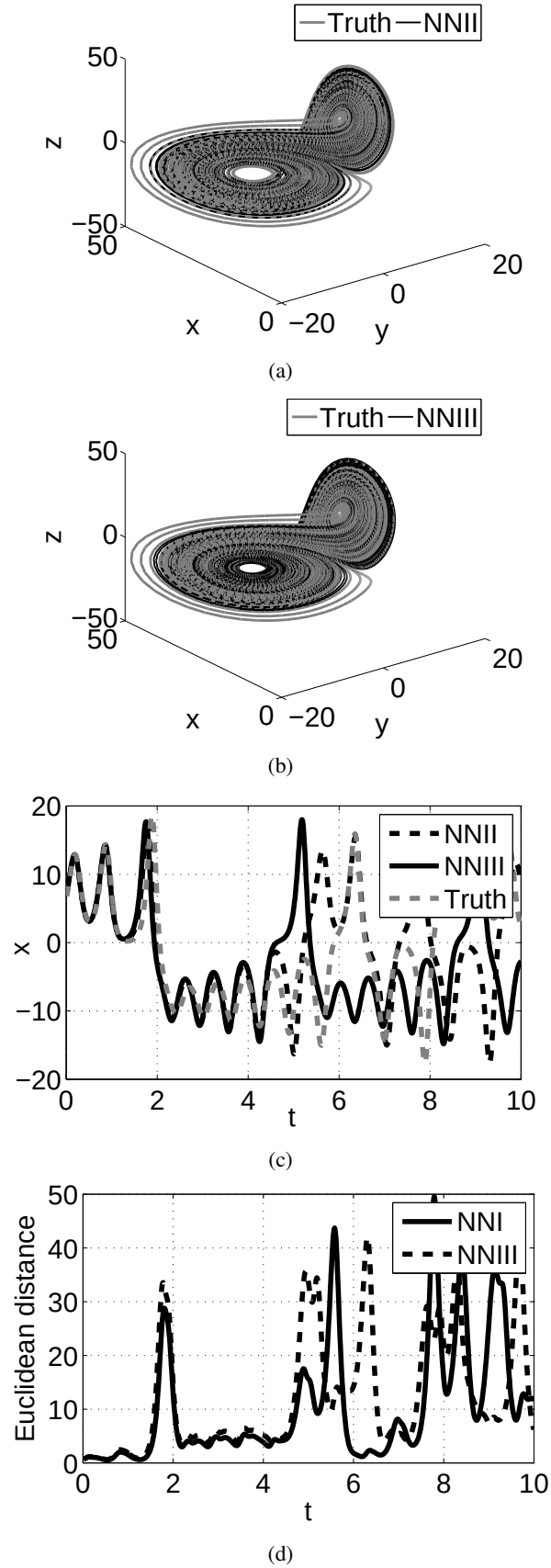


Fig. 3: Comparison of NN II and NN III trained on 40TU with the truth: (a, b) evaluation of NN II and NN III over the last 100 TU of test data; (c, d) the x variables and the Euclidean distance during the last 2TU of training data and the first 10 TU of test data.

TABLE III: Means (μ), standard deviation (σ) and covariance (cov) for the models and the truth, based on 5000TU test run. Training mean-squared error (tMSE) of the neural networks.

	Truth	NN I	NN II	NN III
tMSE	-	2.9e-6	2.1e-5	1.76e-5
μ_x	0.00	-0.12	0.11	-0.02
μ_y	0.00	-0.12	0.11	-0.01
μ_z	25.27	25.28	25.34	25.39
σ_x	8.21	8.21	8.22	8.23
σ_y	9.36	9.36	9.34	9.32
σ_z	8.30	8.29	8.20	8.13
cov_{xy}	67.35	67.39	67.56	67.72
cov_{xz}	0.38	-0.31	0.33	-0.05
cov_{yz}	0.47	-0.39	0.29	0.07
cov_{xx}	67.36	67.39	67.56	67.73
cov_{yy}	87.70	87.63	87.31	86.93
cov_{zz}	69.15	68.8	67.24	66.17

However, neural networks would be currently impractical for application in weather and climate models. Future hardware implementations of neural network could potentially provide faster simulations, which could allow their application to a wider range of systems. One example of such implementations could be based on memristive electronic circuits, such as those presented in [12].

ACKNOWLEDGMENT

This work was supported by the Faculty of Computer Science and Engineering, SS Cyril and Methodius University, Skopje, Macedonia.

REFERENCES

[1] H. D. Abarbanel, R. Brown, J. J. Sidorowich, and L. S. Tsimring, "The analysis of observed chaotic data in physical systems," *Reviews of modern physics*, vol. 65, no. 4, p. 1331, 1993.

[2] J. D. Farmer and J. J. Sidorowich, "Predicting chaotic time series," *Physical review letters*, vol. 59, no. 8, p. 845, 1987.

[3] Quinn, J. C., Abarbanel, H. D.I., "State and parameter estimation using monte carlo evaluation of path integrals," *Q.J.R. Meteorol. Soc.*, vol. 136, no. 652, pp. 1855–1867, 2010.

[4] T. Kuremoto, S. Kimura, K. Kobayashi, and M. Obayashi, "Time series forecasting using a deep belief network with restricted boltzmann machines," *Neurocomputing*, vol. 137, pp. 47–56, 2014.

[5] T. Kuremoto, M. Obayashi, K. Kobayashi, T. Hirata, and S. Mabu, "Forecast chaotic time series data by dbns," in *Image and Signal Processing (CISP), 2014 7th International Congress on*, pp. 1130–1135, IEEE, 2014.

[6] M. Han, J. Xi, S. Xu, and F.-L. Yin, "Prediction of chaotic time series based on the recurrent predictor neural network," *IEEE Transactions on signal processing*, vol. 52, no. 12, pp. 3409–3416, 2004.

[7] J.-S. Zhang and X.-C. Xiao, "Predicting chaotic time series using recurrent neural network," *Chinese Physics Letters*, vol. 17, no. 2, p. 88, 2000.

[8] S. Sarkka, A. Vehtari, and J. Lampinen, "Time series prediction by kalman smoother with cross-validated noise density," in *Neural Networks, 2004. Proceedings. 2004 IEEE International Joint Conference on*, vol. 2, pp. 1653–1657, IEEE, 2004.

[9] van der Berge, L.A., Selten, F.M., Wiegerinck, W., Duane, G.S., "A multi-model ensemble method that combines imperfect models through learning," *Earth Sys Dyn*, vol. 2, no. 1, pp. 161–177, 2011.

[10] Mirchev, M., Duane, G.S., Tang,W.K., Kocarev, L., "Improved modeling by coupling imperfect models," *Commun Nonlinear Sci Numer Simul*, vol. 17, no. 7, pp. 2741–2751, 2012.

[11] E. Lorenz, "Deterministic nonperiodic flow," *J Atmos Sci*, vol. 20, no. 2, pp. 130–141, 1963.

[12] Y. V. Pershin and M. Di Ventra, "Experimental demonstration of associative memory with memristive neural networks," *Neural Networks*, vol. 23, no. 7, pp. 881–886, 2010.

Performance Comparison of Random Forests and Extremely Randomized Trees

Eftim Zdravevski, Petre Lameski, Andrea Kulakov, Vladimir Trajkovik
Faculty of Computer Science and Engineering
Ss.Cyril and Methodius University, Skopje, Macedonia
Email: {eftim.zdravevski, petre.lameski, andrea.kulakov, vladimir.trajkovik}@finki.ukim.mk

***Abstract*—Random Forests (RF) recently have gained significant attention in the scientific community as simple, versatile and efficient machine learning algorithm. It has been used for variety of tasks due it its high predictive performance, ability to perform feature ranking, its simple parallelization, and due to its low sensitivity to parameter tuning. In recent years another tree-based ensemble method has been proposed, namely the Extremely Randomized Trees (ERT). These trees by definition have similar properties. However, there is no extensive empirical evaluation of both algorithms that would identify strengths and weaknesses of each of them. In this paper we evaluate both algorithms of several publicly available datasets. Our experiments show that ERT are faster as the dataset size increases and can provide at least the same level of predictive performance. As for feature ranking capabilities, we have statistically confirmed that both provide the same ranking, provided that the number of trees is large enough.**

***Keywords*—Random Forests, Extremely Randomized Trees, Decision Trees, Ensembles of Trees**

I. Introduction

In recent years, ensemble methods for machine learning has been used extensively in the research community, as well as by industry practitioners. Classification predictive performance have been improved by growing an ensemble of trees and that vote for the most popular class. In particular, a very popular method for ensemble of decision trees is the Random Forests algorithm [1]. Another recently popularized method is the Extremely Randomized Trees [2].

The paper is organized as follows. The next subsection describes briefly these algorithms. Next is described the method for feature extraction used in the experiments. Section IV described the experimental setup and discusses the results. Finally, in section V we conclude the paper.

II. Random Forests and Extremely Randomized Trees

Random forests are a combination of decision trees in which each tree depends on the values of a random feature space sampled independently and with the same distribution for all trees in the forest. The generalization error for forests converges a.s. to a limit as the number of trees in the forest becomes large. The generalization error of a forest of tree classifiers depends on the strength of the individual trees in the forest and the correlation between them.

This work was partially financed by the Faculty of Computer Science and Engineering at the Ss.Cyril and Methodius University, Skopje, Macedonia

In [2] proposed the Extremely Randomized Trees tree-based ensemble method for supervised classification and regression problems. It randomizes both attribute and cut-point choice while splitting a tree node. In the extreme case, it builds totally randomized trees whose structures are independent of the output values of the learning sample. The strength of the randomization can be tuned to problem specifics by the appropriate choice of a parameter.

III. Feature extraction and selection

Extracting robust features from time series is a challenging task, but using a systematic approach, for our experiments we are generating a variety of features. A recent data mining competition for posture recognition of firefighters [3] was able to inspire different feature engineering approaches that are very effective [4], [5], [6]. Additionally, for feature selection we have used a ranking method, proposed in [7], the based that is also able to detect features that are subject to data drift, therefore that can potentially degrade performance over time. As a result of the evaluation of various feature subsets, we can analyze the performance of both classifiers.

