3D Simultaneous Positioning and Mapping in Dark, Closed Spaces with an Autonomous Flying Robot

Abdülkadir Çakır[1], Seyit Akpancar[2]

[1]Department of Electrical and Electronics Engineering, Faculty of Technology, Isparta University of Applied Sciences, Isparta, Turkey
E-mail: abdulkadircakir@isparta.edu.tr

[2]Department of Computer Programming, Atabey Vocational School, Isparta University of Applied Sciences, Isparta, Turkey
E-mail: seyitakpancar@isparta.edu.tr

Abstract: This study describes the steps needed to produce the required hardware and software for the 3D mapping of dark, rugged, and closed spaces such as caves, underground cities and mining pits through a DJI Matrice 100 Flying Robot Platform that has a high bearing capacity of 3600 grams and has no limit of movement in bumpy, hollowed or sloping spaces. In order to obtain the obstacle information around, during the autonomous movement of the air robot used within the scope of this study, 5 ultrasonic sensors – right, left, front, top, bottom – were used. A servomotor driven electromechanical equipment that will be used on the z-axis movement of Hokuyo UST-20LX laser sensor, which provides data in 2D, was developed to help the air robot map its environment in 3D during the autonomous movement. The control of the hardware developed and used in this study is carried out by Robot Operating System (ROS) nodes written in C++ programming language. The mapping studies were carried out by operating the robot autonomously in caves within Atabey District of Isparta Province, Turkey, at the coordinates of 37°53'41.8"N 30°32'58.5"E and 37°53'39.0"N 30°32'42.3"E. It is shown that the 3D maps produced by the system, are realistic and substantial.

Keywords: flying robot; 3D; simultaneous; mapping; dark spaces; closed spaces

1 Introduction

The mapping of space, such as caves or underground cities of an unknown shape, is mostly done using traditional methods, such as, takings manual measurements. However, when it is dangerous for people to enter a closed area, either mapping is not performed or performed by taking risks and entering the closed area.

In the mining accident, which happened in Turkey, on May 13, 2014 and took part in literature as the "Soma Mine Disaster", rescue activities were hampered by gas from the fire in the mine and by the lack of a known plan for the mine [1]. The importance of using autonomous robots actively emerges in such events.

Within the scope of this study, a fast and low-cost system which can operate autonomously in dark, rugged, and closed spaces such as caves, underground cities, and mining pits without any movement limits (bumpy, hollowed, sloping spaces); which can carry out 3D mapping by using 2D laser sensors; and which has a high accuracy level and can collect multipurpose data, was developed.

All the stages of this study were explained in main sections of this paper as Related Work, Description of the System and Conclusions.

2 Related Work

Robots need 3D data (X_O, Y_O, Z_O) of the robot's surroundings to map their surroundings in 3D, and 3D location information (X_L, Y_L, Z_L) of the robot to perform 3D positioning in a 3D area.

In their study, Hinzmann, et al. [2] obtained the odometry data of the robot in accordance with the depth information obtained by using image processing algorithms of the images obtained from the camera placed on flying robots for use in flying robots.

Teixeira, et al. [3] studied on determining the best route with predetermined start and end by using the SLAM algorithms known in the literature as well as telemetry data of Intel's AscTec Firefly drone.

In their study, Iacono and Sgorbissa [4] developed an obstacle avoidance algorithm from the data produced by an RGB-D camera (Microsoft Kinect) they installed on AscTec Firefly drone as an additional algorithm to the algorithms already used in the autonomous controlling of drones.

Kaufman, et al. [5], in their study, integrated variable motion planning algorithm to the Bayesian Probability Mapping algorithm in order to minimize the errors in Bayesian Probability Mapping algorithm that is used in the literature for mapping studies in 3D spaces of unknown shape. They also proposed that the 3D position can be used in the solution of the 3D problem by slicing it into 2D positioning method. Meanwhile, both mapping and reconnaissance algorithms were explained in this study through simulations and quadrotor flight trials.

In their study, Nguyen, et al. [6] developed and proposed a new method for creating a new curve path in case of an obstacle. They concluded that the method can create a safe path that takes into account obstacles detected in real time and prevents collisions.