IV. Results

In order to evaluate the performance of both algorithms, we have selected a publicly available dataset for activity recognition [8]. This dataset, which is extensively described in [9], consists of raw readings of 1 chest-mounted accelerometer with 3 axes. The subjects were performing 7 different actions, but for experimenting the authors that published this dataset have been using only the 5 actions, so we did as well. The total number of instances in the dataset was 73899.

On Fig. 1 and Fig. 2 the average accuracy and execution time for 5 fold cross-validation depending on the number of features and classification algorithm are shown. It is evident that both algorithms are able to cope with redundant features fairly successfully, albeit the Extremely Randomized Trees are offering somewhat better accuracy. Moreover, the Extremely Randomized Trees are always performing faster and this becomes important as the number of features and instances increase.

Next we wanted to investigate the variance and stability of estimated feature scores by both algorithms, so we have repeated the experiments twice with the same feature sets. Table II lists some statistics based on the feature scores listed in Table II. Namely, Table II shows the feature scores of the 24 most common features that were used in the previous

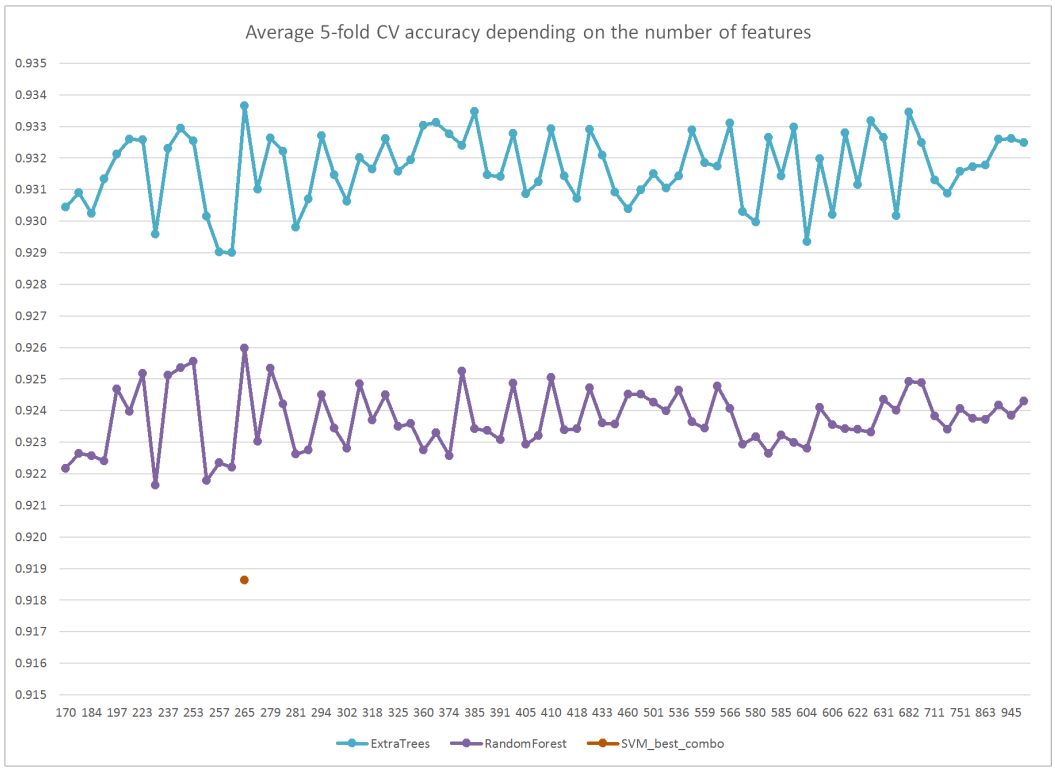


Fig. 1. Average 5-fold CV accuracy per personal model depending on the number of features and classification algorithm.

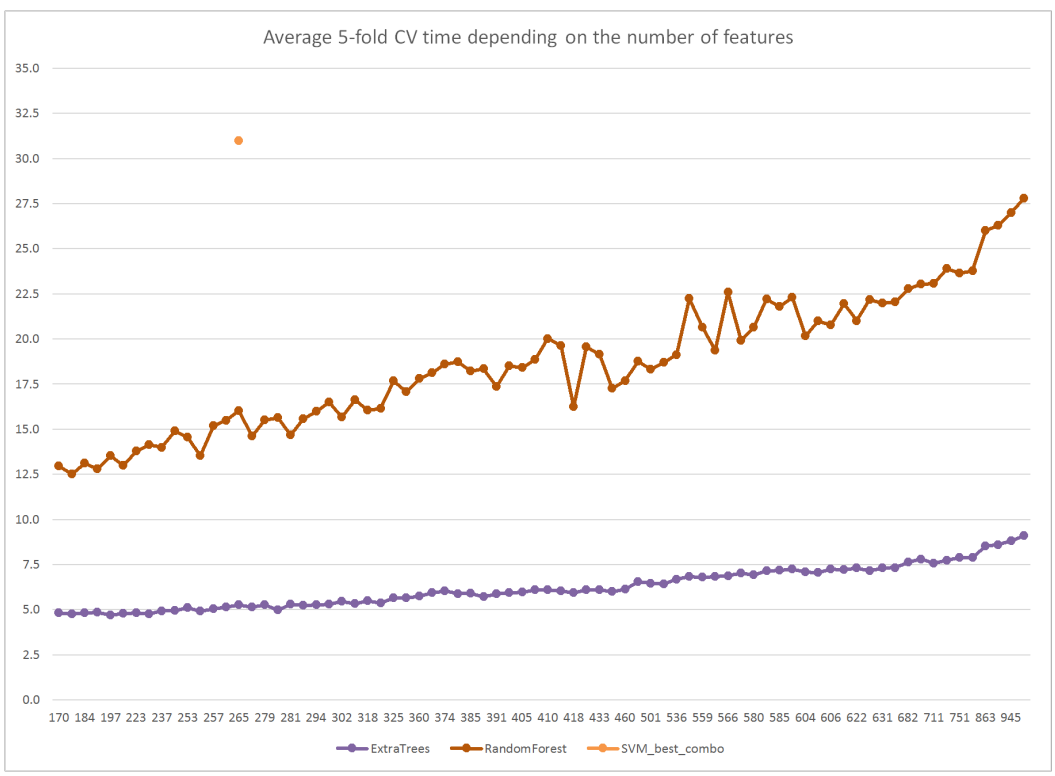


Fig. 2. Average 5-fold CV time in seconds per personal models depending on the number of features and classification algorithm.

evaluation. We performed 2-tailed T-Test of the feature scores obtained by both algorithms, as well as comparing the scores obtained by the 2 classifiers, as they are implemented in [10]. It is evident that the estimated feature scores are very stable due to the large number of trees. Then we calculated the standard deviation from the 2 scores of each feature for each classifier separately (columns ERT_std and RF_std in Table II). Calculating the mean of the standard deviations shows that both values are very low, albeit the Random Forest algorithm experiencing a lower value. Nonetheless, the statistical tests show that both algorithms can be safely used for feature importance estimation.

TABLE I. STATISTIC TESTS BASED ON DATA DESCRIBED IN TABLE II.

Metric	Result
ERT 2-tailed T-Test	0.99999813
RF 2-tailed T-Test	0.99999460
ERT vs RF T-Test	0.99999930
ERT mean of std per feature	0.00305626
RF mean of std per feature	0.00184345

V. CONCLUSION

We can conclude that both algorithms overcome redundant features very successfully, evident by the constant accuracy even if the number of features increases. It turns out that the Extremely Randomized Trees are resulting in somewhat better accuracy and are always performing faster, especially when the number of features and instances increase. Finally, based on the performed statistical tests we conclude that both algorithms provide similar feature importance estimates that are stable across multiple repetitions, regardless of the heavy randomization used intrinsically in the algorithms.

REFERENCES

[1] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32. doi: 10.1023/A:1010933404324. [Online]. Available: http://dx.doi.org/10.1023/A:1010933404324

[2] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," *Machine Learning*, vol. 63, no. 1, pp. 3–42, 2006. doi: 10.1007/s10994-006-6226-1. [Online]. Available: http://dx.doi.org/10.1007/s10994-006-6226-1

[3] M. Meina, A. Janusz, K. Rykaczewski, D. Slezak, B. Celmer, and A. Krasuski, "Tagging firefighter activities at the emergency scene: Summary of aaia'15 data mining competition at knowledge pit," in *Computer Science and Information Systems (FedCSIS), 2015 Federated Conference on*, Sept 2015. doi: 10.15439/2015F426 pp. 367–373.

[4] J. Lasek and M. Gagolewski, "The winning solution to the aaia'15 data mining competition: Tagging firefighter activities at a fire scene," in *Proceedings of the 2015 Federated Conference on Computer Science and Information Systems*, ser. Annals of Computer Science and Information Systems, M. P. M. Ganzha, L. Maciaszek, Ed., vol. 5. IEEE, 2015. doi: 10.15439/2015F418 pp. 375–380. [Online]. Available: http://dx.doi.org/10.15439/2015F418

[5] A. Zagorecki, "A versatile approach to classification of multivariate time series data," in *Proceedings of the 2015 Federated Conference on Computer Science and Information Systems*, ser. Annals of Computer Science and Information Systems, M. P. M. Ganzha, L. Maciaszek, Ed., vol. 5. IEEE, 2015. doi: 10.15439/2015F419 pp. 407–410. [Online]. Available: http://dx.doi.org/10.15439/2015F419

[6] E. Zdravevski, P. Lameski, R. Mingov, A. Kulakov, and D. Gjorgjevikj, "Robust histogram-based feature engineering of time series data," in *Computer Science and Information Systems (FedCSIS), 2015 Federated Conference on*, ser. Annals of Computer Science and Information Systems, M. P. M. Ganzha, L. Maciaszek, Ed., vol. 5. IEEE, Sept 2015. doi: 10.15439/2015F420 pp. 381–388. [Online]. Available: http://dx.doi.org/10.15439/2015F420

[7] M. Boullé, "Tagging fireworkers activities from body sensors under distribution drift," in *Proceedings of the 2015 Federated Conference on Computer Science and Information Systems*, ser. Annals of Computer Science and Information Systems, M. Ganzha, L. Maciaszek, and M. Paprzycki, Eds., vol. 5. IEEE, 2015. doi: 10.15439/2015F423 pp. 389–396.

[8] M. Lichman, "UCI machine learning repository," 2013. [Online]. Available: http://archive.ics.uci.edu/ml/datasets/Activity+Recognition+from+Single+Chest-Mounted+Accelerometer

[9] P. Casale, O. Pujol, and P. Radeva, *Pattern Recognition and Image Analysis: 5th Iberian Conference, IbPRIA 2011, Las Palmas de Gran Canaria, Spain, June 8-10, 2011. Proceedings*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, ch. Human Activity Recognition from Accelerometer Data Using a Wearable Device, pp. 289–296. ISBN 978-3-642-21257-4. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-21257-4_36

[10] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

TABLE II. FEATURE SCORES FOR THE SAME DATASET OF 24 FEATURES WITH RANDOM FOREST (RF) AND EXTREMELY RANDOMIZED TREES (ERT) WITH 2 SEPARATE RUNS

FID	ERT_score1	ERT_score2	ERT_std	RF_score1	RF_score2	RF_std
1	0.0153947	0.0108758	0.0022595	0.0070526	0.0118251	0.0023863
2	0.0420794	0.0385484	0.0017655	0.0204782	0.0182116	0.0011333
3	0.0383765	0.0221985	0.0080890	0.0155013	0.0163048	0.0004018
4	0.0492405	0.0427291	0.0032557	0.0342975	0.0329070	0.0006953
5	0.0307088	0.0258354	0.0024367	0.0021842	0.0058967	0.0018563
6	0.0417594	0.0540148	0.0061277	0.0720881	0.0532936	0.0093973
7	0.1802803	0.2034607	0.0115902	0.1572349	0.1670721	0.0049186
8	0.0437639	0.0549879	0.0056120	0.0391848	0.0485794	0.0046973
9	0.0123589	0.0116197	0.0003696	0.0472597	0.0445050	0.0013774
10	0.0192220	0.0221837	0.0014809	0.0799508	0.0760616	0.0019446
11	0.0041322	0.0031622	0.0004850	0.0010633	0.0009209	0.0000712
12	0.0012066	0.0013322	0.0000628	0.0016406	0.0015975	0.0000216
13	0.0009440	0.0009317	0.0000061	0.0007464	0.0008139	0.0000338
14	0.0012444	0.0008918	0.0001763	0.0006912	0.0006698	0.0000107
15	0.0113268	0.0115136	0.0000934	0.0020207	0.0010306	0.0004951
16	0.0103095	0.0165596	0.0031251	0.0018599	0.0014173	0.0002213
17	0.0562460	0.0443660	0.0059400	0.0177128	0.0200320	0.0011596
18	0.0199697	0.0276082	0.0038193	0.0193860	0.0153704	0.0020078
19	0.0402050	0.0334358	0.0033846	0.0093499	0.0061775	0.0015862
20	0.0593671	0.0591298	0.0001186	0.0602517	0.0729347	0.0063415
21	0.0703908	0.0799185	0.0047639	0.1031800	0.1035752	0.0001976
22	0.0830415	0.0734411	0.0048002	0.1317565	0.1291026	0.0013269
23	0.1626861	0.1555547	0.0035657	0.1703122	0.1666465	0.0018329
24	0.0057459	0.0057007	0.0000226	0.0047964	0.0050541	0.0001289

SHORT PAPERS

Integral operation of mobile wireless sensor networks

Miroslav Mirchev, Lasko Basnarkov and Igor Mishkovski
Faculty of computer science and engineering
Ss. Cyril and Methodius University in Skopje
Rugjer Boshkovikj 16, 1000 Skopje, R. Macedonia
Email: miroslav.mirchev@finki.ukim.mk, lasko.basnarkov@finki.ukim.mk, igor.mishkovski@finki.ukim.mk

I. Extended abstract

Wireless sensor networks have been present in our surrounding for a long time now [1] and [2]. Mobile wireless sensor networks (MWSN) are getting more and more popular nowadays and find wide applications, such as tracking of animals and vehicles or environmental monitoring. In these networks the nodes can be categorized in several types such as static sinks, mobile sensors that make measurements, mobile agents that gather information and facilitate the coordination among the static sinks and the mobile sensor nodes, etc. Some of the classical problems in MWSNs are time synchronization between nodes [3], node localization [4] and measured data exchange (routing) [5].

This work continues the research from our previous projects, in which among the other work we developed methods for cooperative localization in MWSNs. In this contribution we present methods for cooperative time synchronization of nodes, which is necessary for providing the time dimension of the measured data, as well as for the operation of the network. We consider several ways for practical implementation of the previously developed methods for localization, i.e. the procedure for exchanging control information among nodes. Furthermore, we demonstrate schemes for exchange of the measured data in these networks with such an organization, where the data are already labeled with appropriate location and time.

In Fig. 1, we show one of the networks for which we demonstrate its operation, where the nodes' trajectories are generated using sums of several sinusoidal functions with different characteristics, while the beacons (sinks) are placed in fixed positions. All measured data need to be time-stamped, location-labeled and relayed towards the sinks from where it can be sent to a central data repository. The time synchronization and localization are performed in an integral approach using a kind of Extended Kalman filters (EKF) that estimate the time clock, and nodes' positions and velocities. The information exchange could require multiple hops and the routing scheme builds upon the previous filters using the Kalman gain as an indicator for the next hop destination. The closer a node is to the beacons the smaller its Kalman gain is. Hence, when there are multiple possible routes with equal hop count the information will be forwarded towards the nodes that have smaller Kalman gain and are probably closer to the beacons.

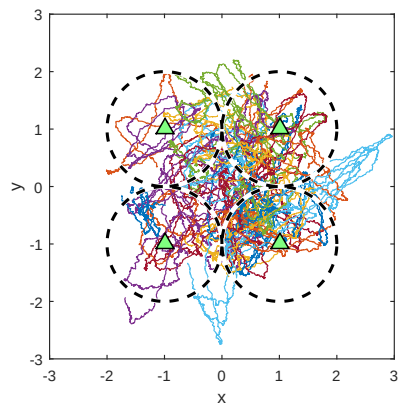


Fig. 1. A mobile sensor network with the nodes' trajectories and four beacons (triangles).

Acknowledgment

This work was supported by the Faculty of Computer Science and Engineering, SS Cyril and Methodius University, Skopje, Macedonia.

References

[1] J. Yick, B. Mukherjee, and D. Ghosal, "Wireless sensor networks survey," vol. 52, pp. 2292–2330, 2008.
[2] I.F. Akyildiz, W. Su, Y. Sankarasubramaniam and E. Cayirci, "Wireless sensor networks: a survey," *Computer Networks*, vol. 38, no. 4, pp. 393–422, 2002.
[3] J. Elson and D. Estrin, "Time synchronization for wireless sensor networks.," in *IPDPS*, vol. 1, p. 186, 2001.
[4] I. Amundson, and X.D. Koutsoukos, "A survey on localization for mobile wireless sensor networks," in *Mobile Entity Localization and Tracking in GPS-less Environnments*, pp. 235–254, Springer, 2009.
[5] K. Akkaya and M. Younis, "A survey on routing protocols for wireless sensor networks," *Ad hoc networks*, vol. 3, no. 3, pp. 325–349, 2005.

Comparison of several algorithms for plant segmentation

Petre Lameski*, Eftim Zdravevski* and Andrea Kulakov*,
*University of Sts. Cyril and Methodius in Skopje, Faculty of Computer Science and Engineering

***Abstract*—Segmentation of vegetation from the ground in images is one of the main steps in image processing for agricultural and vegetation images. The main goal of this step is to calculate which areas of the images are populated with vegetation and which belong to other objects and land. In this paper we compare several algorithms for segmentation for their ability to successfully differentiate between plants and other object in images. The algorithms were tested on a dataset of plant images taken under controlled lighting conditions and in the visible and near infra red (NIR) spectrum. From these images the NDVI index is calculated and the result is used for the segmentation. The segmentation algorithms and the obtained results are presented in this paper.**

I. Introduction

The food production has always been an issue in human history. The current trends show that by 2050 the human population will increase so much that the food production will not be able to satisfy the demand [1]. One of the ways to overcome this problem is to increase the automation in the agricultural food production. To do this an efficient way for plant detection and recognition is needed for the automated machines to be able to perform their tasks. The process of weeding is one of the most important processes in agriculture that significantly increases the yield of plants per land unit. The weeding is consisted of removing unwanted plants physically or trough chemical treatment of plants. To automate this process, an efficient algorithms for weed detection and localization are needed. One way to solve this problem is to use camera as a sensor and process the images from the camera. The first step in weed detection is segmentation of the plants from the background land mass. In this paper we experiment with several algorithms for segmentation that could be used for segmentation of plants from land mass in images.

II. System architecture

The proposed system for weed detection from images is consisted on three parts. After the images are taken, the first part segments the image on plant pixels and land pixels. After that, the second part classifies the patches of the image a pre-trained model for weed and non-weed patch. Then the the last part merges the patches together and a weed and non-weed marked image is formed that can be used as an input for a weed removal system. In this paper we are focused on the first part of the system - the plant segmentation. The plant segmentation problem can be defined as a segmentation problem where the segments are defined as plant and non-plant segments. The two class segmentation can be accomplished with several existing algorithms.

III. Segmentation algorithms

One of the most common algorithms for image segmentation is the thresholding. The thresholding is executed by defining a threshold value that is used as a boundary and all pixel values larger than the threshold value are marked as one class and the others are marked as the second class. The thresholding is usually used to divide the image on foreground and background. Since in the plant segmentation problem, the plants are green and the land is brown, the segmentation is performed by defining a threshold value that should easily distinguish between the two parts. The problem of thresholding is the finding the right value for the threshold. One of the most used automated algorithms for finding the optimal threshold value based on the value distribution of the image intensities is the Otsu algorithm [2]. One of the state of the art segmentation algorithms is the GrabCut algorithm for image segmentation [3]. This algorithm requires a user input for initial image segmentation and then searches for the optimal boundaries for the segments. Another algorithm used for pixel-wise segmentation of images is the k-d trees [4] based segmentation. The k-d trees are used for finding the nearest neighbor based on the pixel values. It significantly speeds up the search process by building trees based on the initial segments and then finds the nearest neighbor of the unsegmented pixels in the segmented pixel. The nearest pixel segment is considered to be the segment of the unsegmented pixel. The watershed algorithm [5] is also used for image segmentation with prior initialization of the segments and also requires initial segmentation of the image.