Yu, et al. [7] performed 2D positioning and 3D mapping in a 3D environment with obstacles by using Rplidar and Kinect sensor with ROS operating system control.

In the study conducted by Yu, et al. [7], they created a composite coordinate positioning system by combining the indoor 2D map of the object created by the Gmapping algorithm and the 3D point cloud image information of the object. According to the result of the empirical studies conducted by Yu, et al. [7], they concluded that the positioning precision of the composite coordinate positioning system in this study is 6.7% higher than widespread ultrasonic and infrared positioning systems [8]; 20% higher than Bluetooth angle estimation positioning system [9]; and 72% higher than ultra-wideband positioning system [10].

Nellithimaru and Kantor [11] conducted a study concerning the classification of agricultural products using the SLAM algorithm they developed by adding a stereo camera. They stated that they achieved positive results thanks to this method in outdoor environments (without any pipeline used in such applications) without any constant lighting conditions or scene dynamics. They conducted this study in vineyards in order to count the crops of the farmers and estimate the yield of the crop to be obtained from the land.

When the studies in the literature are examined, it is seen that there are robots used for mapping from the ground and air. Ground and air robots that are used for mapping of indoor and outdoor spaces map their environments either by the control of an operator or by navigating autonomously [12] [13] [14] [15] [16] [17].

As for the terrestrial robots commonly used in the literature, it is seen that mapping reaches a high level of precision. However, when the physical conditions of the spaces such as mine pits, underground cities or caves where the study will be performed are examined, it is concluded that terrestrial robots are not suitable for these physical conditions due to fact that these robots must be used on flat or close-to-flat surfaces.

3 Description of System

In this study conducted, through the M100 Robot model created under Gazebo simulator, the control of the M100 Robot in real-world under the ROS environment is carried out by the developed nodes. The nodes developed, their relations with each other as well as the publisher and subscriber topics by these nodes are shown in Figure 1.

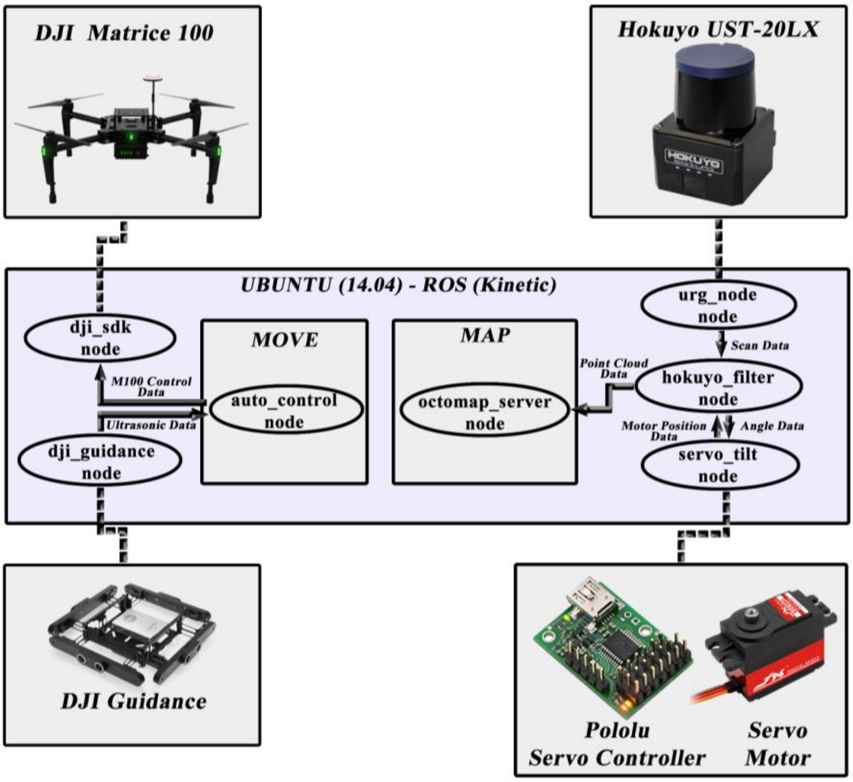


Figure 1
Block diagram of the study

3.1 Equipment

A DJI Guidance ultrasonic sensor kit, Hokuyo UST-20LX laser sensor, as well as, a Mini PC, running at 2.8 GHz processor (i5-6200U) with 4GB DDR3 Ram and 256 GB SSD hard disk, were placed on the M100 Robot (Figure 2).

Ubuntu 14.04 operating system was installed on the Mini PC due to the fact that the ROS Kinetic operating system used for the control of the M100 robot supports the Ubuntu 14.04 operating system. The control of M100 Robot and Hokuyo UST-20LX laser sensor is conducted by the ROS Kinetic operating system installed on the Mini PC.

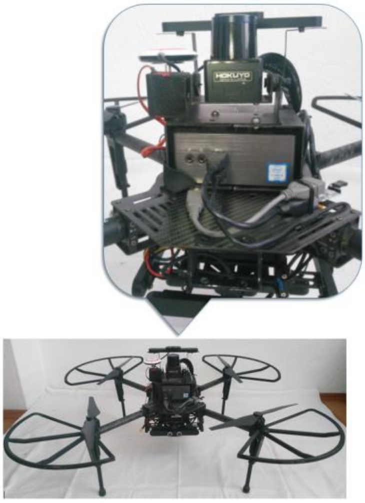


Figure 2
3D Simultaneous Positioning and Mapping Robot in Dark, Indoor Environments with Autonomous Flying Robot

The Hokuyo UST-20LX laser sensor is capable of area scanning on the y-axis at an angle of 270° within a period of 25 ms. The resolution of this sensor is 0.25° (Equation 1) and the measuring range is 20-30 meters depending on the amount of light in the environment. In this study, the data transfer between Hokuyo UST-20LX laser sensor and Mini PC is carried out through RJ45 connection [18].

$$270° / 1080 \text{ (step)} = 0.25° \qquad (1)$$

The servomotor driven electromechanical system in Figure 3 was designed so that the Hokuyo UST-20LX laser sensor can capture distance information by scanning its 3D environment in multi-axis. Thanks to this design, the Hokuyo UST-20LX sensor, which can scan in 2D, is equipped with 3D scanning capability.

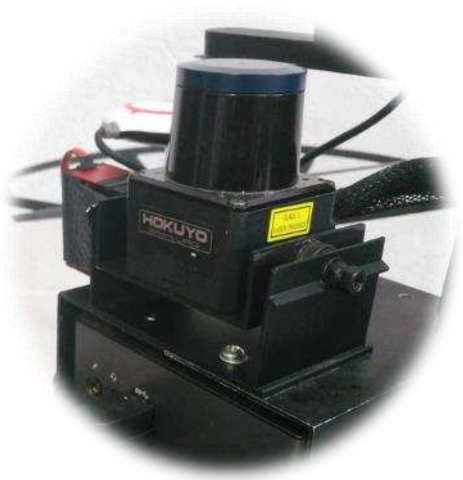


Figure 3
Electromechanical system developed for 3D scanning of the Hokuyo UST-20LX sensor

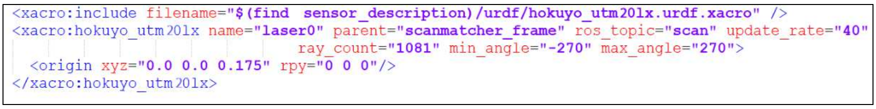
A servo motor driver board in Figure 4 was used in order to control the movement of the Hoyuko UST-20LX sensor on the electro-mechanics system in Figure 3 between the 70° and 80° on Z axis through the Mini PC [19].

Figure 4
PC controlled, Pololu Micro Maestro 6-Channel USB Servo Controller

Through this moving system, the Hokuyo UST-20LX sensor taking measurements from horizontal axis was provided with the ability to make measurements from vertical axis (z axis) in addition to the horizontal axis.