IV. Results and conclusion

We analyse the proposed segmentation algorithms on 60 NIR-Red images taken from plants under constant illumination. The segmentation is performed on the NDVI [6] index images. The NDVI index has larger values where there are green patches and lower values where the image contains land. We initialize the algorithms on the top 10% values as certain green and bottom 10% values as the certain land prior to the segmentation algorithms that need initialization. The results obtained from the segmentation algorithms are presented in Table 1. The results show that the Otsu thresholding technique outperforms the other segmentation methods on the analysed

TABLE I
PER-PIXEL SEGMENTATION RESULTS

Algorithm	Accuracy	Sensitivity	Specificity
Otsu	0.9909	0.9505	0.9945
GrabCut	0.9039	0.9686	0.8982
K-d trees	0.9530	0.8858	0.9589
Watershed	0.9270	0.8454	0.9341

images. Additional experiments should be performed for other types of images taken in different color spaces.

ACKNOWLEDGMENT

The work presented in this paper was partially financed by the Faculty of Computer Science and Engineering at the Ss. Cyril and Methodius University in Skopje, Macedonia.

REFERENCES

[1] D. K. Ray, N. D. Mueller, P. C. West, and J. A. Foley, "Yield trends are insufficient to double global crop production by 2050," *PloS one*, vol. 8, no. 6, p. e66428, 2013.

[2] N. Otsu, "A threshold selection method from gray-level histograms," *Automatica*, vol. 11, no. 285-296, pp. 23–27, 1975.

[3] C. Rother, V. Kolmogorov, and A. Blake, "Grabcut: Interactive foreground extraction using iterated graph cuts," in *ACM transactions on graphics (TOG)*, vol. 23, no. 3. ACM, 2004, pp. 309–314.

[4] M. Greenspan and M. Yurick, "Approximate kd tree search for efficient icp," in *3-D Digital Imaging and Modeling, 2003. 3DIM 2003. Proceedings. Fourth International Conference on*. IEEE, 2003, pp. 442–448.

[5] S. Beucher and F. Meyer, "The morphological approach to segmentation: the watershed transformation," *OPTICAL ENGINEERING-NEW YORK-MARCEL DEKKER INCORPORATED-*, vol. 34, pp. 433–433, 1992.

[6] T. N. Carlson and D. A. Ripley, "On the relation between ndvi, fractional vegetation cover, and leaf area index," *Remote sensing of Environment*, vol. 62, no. 3, pp. 241–252, 1997.

Comparison of software packages for modeling nonlinear social and economic systems

Hristijan Trifunoski, Marija Mihova
Faculty of Computer Science and Engineering
Ss. Cyril and Methodius University
Rugjer Boshkovikj 16, PO Box 393, 1000 Skopje, Macedonia
{marija.mihova}@finki.ukim.mk

***Abstract*— This research focused on analysis of existing software packages available for use at this moment, for modeling nonlinear social/economic systems with a limited number of social/economic actors. The aim of research is to select suitable software package, tool, with which we can effectively, precisely and quickly model nonlinear systems, for example. social networks or certain economic community, separating the important parameters affecting the overall social and economic results of the community, finding a correlation formula between essential variables in those nonlinear systems and check and compare measured results in accordance with historical values. The models can be put on trial, on test of time, so that they can be continuously improved with iterative approximations verifying that real results are in full compliance with modeled values.**

Keywords – software packages, economic networks, social networks.

I. INTRODUCTION

Developing computer programs, tools and software packages for analyzing modeling and simulation of economic and social processes become very popular in recent years. There are wide range of such software packages based on different types of mathematical models [1, 2], models based on neural networks and the most popular nowadays, software based on agent systems.

All human decisions and reasonable actions are based on deductive knowledge about some phenomenon that is applicable to given situation. Analog to physicist quest, trying to find unified formula binding all four natural forces, economists are searching unified formula describing economic output/progress visa-vi essential variables that defines some numbered socio-economic group. Although there is overwhelming research in economic fields and success in partial fields of economics, no one can firmly answer, what should e.g. some state do, to maximize economic output in every turn of the economic cycle. This problem was absolved in numerous research papers, and almost every time, the approach was top-to-bottom. Now, with agent based modeling we have an opportunity to challenge that problem with bottom-up approach.

Social networks form our social and economic lives. They play a central role in the transmission of information about job opportunities and are critical to the trade of many goods and services. They are important in determining which products we buy, which languages we speak, how we vote, as well how much education we obtain, and our likelihood of succeeding professionally. The countless ways in which network structures affect our well-being make it critical to understand how social network structures impact behavior, which network structures are likely to emerge in a society, and why we organize ourselves as we do. The emergence of software modeling technics and packages helps research these social and economic networks, drawing studies by sociologists, economists, computer scientists, physicists, mathematicians. In this sense we can treat economic behavior and economic networks as a particular specific subset of social networks.

The research is focused on analyzing software packages and tools that are available today to scientists for modeling nonlinear and complex social/economic systems and behavior of some chosen group of interest. Most of the packages that are observed, are far more applicable than only in mentioned areas. They can also successfully model biological, chemical, physical and business processes, which are only few to mention from the possible applications.

Section II explains the motivation undertaking this research, in section III we make some effort to mention the first scientific steps toward addressing the social and economic problems using modeling technics. Also we address some chronological development of agent based approach. In section IV we analyze the specifics of modeling economic and social networks, types of approaches to these problems in the history and the methods that gave best results. In section V we make comparison of the most popular ABM software's available today for running in standalone/desktop environment. Agents running of HPC are out of the scope of this research.

II. MOTIVATOIN

The quality of life of the citizens in a specific country mostly depends of the economy factor. Macedonia today is experiencing a hard and turbulent economic times, so one of the most important question is how to improve the economy. In order to give some reasonable solution we need to analyze the factors that have influence on the national economy, how much the people that drive the country are responsible and are there some other natural limitations in our economic situation that cannot be surpassed by implementing any reasonable economic policy. Consulting some of the departments in the state institutions for analytics and development, we can conclude

that the one of the reasons for difficult economic situation of Macedonia is due to absence of serious modeling of the economy. Such models are of crucial importance in creating economical politics, so, it is not clear how does economic decisions take place.

In order to detect the current situation in the field of economy analyzing and modeling, we made some consultations and discussions with the ministry of finance and national bank. So, we found that there is a department in national bank that is modeling some macroeconomic indicators. The department was formed with the help of foreign aid and the main working model is focused on GDP and inflation as the target for research and modeling. The model they use for analyzing and predicting the trends in economy is based on a static formula with no more than 6 variables. Changing this variables they simulate different scenarios depending of changes in those variables. Some examples of this models are formulas for foreign reserves

fr(t) = fr(t-1)+ export(t)- import(t)+ private transfers(t)+ foreign investments (t)

And interest rates

i.r.= euribor rate + country risk+ interbank rates

The tool that they used was time series based model in Matlab/Dynare, and the product they named MAKPAM. The main target for this modeling and simulation was to keep the inflation in predicted limits and stable foreign exchange rate. The GDP growth was/is only secondary/collateral priority/effect to modeling interest.

Having all previous in mind, quest for solving this problem, naturally brought us to approaches of modeling the problem and therefore surveying existing software packages for modeling. The reason behind this was that there is too much parameters that can be put by theoretical approach, time series in past and DSGE models currently have some draw backs and also we can’t use experimental approach for obvious reason. We are planning to model the economy starting with some basic DSGE model with few parameters. After that, we will add agents that will statistically represent people, firms and other important economic actors. The process of calibrating the model will be trough iterations, comparing real data from time series and modeled results, refining the formula that will emerge by correlating included variables in each step.

III. Related Work

Agent-based models are (computational) models of a heterogeneous population of agents and their interactions. The micro-level interactions can result to interesting macro-level behavior like cooperation, segregation, fashion, culture, etc. ABMs are also known as multi-agent systems, agent-based systems etc., and is an important field in computer science where agent-based models are developed to do tasks, like searching for information on the internet. Within social science we are interested in agent-behavior that is based on our understanding of human decision making. Agents can represent individuals, households, firms, nations, depending on the application. The heterogeneity of agents is an important aspect of ABM. [3] An **agent-based model** (**ABM**) is one of a class of computational models for simulating the actions and interactions of autonomous agents (individual or collective entities such as organizations or groups) with a view to assessing their effects on the system as a whole. It combines elements of game theory, complex systems, emergence, computational sociology, multi-agent systems and evolutionary programming. Monte Carlo methods are used to introduce randomness. Particularly within ecology, ABMs are also called **individual-based models** (**IBMs**)[4], and individuals within IBMs may be simpler than fully autonomous agents within ABMs. A review of recent literature on individual-based models, agent-based models, and multi agent systems shows that ABMs are used on non-computing related scientific domains including biology, ecology and social science [5]. This work is related to a number of literatures which study networks in a social/economic science context. The earliest steps in the agent-based models can be traced back to the Von Neumann machine, a theoretical machine capable of reproduction. The device von Neumann proposed would follow precisely detailed instructions to fashion a copy of itself. The concept was then improved by von Neumann's friend Stanislaw Ulam, also a mathematician. Ulam suggested that the machine be built on paper, as a collection of cells on a grid. The idea intrigued von Neumann, who drew it up—creating the first of the devices later termed cellular automata[6]. Another advance was introduced by the mathematician John Conway. He constructed the well-known Game of Life. Unlike von Neumann's machine, Conway's Game of Life operated by tremendously simple rules in a virtual world in the form of a 2-dimensional checkerboard [7].

One of the earliest agent-based models in concept was Thomas Schelling's segregation model, which was discussed in his paper "Dynamic Models of Segregation" in 1971. Though Schelling originally used coins and graph paper rather than computers, his models embodied the basic concept of agent-based models as autonomous agents interacting in a shared environment with an observed aggregate, emergent outcome[8].

There is an extensive literature on modeling social networks from a sociological perspective covering issues ranging from the interfamily marriage structure in 15th century Florence to the communication patterns. Occasional contributions to microeconomic theory have used network structures for such diverse issues as the internal organization of firms, systems compatibility, information transmission, and the structure of airline routes. There is a formal game theoretic literature which includes the marriage problem and its extensions, games of flow, and games with communication structures. Finally, the operations research literature has examined the optimization of transportation and communications networks[9]. One area of that research studies the allocation of costs on minimal cost spanning trees and makes explicit use of cooperative game theory.