3.2 Installing Hokuyo UST-20LX on the DJI Matrice 100 Flying Robot Platform

The Hokuyo UST-20LX laser sensor model of ROS's ‘hukuyo_utm20lx’ package was added to the x = 0, y = 0, z = 0.175 (mt) position of ‘*scanmatcher_frame*’ frame where the M100 Robot that was modelled in Blender program is located through the XML codes in Figure 5.

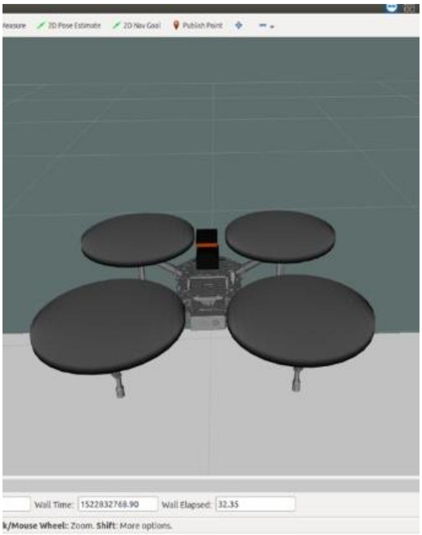


```
<xacro:include filename="$(find  sensor_description)/urdf/hokuyo_utm20lx.urdf.xacro" />
<xacro:hokuyo_utm20lx name="laser0" parent="scanmatcher_frame" ros_topic="scan" update_rate="40"
                             ray_count="1081" min_angle="-270" max_angle="270">
  <origin xyz="0.0 0.0 0.175" rpy="0 0 0"/>
</xacro:hokuyo_utm20lx>
```

Figure 5

Installing Hokuyo UST-20LX on M100 Robot model

When the M100 Robot model as well as the Hokuyo UST-20LX model installed on the M100 Robot model are opened in the RViz visualizer, they appear on the RViz visualizer screen as in Figure 6.



Wall Time: 1522832768.90 Wall Elapsed: 32.35

k/Mouse Wheel: Zoom. Shift: More options.

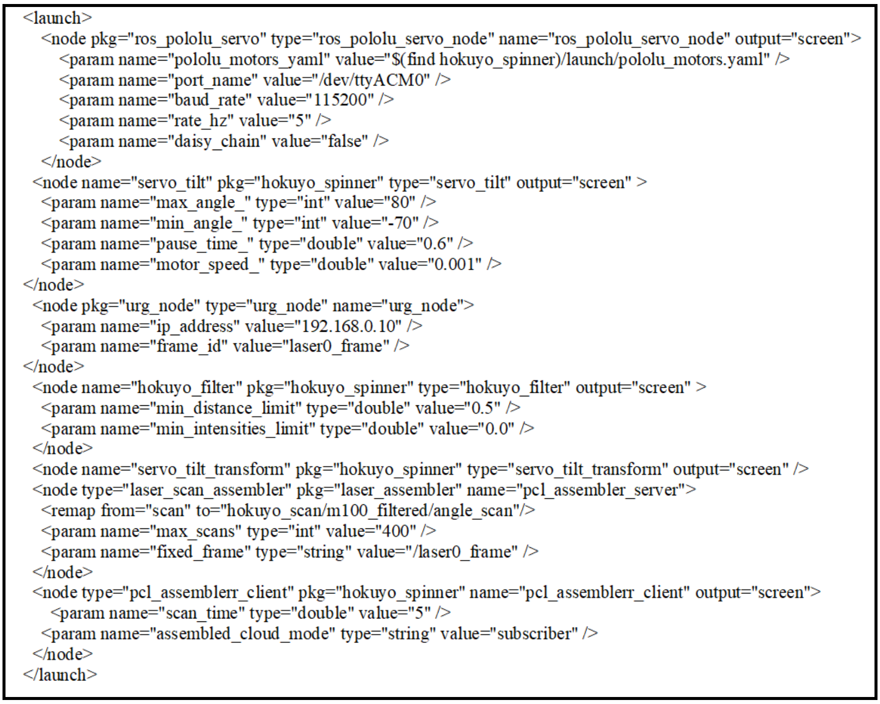
Figure 6

Image of M100 Robot and M100 Robot model in RViz visualizer

The communication between the Hokuyo UST-20LX laser sensor and the PC is carried out via the TCP/IP protocol using the static IP address of 192.168.0.10.

3.3. Obtaining 3D Data from 2D Data taken from Hokuyo UST-20LX

The "create_point_cloud.launch" in Figure 7 is an XML file that was created for use in 3D mapping and that combines the nodes which are necessary to obtain 3D data from 2D data.



```
<launch>
  <node pkg="ros_pololu_servo" type="ros_pololu_servo_node" name="ros_pololu_servo_node" output="screen">
    <param name="pololu_motors_yaml" value="$(find hokuyo_spinner)/launch/pololu_motors.yaml" />
    <param name="port_name" value="/dev/ttyACM0" />
    <param name="baud_rate" value="115200" />
    <param name="rate_hz" value="5" />
    <param name="daisy_chain" value="false" />
  </node>
 <node name="servo_tilt" pkg="hokuyo_spinner" type="servo_tilt" output="screen" >
  <param name="max_angle_" type="int" value="80" />
  <param name="min_angle_" type="int" value="-70" />
  <param name="pause_time_" type="double" value="0.6" />
  <param name="motor_speed_" type="double" value="0.001" />
</node>
 <node pkg="urg_node" type="urg_node" name="urg_node">
  <param name="ip_address" value="192.168.0.10" />
  <param name="frame_id" value="laser0_frame" />
</node>
 <node name="hokuyo_filter" pkg="hokuyo_spinner" type="hokuyo_filter" output="screen" >
  <param name="min_distance_limit" type="double" value="0.5" />
  <param name="min_intensities_limit" type="double" value="0.0" />
 </node>
 <node name="servo_tilt_transform" pkg="hokuyo_spinner" type="servo_tilt_transform" output="screen" />
 <node type="laser_scan_assembler" pkg="laser_assembler" name="pcl_assembler_server">
  <remap from="scan" to="hokuyo_scan/m100_filtered/angle_scan"/>
  <param name="max_scans" type="int" value="400" />
  <param name="fixed_frame" type="string" value="/laser0_frame" />
 </node>
 <node type="pcl_assemblerr_client" pkg="hokuyo_spinner" name="pcl_assemblerr_client" output="screen">
   <param name="scan_time" type="double" value="5" />
  <param name="assembled_cloud_mode" type="string" value="subscriber" />
 </node>
</launch>
```

Figure 7

The "create_point_cloud.launch" XML file written for transition from 2D to 3D

While the '*ros_pololu_servo_node*' node in the "create_point_cloud.launch" file make the connection with the servo-motors of the drive board in ROS, the '*servo_tilt*' node ensures movement in Hokuyo UST-20LX on the vertical axis (The servo-motor is continuously moved between -70° and 80°). After the obstacle information taken from Hokuyo UST-20LX laser sensor by the '*urg_node*' node is filtered by the '*hokuyo_filter*' node, this information is shared in ROS environment in the topic form of '*/hokuyo_scan/m100_filtered/angle_scan*'. The 3D point cloud is created by combining the angle created by both '*pcl_assembler_server*' node and '*servo_tilt_transform'* (the tf angle of layers created in Figure 8) node as well as the topic of *'/hokuyo_scan/m100_filtered/angle_scan'*.

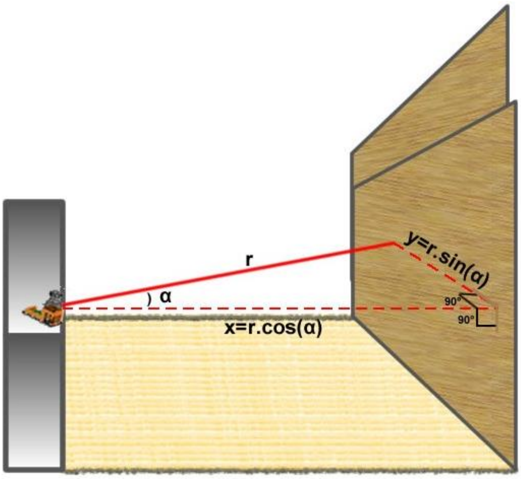


Figure 8
The 3D tf conversion of Hokuyo UST-20LX laser sensor

Equation 2 was used for 3D conversion of information from the Hokuyo UST-20LX laser sensor along the Y axis according to the view angle of M100 Robot.