Agent based modelling and simulation today can be accomplished by using a desktop computer, computing clusters, or clusters on computational Grids depending on the number of interacting agents and complexity of the model. Typically, desktop agent based models do not scale to what is

required for extremely large applications in the study of realistic complex systems. At one of the TEDx events UVM 2011 - Rob Axtell made a presentation on his work- Modeling the Economy with 150 Million Agents. Rob Axtell works at the intersection of economics, behavioral game theory, and multi-agent systems computer science. His most recent research attempts to emerge a macro economy from tens of millions of interacting agents. He is Chair of the Dept. of Computational Social Science at George Mason University and External Professor at the Santa Fe Institute[10].

IV. ECONOMICAL AND SOCIAL MODELS

ABMs are especially used to study complex systems. With complex systems we refer to systems that use simple micro-level rules that generate macro-level phenomena. These emergent phenomena can't be explained by the micro-level units alone. The interactions of the units lead to a nonlinear transformation to macro-level phenomena. For example, we can't understand the emergence of ant-colonies by studying one ant. Complex systems are different from complicated systems. Complex systems are in fact simplified versions of very complicated systems. Scholars studying complex systems and complexity are interested in discovering the basic underlying rules that describe most of the phenomena, not the details for specific empirical applications[11].

ABMs in economics and social sciences

Prior to, and in the wake of the financial crisis, interest has grown in ABMs as possible tools for economic analysis [12,13]. ABMs do not assume the economy can achieve equilibrium and "representative agents" are replaced by agents with diverse, dynamic, and interdependent behavior including herding. ABMs take a "bottom-up" approach and can generate extremely complex and volatile simulated economies. ABMs can represent unstable systems with crashes and booms that develop out of non-linear (disproportionate) responses to proportionally small changes [14].

A July 2010 article in *The Economist* looked at ABMs as alternatives to DGSE models [15]. The journal *Nature* also encouraged agent-based modeling with an editorial that suggested ABMs can do a better job of representing financial markets and other economic complexities than standard models [16] along with an essay by J. Doyne Farmer and Duncan Foley that argued ABMs could fulfill both the desires of Keynes to represent a complex economy and of Robert Lucas to construct models based on micro foundations [17] Farmer and Foley pointed to progress that has been made using ABMs to model parts of an economy, but argued for the creation of a very large model that incorporates low level models [18].

Since the beginning of the 21st century ABMs have been deployed in architecture and urban planning to evaluate design and to simulate pedestrian flow in the urban environment [19].

Agent-based models have been used since the mid-1990s to solve a variety of business and technology problems. Examples of applications include the modeling of organizational behavior and cognition [20], team working [21], supply chain optimization and logistics, modeling of consumer behavior, including word of mouth, social network effects, distributed computing, workforce management, and portfolio management. They have also been used to analyze traffic congestion [22]

V. CLASSES OF SOFTWARE TOOLS

Having in mind previous attempts and approaches to solve economic problems e.g. macroeconomic, we can spot several methods and classes of software. One is clear simple theoretical approach with few formulas and diagrams that represent static or sometimes dynamic relationships between correlating variables (GDP, employment, production etc.). Secondly, there are empirical forecasting models that use time series analysis - regression analysis, for building large scale macro econometric models, third would be Dynamic stochastic general equilibrium models (DSGE), with one notable branch computable general equilibrium (CGE) and lately Agent-based computational macroeconomic models. All of these, except the first one, use and make computers power essential for solving those models.

Following the development of Keynesian economics, applied economics began developing forecasting models based on economic data including national income and product accounting data. In contrast with typical textbook models, these large-scale macro econometric models used large amounts of data and based forecasts on past correlations instead of theoretical relations. These models estimated the relations between different macroeconomic variables using regression analysis on time series data. These models grew to include hundreds or thousands of equations describing the evolution of hundreds or thousands of prices and quantities over time, making computers essential for their solution. While the choice of which variables to include in each equation was partly guided by economic theory (for example, including past income as a determinant of consumption, as suggested by the theory of adaptive expectations), variable inclusion was mostly determined on purely empirical grounds. Large-scale macro econometric model consists of systems of dynamic equations of the economy with the estimation of parameters using time-series data on a quarterly to yearly basis.

Macro econometric models have a supply and a demand side for estimation of these parameters. Kydland and Prescott call it the system of equations approach [23].

In the Central Bank-sphere internationally, as in Macedonia, the following software's are quite popular: EViews [24], MATLAB/Dynare [25], TROLL, RATS, and R Studio. Basically, that is a software for large-scale econometric models, some of the national banks use DSGE models, time-series models (SVARs, various state-space models, etc.), and Bayesian techniques. The Federal Reserve Bank of New York (FRBNY) has built a DSGE model as part of its efforts to forecast the U.S. economy. They recently moved its code-base from Matlab to Julia[26]. Other, FRB in US use models in EViews. DSGE models are also quite abstract representations of reality, however, which in the past severely limited their empirical appeal and forecasting performance. This started to change with work by Schorfheide [27] and Smets and Wouters [28]. First, they popularized estimation (especiallyBayesian estimation[29]) of these models, with parameters chosen in a

way that increased ability of these models to describe the time series behavior of economic variables. Next, these models were enriched with both endogenous and exogenous propagation mechanisms that allowed them to better capture patterns in the data. For this reason, estimated DSGE models are increasingly used within the Federal Reserve System (the Board of Governors[30] and the Reserve Banks of Chicago [31] and Philadelphia [32] have versions) and by central banks around the world (including the New Area-Wide Model [33] developed at the European Central Bank, and models at the Norges Bank [34] and the Sveriges Riksbank [35]). The FRBNY DSGE model is a medium-scale model in the tradition of Christiano, Eichenbaum, and Evans[36] and Smets and Wouters that also includes credit frictions as in the financial accelerator model developed by Bernanke, Gertler, and Gilchrist [37] and further investigated by Christiano, Motto, and Rostagno [38]. TROLL is used at Bank of Canada.

VI. SOFTWARE ANALYSIS

There are a lot of software packages and libraries that have been developed in recent years addressing agent based modeling issues. Currently on Wikipedia, there is a list of 89 such software packages [39], compared on few basic attributes. Also, on Leigh Tesfatsion's Web [40] site and AgentLink EU project [41] site, we find even more lists of software, some of them highly specialized for ABM research in particular fields, as for example a review of ABMs packages relevant for geo-spatial analysis [42]. Also, some individuals have made attempts to compare this software to each other on different attributes [43,44,45,46].

Basic documentation for most of the packages reviewed is largely incomplete. As an example, the Swarm community have a Web site and Wiki which contains a lot of information but is hard to navigate. The Repast development program is highly productive and helpful to the scientific community, but tidying up and fully documenting all the packages in their current form would be very useful. There are indeed a number of challenges which must be faced in order to make ABMs a mainstream computational science technology.

Most of the commonly used ABM platforms follow the ``framework and library" paradigm, providing a framework, a set of standard concepts for designing and describing ABMs, along with a library of software implementing the framework and providing simulation tools. This group starts with Swarm, the libraries of which were written in Objective-C, and also Java counterpart of Swarm, which allows use of Swarm's Objective-C library from Java. Repast is a fork that started as a Java implementation of Swarm but has since diverged significantly from Swarm. More recently, MASON has been developed as a new Java platform. This group of software's requires extensive knowledge of the platform, making learning curve and startup of using them by individual researcher pretty hard, even if the researcher is a person with very good IT skills. Indeed, without a proper tool programming agent behaviors with frameworks and libraries is a tedious, time consuming task involving multiple code, test, and debug cycles and creating agents behaviors requires significant programming ability. The reason that we didn't excluded them in research is that they were/are largely used by universities and therefore still popular.

Most of the ABM software's were and still are initiated by necessities and work at universities and scientific community purely driven by a quest to make or find suitable tool, for easy research, explore and survey different phenomena and scenarios, although ABM gain lot of attraction and interest in current years by commercial companies, that also saw a potential in ABM for solving business like tasks. So, no wonder most of the platforms are open source, dominantly based on Linux and Java as programing language, and some of them which are more advanced are proprietary with an option for a free personal or educational editions. The main reason that some ABM packages succeeded, and this has to be a future norm for any such software, is because they provided standardized software designs and tools without limiting the kind or complexity of models that can be implemented. A review of Java Swarm and Repast (along with two less used platforms) ranked them numerically according to well defined criteria [47]. Criteria were evaluated from documentation and other information about each platform. The review indicated important weaknesses including: difficulty of use; insufficient tools for building models, especially tools for representing space; insufficient tools for executing and observing simulation experiments; and a lack of tools for documenting and communicating software. The review by Railsback [48] lists five ABMS platforms: Mason, NetLogo, Repast and Swarm in its Objective-C and Java flavors. This work effectively sets a benchmark for ABMS packages based on 16 flavors of what the authors called StupidModel.

From all the previous research and papers on this field, that we examined, it's obvious that Swarm was lauded as probably the most powerful and flexible simulation platform, even though, some researchers noted that it comes with a price. In practice, Swarm has a very steep learning curve, so the only pre-requisite for modeling some phenomena is finding a programmer experienced enough to be able to implement what was needed [49]. It is necessary to have experience of Objective-C and possibly Java, be familiar with the object orientation methodology and be able to learn some Swarm code. Swarm models can run inside a Web browser, if JAVA applets are supported and enabled (currently they can't in Chrome at least). Swarm runs on any platform which has Objective-C.

As Charles M. Macal and Michael J. North noted [50], we distinguish several approaches to building ABMS applications in terms of the scale of the software that one can apply according to the following criteria's:

Desktop Computing for ABM Application Development:

- Spreadsheets: Excel using the macro programming language VBA , AgentSheets etc
- Dedicated Agent-based Prototyping Environments: Repast Simphony, NetLogo, StarLogo
- General Computational Mathematics Systems: MATLAB/Dynare, Mathematica
- Large-Scale (Scalable) Agent Development Environments: Ascape -Sugarscape, Repast, Swarm, MASON, AnyLogic

From all the information that we researched and gathered on-line and by testing some of the most appealing ABM software, we have concluded and narrowed our focus to only one, that has not been properly researched, evaluated and compared to other popular ABM software's in the field. That is the Anylogic software that we currently believe to be the Avant-guard of ABM packages, with particular relevance to scientific modelling and simulation for an individual researchers. AnyLogic-7.3 is a proprietary software that incorporates a range of functionality for the development of agent based models. Models can dynamically read and write data to spreadsheets or databases during a simulation run, as well as charting model output dynamically. Furthermore, external programs can be initiated from within an AnyLogic model for dynamic communication of information, and vice versa. AnyLogic models can run on Linux, Windows and OS.