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} r\ cos\alpha \\ r\sin\alpha \\ Z_{imu} \end{pmatrix} \tag{2}$$

Then, the 3D point cloud generated by the '*pcl_assembler_server*' node was shared in the topic form of '*/hokuyo_scan/m100_filtered/assembled_cloud*' in '*pcl_assemblerr_client*' node to be mapped in ROS environment.

Filtering data from Hokuyo UST-20LX: The Hokuyo UST-20LX laser sensor is capable of measuring within 270° along the Y-axis [18]. The data within the limits of 46°-225° were evaluated by filtering the measurement gaps of 0°-45° and 225°-270° due to the fact that the hardware of M100 robot runs into the view angle of Hokuyo UST-20LX laser sensor between 0°-270° measurement gap. In addition, during the movement of the electromechanical system in Figure 3 created to obtain 3D data from 2D data of the Hokuyo UST-20LX laser sensor, since some angle values on the z-axis correspond to the M100 Robot's equipment like wings, the measuring range was determined to be taken as 0.5 m and above. Thus, topic of '*/scan*' shared by '*urg_node*' in ROS is published in ROS environment being filtered by the '*hokuyo_filter*' node in order to be used with tf angular conversion as topic form of '*/hokuyo_scan/m100_filtered/angle_scan*'.

Angular tf conversion for transition from 2D to 3D: In order to perform the mapping in the RViz visualizer, the real-world M100 Robot and the M100 Robot model visualized in the RViz visualizer are programmed to work simultaneously with ROS. M100 Robot model and peripheral equipment models (links: laser sensor, servo-motor mechanism, etc., which moves the laser sensor in the z axis) were placed on different layers in Rviz visualizer to be controlled by ROS in the visualizer. While the M100 Robot model is on the '*scanmatcher_frame*' layer, the Hokuyo UST-20LX laser sensor model is placed on the '*laser0_frame*' layer. Thus, the Hokuyo UST-20LX laser sensor information on the y-axis, which is received during the movement of the real-world Hokuyo UST-20LX laser sensor in the range of -70° to 80° on the z-axis, is converted into a 3D point cloud in ROS environment (Figure 9) by distributing the information to the '*laser0_frame*' layer that moves between -70° and 80° along the z-axis of the location, where the M100 Robot is positioned.

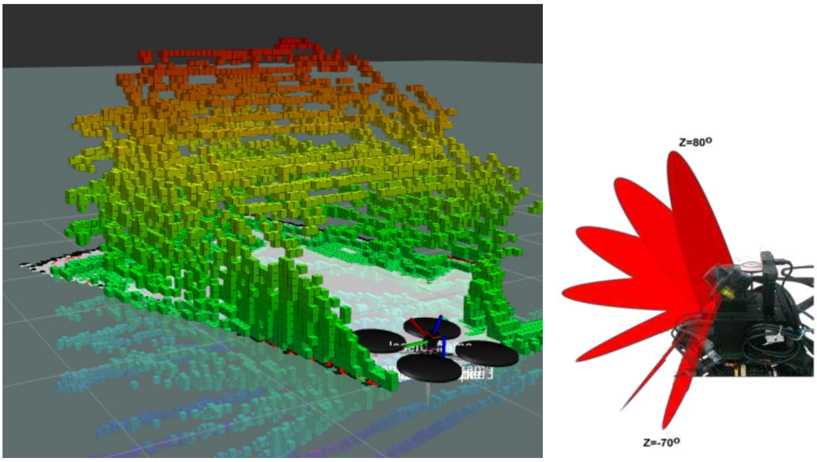


Figure 9

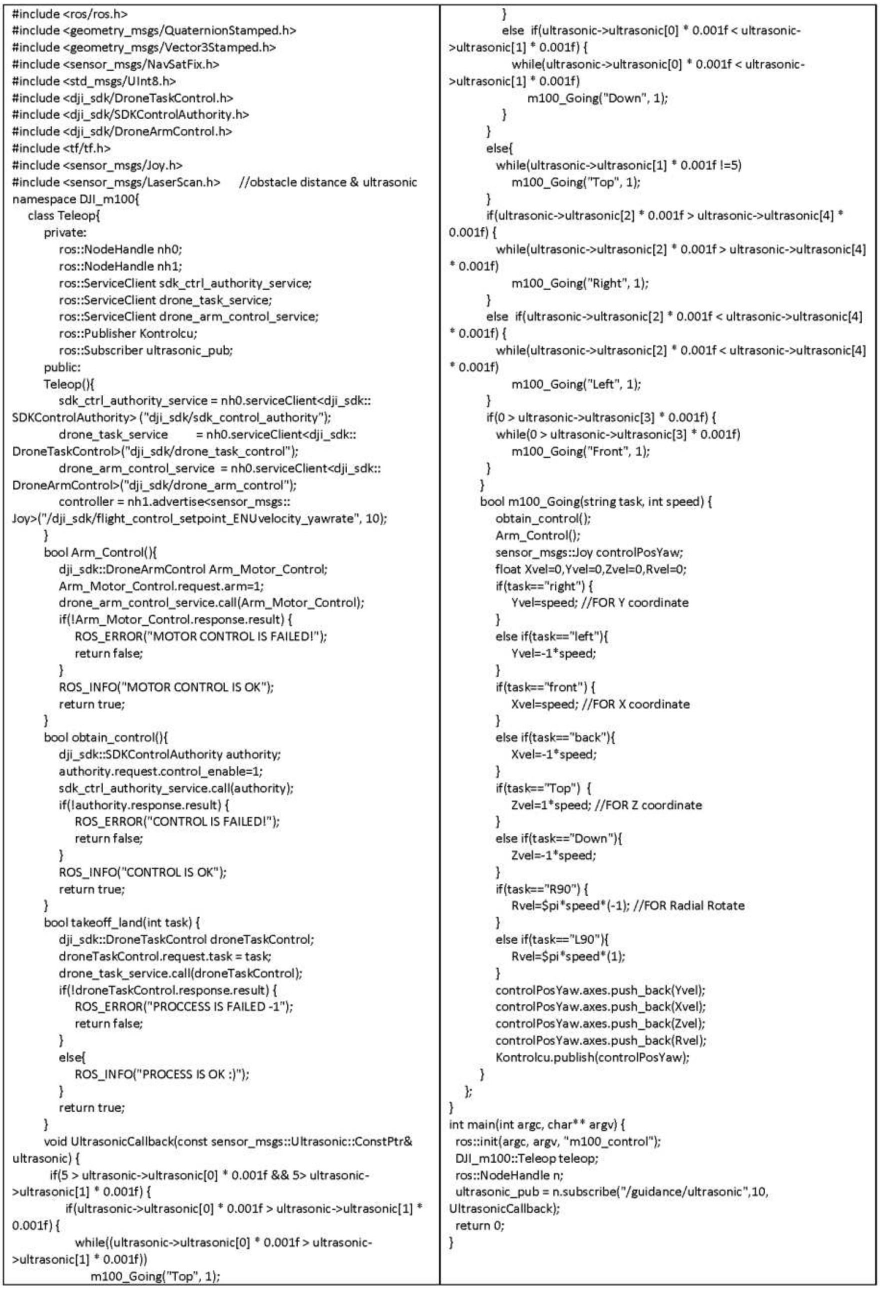
Conversion from 2D to 3D with Hokuyo UST-20LX laser sensor

3.4 The Control of M100 Robot

The autonomous control software for the autonomous movement of the M100 Robot is as in Figure 10. Autonomous movement of the M100 Robot was conducted by using obstacle information obtained from ultrasonic sensors.