The criteria's that filtered AnyLogic as currently best ABM software was documentation, samples included with installation (or easily accessible online), flexibility / scalability, intuitiveness of GUI and IDE, and popularity of used programing language. Lot of the ABM software's previously mentioned were and still are concepts, libraries, frameworks, and lot of them without active support and development, some with lack of documentation and versions that are built more than 5 years ago, so they were the first one, which we excluded from further analysis. Anylogic is a complete product. Once installed (which goes straight-trough), the IDE is so intuitive, you have feeling you are fully in control and everything is as it is supposed to be, from individual ABM researcher perspective.

Included example models with installation of version 7.3 - currently 95 from more than 14 areas (ranging from social, urban and eco dynamics, ecosystem dynamics e.g. a predator prey system, planning of healthcare schemes e.g. the impact of safe syringe usage on HIV diffusion, computer and telecommunication networks e.g. the placement of cellular phone base stations, the location of emergency services and call centers, pedestrian dynamics, to simulating air-defense system), 123 models from the book "The Big Book of Simulation", 67 how to models and 48 models from book "Business dynamics". Also, accessible on internet through RuntheModel.com there are even more than 750 (although not all of them with source code). Integrated help within application, has also good quality and organization, and also on their website, there are 4 e-books that helps and make step by step learning of modeling with AnyLogic a lot easier. Also they have organized a great training program, with options for on-site training and testimonials of user experience. They have posted a great case studies, educational videos, organized and scheduled training courses, active blogs and groups of users at social networks Facebook, Linkedin, Youtube. There are more than 100 scientific papers posted on their site, about research made with AnyLogic software, all of them easily accessible on their website. Video tours, consulting services, academic partnership, workshops and conferences are proof that Anylogic is aimed at getting an acknowledgement as a mainstream ABM tool surpassing all other ABMs by far in recent years. Anylogic haven't been properly analyzed and compared in papers with other ABM software's in recent years, which gap we want to fill with our research. On their website they immediately declare that Anylogic is the only simulation tool today that supports Discrete Event, Agent Based, and System Dynamics Simulation. Our extensive online research to find some contra arguments yielded in no evidence that this declaration it is not correct.

The Anylogic programing paradigm is object oriented that is JAVA based, which allows reusing same properties and methods of once developed classes of objects. The libraries that comes with the installation are quite rich having out of the box – usable objects in many different areas.

The software supports connecting reading and writing data from and to external databases. It can find and correlate different variables, it can simulate different scenarios, and make automatic experiments based on different set of startup conditions, and all those results can be saved to database for further analyses. 2D and 3D models and presentation are fully supported even though we didn't tried to create and run 3D model ourselves. Included samples were more than enough to convince us, that this tool is not limited in features and properties, for modeling even most complicated and complex problems.

Their reference list shows more than 800 world scientific institutions use Anylogic as a tool for their research. The number and names of companies (some of the biggest companies in the world) that are users of this software also gives impression that this is a mature product with large number and diverse types of users.

VII. Conclusion

This research focused on analysis of existing software packages available for use at this moment, for modeling nonlinear social/economic systems with a limited number of social/economic actors. The aim of research is to select suitable software package, tool, with which we can effectively, precisely and quickly model nonlinear systems, for example. social networks or certain economic community

For an independent individual researcher we highly recommend usage of AnyLogic, because it have competitive

advantages in easiness, its personal edition is a freeware, it have great documentation, its Java based, it has great interface and great number of diverse samples included in installation to enable effective and quick start of the researcher. It's flexible scalable, and fit to answer almost every modern simulation need before individual researcher. Currently we can't mention any scientific problem in nonlinear systems that can't be modeled with AnyLogic.

REFERENCES

[1]STELLA, http://www.iseesystems.com/softwares/stella-pro/v1.aspx
[2]Aspentech, Aspen Economic Evaluation Family, 2014, http://www.aspentech.com/products/economic-evaluation/aspen-process-economic-analyzer
[3]https://www.openabm.org/faq-page#t780n3730
[4]Grimm, Volker; Railsback, Steven F. (2005). Individual-based Modeling and Ecology. Princeton University Press. p. 485.ISBN 978-0-691-09666-7.
[5]Niazi, Muaz; Hussain, Amir (2011). "Agent-based Computing from Multi-agent Systems to Agent-Based Models: A Visual Survey" (PDF). Scientometrics (Springer) 89 (2): 479–499. doi:10.1007/s11192-011-0468-9.
[6]https://en.wikipedia.org/wiki/Agent-based_model
[7]https://en.wikipedia.org/wiki/Agent-based_model
[8]Schelling, Thomas C. (1971). "Dynamic Models of Segregation" (PDF). Journal of Mathematical Sociology 1 (2): 143–186. doi:10.1080/0022250x.1971.9989794.
[9]Bhaskar Dutta, Matthew O. Jackson, "Networks and Groups: Models of Strategic Formation", Springer Science & Business Media, Mar 9, 2013
[10]https://www.youtube.com/watch?v=c-sieJVR5TI
[11]https://www.openabm.org/faq-page#t780n3732
[12]Testfatsion, Leigh; Judd, Kenneth, eds. (May 2006).Handbook of Computational Economics 2. Elsevier. p. 904.ISBN 978-0-444-51253-6.
[13]Page, Scott E. (2008). Agent-Based Models. The New Palgrave Dictionary of Economics (2 ed.).
[14]Agents of change". The Economist. July 22, 2010.
[15]"Agents of change". The Economist. July 22, 2010.
[16]"A model approach" (Editorial). Nature 460 (7256): 667. August 6, 2009. Bibcode:2009Natur.460Q.667.doi:10.1038/460667
[17]Farmer & Foley 2009, p. 685.
[18]Farmer & Foley 2009, p. 686
[19]Aschwanden, G.D.P.A; Wullschleger, Tobias; Müller, Hanspeter; Schmitt, Gerhard (2009). "Evaluation of 3D city models using automatic placed urban agents". Automation in Construction 22: 81–89. doi:10.1016/j.autcon.2011.07.001.
[20]Hughes, H. P. N.; Clegg, C. W.; Robinson, M. A.; Crowder, R. M. (2012). "Agent-based modelling and simulation: The potential contribution to organizational psychology". Journal of Occupational and Organizational Psychology 85 (3): 487–502.doi:10.1111/j.2044-8325.2012.02053.x.
[21]Crowder, R. M.; Robinson, M. A.; Hughes, H. P. N.; Sim, Y. W. (2012). "The development of an agent-based modeling framework for simulating engineering team work". IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans 42 (6): 1425–1439.doi:10.1109/TSMCA.2012.2199304
[22]"Application of Agent Technology to Traffic Simulation".United States Department of Transportation. May 15, 2007.
[23]Kydland, Finn E.; Prescott, Edward C. (1991). "The Econometrics of the General Equilibrium Approach to Business Cycles". The Scandinavian Journal of Economics 93 (2): 161–178. JSTOR 3440324
[24]http://www.eviews.com/
[25]http://www.dynare.org/more-resources/bibliography
[26]http://libertystreeteconomics.newyorkfed.org/2015/12/the-frbny-dsge-model-meets-julia.html#.Vwg4_fl95D9
[27]http://www.jstor.org/stable/2678564
[28]https://www.aeaweb.org/articles.php?doi=10.1257/aer.97.3.586
[29]http://books.google.com/books?id=da0mve_vB5IC&printsec=frontcover&dq=Handbook+of+Bayesian+econometrics&hl=en&sa=X&ei=7qHBU7-ACMOvyAS814Eg&ved=0CB4Q6AEwAA
[30]http://www.federalreserve.gov/econresdata/edo/edo-models-about.htm
[31]http://www.chicagofed.org/webpages/publications/working_papers/2012/wp_02.cfm
[32]https://www.philadelphiafed.org/research-and-data/real-time-center/PRISM/
[33]https://www.ecb.europa.eu/pub/pdf/scpwps/ecbwp1145.pdf?e3511df65c93ce87e656eff7fc2393e2
[34]http://www.norges-bank.no/Upload/66495/EN/NEMO.pdf
[35]http://www.riksbank.se/en/Press-and-published/Published-from-the-Riksbank/Other-reports/Occasional-Paper-Series/2013/Occasional-Paper-Series-No-12-Ramses-II--Model-Description/
[36]http://benoitmojon.com/pdf/Christiano%20%20Eichenbaum%20Evans%202005%20JPE.pdf
[37]http://faculty.wcas.northwestern.edu/~lchrist/course/Czech/BGG%201999%20Handbook%20chapter.pdf
[38]http://www.aeaweb.org/articles.php?doi=10.1257/aer.104.1.27
[39]https://en.wikipedia.org/wiki/Comparison_of_agent-based_modeling_software
[40]http://www.econ.iastate.edu/tesfatsi/acecode.htm.
[41]http://eprints.agentlink.org/view/type/software.html
[42]http://www.spatialanalysisonline.com/output/html/Simulationmodellingsmsystemsforagent-basedmodelling.html
[43]Nikolai, C.; G. Madey (2008). "Tools of the Trade: A Survey of Various Agent Based Modeling Platforms"
[44]Salamon, T. (2011). Design of Agent-Based Models : Developing Computer Simulations for a Better Understanding of Social Processes
[45]Castle, C; A. Crooks (2006). "Principles and Concepts of Agent-Based Modelling for Developing Geospatial Simulations". Working paper 110 (University College London).
[46]Au. R.J. Allan (2009) Survey of Agent Based Modelling and Simulation Tools
[47]http://www.grids.ac.uk/Complex/ABMS/ABMS.html#TH2004
[48]http://www.grids.ac.uk/Complex/ABMS/ABMS.html#RLJ2005
[49]http://www.grids.ac.uk/Complex/ABMS/
[50]http://www.informs-sim.org/wsc09papers/009.pdf

Using Virtual Reality to Analyze Handball Goalkeeper Performance

Natasha Antikj
School of Computer Science and Information Technology
University American College Skopje
Skopje, Macedonia
natashaantikj@gmail.com

Adrijan Bozinovski
School of Computer Science and Information Technology
University American College Skopje
Skopje, Macedonia
bozinovski@uacs.edu.mk

Veno Pachovski
School of Computer Science and Information Technology
University American College Skopje
Skopje, Macedonia
pachovski@uacs.edu.mk

Abstract— Virtual reality is now widely used in many domains such as video games and other applications for entertainment, but also in serious applications, such as analyzing the behaviors of subjects or improvement of training skills in highly controlled environments.
There have been some experiments with virtual reality technologies used in handball, specifically with handball goalkeepers. In order to improve a goalkeeper's training by allowing insights into her/his complex perception, learning and action processes, virtual reality (VR) technologies provide a way to standardize experimental sport situations. However, few studies have been carried out in order to identify whether the subjects reacted the same way in virtual reality as they do in the real world.
The purpose of this paper is to verify if the movements performed by the handball goalkeepers in the real and virtual environment are similar and whether virtual reality could be used in sports as a training tool.

Keywords-***virtyal reality; sport; handball; goalkeeper***

I. INTRODUCTION

Virtual can be a valid training tool for sports. Sports like handball require a set of cognitive skills of players, such as anticipation, techniques, decision-making and also social behavior. For them, improving these skills is mandatory, due to the competition of skilled performers. Therefore, in sport sciences as in others, controlled conditions for experiments are needed to gain reproducible and reliable measurements of the performance of players. Virtual reality technologies provide a way to standardize experimental situations and are already used by researchers around the world [1].
The main goal of this research is to verify if the movements performed by the goalkeeper in the real and virtual environments are similar and, thus, if virtual reality could be used in sports as a training and research tool. In this paper the experiments by Bolte et al. [1] and Bideau et al. [2] will be described, which target handball goalkeepers. Goalkeeper training has already been implemented on an artificial intelligence level [3].

II. VR EXPERIMENTS

A. *The Bolte et al. Experiment*

1) *Apparatus*

B. Bolte, F. Zeidle, G. Bruder, F. Steinicke, K. Hinrichs, L. Fischer and J. Schorer [1] used a pre-recorded video snapshot of a shooter aiming one of the corners of the goal and a handball goalkeeper equipped with a light-weight Carl Zeiss Cinemizer Plus head-mounted display. A low-cost Philips SPC1030 USB webcam with 60 frames per second (FPS) was used for markerless computer vision based detection of the goalkeeper's reactions and motions. They captured the goalkeeper's body in front of uniformly colored patches of rectangular shape, which were attached to a goal-sized wall or goalposts (Figure 1(a)). And for the visual stimulus presentation and the goalkeeper analysis system they used an Intel computer with Core2 Duo processor, 4GB of main memory and nVidia GeForce G210M graphics card [1].

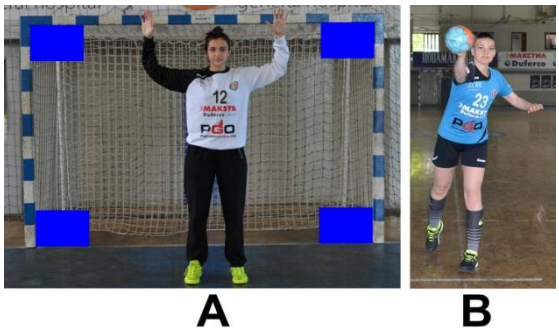


Figure 1 - (a) Handball goalkeeper (b) Shooter aiming at one of the corners of the goal – similar to those of the Bolte et al experiment

2) Goalkeeper Reaction Analysis

The goalkeeper reaction process was divided into 3 parts, a calibration, visual stimulus and a detection phase. In the calibration phase, target points were detected.
In the visual stimulus phase, pre-recorded videos of a shooter aiming at one of the corners of the goal (Figure 1(b)) was presented to the goalkeeper.
The detection phase begins after the visual stimulus ends and in this phase successful saves were counted [1].

B. The Bideau et al. Experiment

1) Apparatus

To ensure that the goalkeeper could move freely, Bideau et al. used an immersive system which included three Barco 1208s synchronized video projectors, driven by Silicon Graphics' Onyx2 InfiniteReality, projecting the 3D sports hall onto a large cylindrical screen (3.80-m radius, 2.38-m height, and 135° field of view). They attained 3D vision using stereoscopic glasses activated at 60 Hz (30 Hz for each eye).
To improve the subject's presence, goalkeepers stood in a real goal in front of a screen, which was in the same position it would be in the virtual stadium (Figure 2) [2].

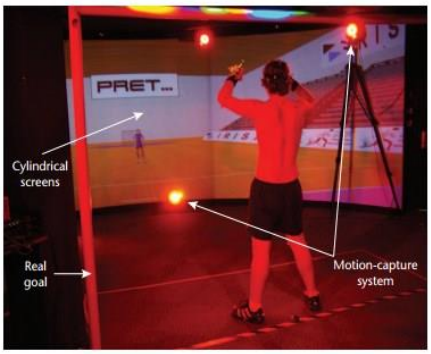


Figure 2 - A real goalkeeper facing a virtual thrower in the handball case study by Bideau et al. [2]

2) Procedure

Bideau et al. asked the goalkeeper to move as if he was stopping a ball in a real match situation, and they recorded his movements during each trial. They gave each participant a training period to become familiar with the environment and task. After this period, he had to stop 50 virtual throws, presented in random order [2].

3) Data Analysis

Bideau et al. evaluated anticipation skill through success rates and response times. They considered a movement successful when, according to motion-capture data, at least one part of the goalkeeper's body collided with the virtual ball—giving an overall success rate for each throw in each zone. Successful movements were characterized by faster response times (Figure 3)[2].

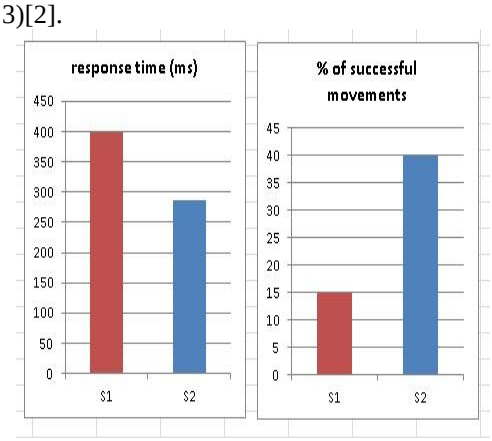


Figure 3 - Results for the handball case study by Bideau et al.: (a) response time for two subjects S1 and S2 and (b) their percentage of successful movements (Diagram created by authors)

III. Conclusion and Future Work

A computer-generated visual environment lets experimenters control factors that would otherwise be extremely difficult, if not impossible, to control in the real world. But there are some things in real life that can influence a goalkeeper's perception that virtual reality can't control, such as defense players which are of great help for the goalkeepers, the lightning in the sports hall, the temperature, the audience and so on. Therefore, the experiments explained in this paper are just approximations of what a real-life goalkeeper might expect during training or a match.
In the future, the authors plan on making such an experiment as Bolte or Bideau did, collaborating with Macedonian handball goalkeepers and making a positive impact on Macedonian handball. Handball goalkeepers and shooters volunteers will be needed to participate in the experiment. Also, help from developers and virtual reality experts will be needed for designing the VR arena, animating the throwing actions and for capturing the movements performed by the goalkeepers.

References

[1] B. Bolte , F. Zeidler , G. Bruder , F. Steinicke , K. Hinrichs , L. Fischer and J. Schorer: A Virtual Reality Handball Goalkeeper Analysis System. Joint Virtual Reality Conference (JVRC) (Poster Presentation), page 1--2 , 2010

[2] B. Bideau, R. Kulpa, N. Vignais, S. Brault, F. Multon: Using virtual reality to analyze sports performance, IEEE Computer Graphics and Applications, page 14–21, 2010

[3] Bozinovski S., Jaeger H., Schoel P. "Engineering goalkeeper behavior using emotion method", In S. Sablatnoeg, S. Enderle (eds.) Proceedings of RoboCup Workshop '99. Deutsche Jahresgagung fuer Kuenstliche Intelligenz (KI99), Bonn, pp 48-56, 1999

A Concept for an Intelligent Transportation System for Managing Traffic Lights and Optimal Road Control

Bogdan Jeliskoski, Adrijan Božinovski, Ivana Stojanovska
School of Computer Science and Information Technology
University American College Skopje
Skopje, Macedonia
bogdan.jeliskoski@live.com bozinovski@uacs.edu.mk ivana.stojanovska@uacs.edu.mk

***Abstract*— Traffic chaos in the urban areas is a well-known problem: long timing cycles between the traffic lights, everyday increase of the traffic jam and, most importantly, the traffic accidents occurring on a daily basis. In the areas where increasingly more vehicles are noticed, implementing a smart and an adaptive system for road control is necessary. In this paper, a concept for an intelligent transportation management system is proposed, based on cluster computing and traffic prediction. This system should improve the safety of the drivers, and the efficiency of the transport, and generally benefit the population.**

Keywords*—Traffic light control; cluster computing; traffic prediction; road control; intelligent transportation; crash prevention.

I. INTRODUCTION

According to *The National Council for Traffic Safety Roads of Republic of Macedonia*, in the year 2015 the total death toll in traffic accidents on the roads in Republic of Macedonia was 144, which is a significant increase compared to the year 2014 [1].

The timing of the traffic lights in the transportation system is the most important feature for safe driving and optimal road control. The increase in the traffic congestion in the urban areas lead to the emergence of innovative solutions for management of the traffic flow.

Using cloud computing based on cluster infrastructure as a basis, a traffic prediction model for managing traffic lights can be designed. Cloud computing is one of the best known service for the transfer of various resources, software, design platforms and combined data [2].

Cluster computing is used for Internet service workloads, because clusters are working in parallel. For these workloads, large clusters can be more power-efficient, performing even better than the largest machines [3].

A very robust fault tolerance and error correction system will be needed for the structure reliability, due to the sensitivity of the area where human lives are concerned.

With today's increased number of vehicles, an adaptive method for timing of the traffic lights is needed, because the traditional methods are becoming obsolete. To cope with this issues, use of fuzzy logic structure and fuzzy algebra is proposed.

Fuzzy logic can be used for implementation of human rules for controlling a specific junction. Human terms such as: longer, less and heavy, can be successfully carry out in the fuzzy algebra, and be understood by the computer [4].

II. CLUSTER-BASED INFRASTRUCTURE

High availability system is needed for the proposed traffic prediction model, with 24/7 availability. Cluster computing is capable of fast and secure processing of data [3].

Cluster infrastructure is highly reliable, modular and redundant both on software and hardware level. The independence of the clusters' nodes gives the ability of high redundancy. Also, use of "hot-upgrade" is advantage in this structure, meaning some modules can be replaced without shutting down the nodes [3].

Data mining and analyzing the traffic safety information can also be achieved with cluster-based infrastructure [5].

III. *HARDWARE REDUNDANCY FOR HIGH AVAILABILITY*

In cluster computing, predefined number of nodes are made to run as a single entity. The connection between nodes is with high speed local area connection networks. Modularity, performance and fault tolerance are the main reasons for cluster-based infrastructure to be used [6].

Experiments and testing have been conducted for the purpose of this paper on two important modules in cluster infrastructure, storage and memory.

As can been seen from Table I, testing of two enterprise storage devices has been conducted; traditional hard disk drive, model Samsung HD103UJ and a solid-state disk drive, namely Samsung SSD 840 Pro. The HD Tune Pro version 5.00 testing software was used for benchmarking with random access mode.

The results show that the solid-disk state drives are clearly superior to the traditional hard disk drives in every tested aspect.

The proposed basis for the traffic prediction model requires cluster based infrastructure with solid-state disks in every node of the structure, as they are much faster, more stable and more reliable from the traditional hard drives.

The need for high-speed memory is required for the cluster structure. The memory should be able to handle all of the given data, because the traffic prediction model will receive information in real time.

Testing of the enterprise models of today's fastest popular DDR3 memory architectures, quad and dual channel, is conducted. AIDA 64 Extreme Edition Version 5.20.3400

software is being used for testing. Table II shows the memory read test, and Table III shows the memory write test.

The memory working speed should not be the deciding factor when choosing the cluster structure's memory modules. From Table II. and Table III it can be seen that higher memory speed does not reflect the actual speed of reading and writing. Testing should be performed before installing the memory modules on the structure's nodes.

Table I. Testing traditional hard drive and solid-state drives

Random access on model		Samsung HD103UJ		
Transfer size	Operations/sec (IOPS)	Average access time (ms)	Max. access time (ms)	Average speed (MB)
512 Bytes	73	13.557	26.241	0.036
4 KB	73	13.543	27.842	0.288
64 KB	70	14.198	27.701	4.402
1 MB	36	27.204	43.290	36.759
Random	48	20.557	86.249	24.682
Random access		Samsung SSD 840 PRO		
Transfer size	Operations/sec (IOPS)	Average access time (ms)	Max. access time (ms)	Average speed (MB)
512 Bytes	14496	0.068	0.128	7.7078
4 KB	12287	0.081	1.168	47.999
64 KB	4891	0.204	1.005	305.698
1 MB	497	2.011	2.880	497.261
Random	927	1.078	2.384	470.510

Table II. Table of results obtained from the AIDA 64 software, showing that higher memory speed does not reflect the actual memory reading speed

CPU Clock (MHz)	Motherboard	Chipset	Memory (DDR3) (MHz)	CL-RCD-RP-RAS	Read Speed (MB/s)
2600	Supermicro XDDR6-F	C600	Quad-1333	9-9-9-24	77243
2200	Supermicro X8DGI-F	SR5690	Dual-1600	11-11-11-28 CR1	67619
3400	Gigabyte GA-X79-UD3	X79	Quad-1866	9-10-9-27-CR2	52396
3800	Asus P9X79 Deluxe	X79	Quad-2400	11-30-30-30 CR2	47508

Table III. Write test performed with AIDA 64 Extreme Edition Version 5.20.3400

CPU Clock (MHz)	Motherboard	Chipset	Memory (DDR3) (MHz)	CL-RCD-RP-RAS	Write Speed (MB/s)
2600	Supermicro XDDR6-F	C600	Quad-1333	9-9-9-24	77060
2200	Supermicro X8DGI-F	SR5690	Dual-1600	11-11-11-28 CR1	57844
3800	Asus P9X79 Deluxe	X79	Quad-2400	11-30-30-30 CR2	53524
3400	Gigabyte GA-X79-UD3	X79	Quad-1866	9-10-9-27-CR2	52243

IV. FUZZY LOGIC SIGNAL CONTROL METHOD

An adaptive method is needed for the most accurate traffic flow. The conventional methods for managing the traffic lights control cannot keep up with the ever increasing number of vehicles.

The traditional traffic light controller provides predefined times for the traffic lights cycles. This results in traffic-wise unfavorable use of red and green lights, which in turn leads to congested roads and lanes.

Time of switching and skipping the states of the traffic lights is achieved with a fuzzy logic controller according to the traffic load. The number of waiting and arriving vehicles are allotment into different fuzzy logic variables [7].

The basic function of the fuzzy rule is an IF-THEN rule, combined with "AND" and "OR" operators [8].

A set of fuzzy rules can be done for the intersection, as can be seen from Fig.1, following this simplified algorithm:

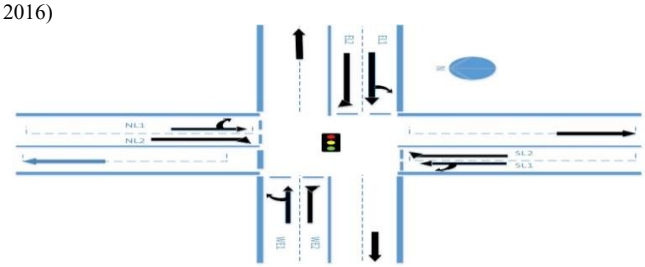


Figure 1. A traditional intersection with four phases and eight lanes (diagram recreated by authors)

```
procedure checkForCars(direction)
{
    if(numberOfCars(direction) == many)
    {
        lights(green, direction);
    }
    else if(lights(green, direction))
    {
        lights(yellow, direction);
        waitSeconds(5);
        lights(red, direction);
        checkForCars(nextDirection(CW));
    }
}
```

In the algorithm presented, the value *many* is the integer value, calculated with a specified set of rules [9].

V. VEHICLES DETECTION

Video camera image devices placed on the intersections will be is used for detecting and counting vehicles. The proposed approach to vehicle detection is by parts. Strong classifiers and trained active learning are used for detecting front and rear parts of vehicles. Then, the detected parts are paired using "Support Vector Machines". The final result is detected combination by parts [10]. Fig. 2 shows how the actual detection looks like. Using vehicle detection by parts, accurate predictions can be made.

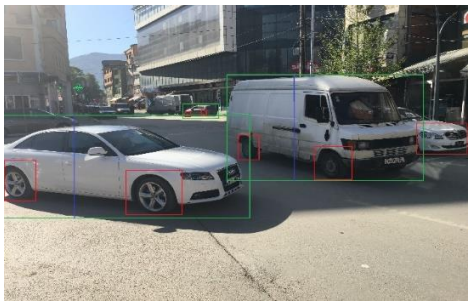


Figure 2. Detecting vehicles by parts

V. CONCLUSION

In this research paper we have presented a concept for an intelligent transportation system for managing the traffic lights for improving the traffic congestion and safety in the Republic of Macedonia. Our approach is with a stable, reliable and low cost cluster-based infrastructure as a basis, while using cloud computing for managing, controlling and analyzing the video image processors, and also using fuzzy logic controllers. Detection of the vehicles is accomplished with video camera image processors, with detection by parts. This is done for improved and more accurate detection.

Such a system would provide a huge step towards smart managing traffic lights, and should be implemented for common good, where the citizens would benefit the most.

REFERENCES

[1] "*National Council for Traffic Safety Roads of Republic of Macedonia"* accessible at http://www.rsbsp.org.mk/ (last accessed on 25 January, 2016)

[2] Alifia Fithritama and Ramya Narayanan *"Intelligent Transportation System (ITS) using Cloud Computing"* accessible at http://sdn.ifmo.ru/education/perccom/cluster-grid-clouds/2014/students-presentations/Intelligent%20Transportation_Alifia_Ramya.pdf (last accessed on 24 January, 2016)

[3] Armando Fox, Steven D Gribble, Yatin Chawath, Eric A. Brewer, Paul Gauthier *"Cluster-Based Scalable Network Services*" accessible at http://courses.cs.washington.edu/courses/cse454/13wi/papers/cluster_scalable.pdf (last accessed on 25 January, 2016)

[4] Milan Koukol, Lenka Zajíčková, Lukáš Marek and Pavel Tuček *" Fuzzy Logic in Traffic Engineering: A Review on Signal Control"* Hindawi Publishing Corporation, Mathematical Problems in Engineering Volume 2015, Article ID 979160, 14 pages accessible at http://dx.doi.org/10.1155/2015/979160 (last accessed on 1 October, 2015)

[5] Xian Xiaobing, Bao Chao and Chen *Feng "An insight into traffic safety management system platform based on cloud computing"* Procedia - Social and Behavioral Sciences 96 (2013), pp. 2643 – 2646 [13th COTA International Conference of Transportation Professionals, 2013]

[6] Kiranjot Kaur and Anjandeep Kaur Rai *"A Comparative Analysis: Grid, Cluster and Cloud Computing"*, [International Journal of Advanced Research in Computer and Communication Engineering Vol. 3, Issue 3 March 2014]

[7] Marco Wiering, Jelle van Veenen, Jilles Vreeken and Arne Koopman *"Intelligent Traffic Light Control",* accessible at http://www.cs.uu.nl/research/techreps/repo/CS-2004/2004-029.pdf (last accessed on 27 October, 2009)

[8] Madhavi Arora and V.K. Banga *"Real Time Traffic Light Control System Using Morphological Edge Detection and Fuzzy Logic",* 2012 [2nd International Conference on Electrical, Electronics and Civil Engineering (ICEECE'2012) Singapore April 28-29, 2012]

[9] Javed Alam and Pandey *"Design and Analysis of a Two Stage Traffic Light System Using Fuzzy Logic",* [J Inform Tech Softw Eng 2015], accessible at http://dx.doi.org/10.4172/2165-7866.1000162 (last accessed on 09 February 9, 2016)

[10] Sayanan Sivaraman and Mohan M. Trivedi *"Real-Time Vehicle Detection Using Parts at Intersections",* [2012 15th International IEEE Conference on Intelligent Transportation Systems Anchorage, Alaska, USA, September 16-19, 2012]

9 786084 699064