The "obtain_control()" function in Figure 10 is a C ++ code block created to make the M100 Robot controllable through shared messages on ROS. Communication of the M100 Robot with the PC is carried out via the UART port of the M100 Robot. The M100 Robot's motors are activated by the "Arm_Control()" function.

The "m100_Going(string task, int speed)" function created for the autonomous control of the M100 Robot enables the movement of the M100 Robot in the direction of the '*task*' parameter that corresponds to the function as a parameter and at the speed of '*speed*' parameter that also corresponds to the function as a parameter.



```
#include <ros/ros.h>
#include <geometry_msgs/QuaternionStamped.h>
#include <geometry_msgs/Vector3Stamped.h>
#include <sensor_msgs/NavSatFix.h>
#include <std_msgs/UInt8.h>
#include <dji_sdk/DroneTaskControl.h>
#include <dji_sdk/SDKControlAuthority.h>
#include <dji_sdk/DroneArmControl.h>
#include <tf/tf.h>
#include <sensor_msgs/Joy.h>
#include <sensor_msgs/LaserScan.h>     //obstacle distance & ultrasonic
namespace DJI_m100{
   class Teleop{
      private:
         ros::NodeHandle nh0;
         ros::NodeHandle nh1;
         ros::ServiceClient sdk_ctrl_authority_service;
         ros::ServiceClient drone_task_service;
         ros::ServiceClient drone_arm_control_service;
         ros::Publisher Kontrolcu;
         ros::Subscriber ultrasonic_pub;
      public:
      Teleop(){
         sdk_ctrl_authority_service = nh0.serviceClient<dji_sdk::
SDKControlAuthority> ("dji_sdk/sdk_control_authority");
         drone_task_service       = nh0.serviceClient<dji_sdk::
DroneTaskControl>("dji_sdk/drone_task_control");
         drone_arm_control_service  = nh0.serviceClient<dji_sdk::
DroneArmControl>("dji_sdk/drone_arm_control");
         controller = nh1.advertise<sensor_msgs::
Joy>("/dji_sdk/flight_control_setpoint_ENUvelocity_yawrate", 10);
      }
      bool Arm_Control(){
         dji_sdk::DroneArmControl Arm_Motor_Control;
         Arm_Motor_Control.request.arm=1;
         drone_arm_control_service.call(Arm_Motor_Control);
         if(!Arm_Motor_Control.response.result) {
            ROS_ERROR("MOTOR CONTROL IS FAILED!");
            return false;
         }
         ROS_INFO("MOTOR CONTROL IS OK");
         return true;
      }
      bool obtain_control(){
         dji_sdk::SDKControlAuthority authority;
         authority.request.control_enable=1;
         sdk_ctrl_authority_service.call(authority);
         if(!authority.response.result) {
            ROS_ERROR("CONTROL IS FAILED!");
            return false;
         }
         ROS_INFO("CONTROL IS OK");
         return true;
      }
      bool takeoff_land(int task) {
         dji_sdk::DroneTaskControl droneTaskControl;
         droneTaskControl.request.task = task;
         drone_task_service.call(droneTaskControl);
         if(!droneTaskControl.response.result) {
            ROS_ERROR("PROCCESS IS FAILED -1");
            return false;
         }
         else{
            ROS_INFO("PROCESS IS OK :)");
         }
         return true;
      }
      void UltrasonicCallback(const sensor_msgs::Ultrasonic::ConstPtr&
ultrasonic) {
        if(5 > ultrasonic->ultrasonic[0] * 0.001f && 5> ultrasonic-
>ultrasonic[1] * 0.001f) {
           if(ultrasonic->ultrasonic[0] * 0.001f > ultrasonic->ultrasonic[1] *
0.001f) {
            while((ultrasonic->ultrasonic[0] * 0.001f > ultrasonic-
>ultrasonic[1] * 0.001f))
               m100_Going("Top", 1);
           }
           else  if(ultrasonic->ultrasonic[0] * 0.001f < ultrasonic-
>ultrasonic[1] * 0.001f) {
            while(ultrasonic->ultrasonic[0] * 0.001f < ultrasonic-
>ultrasonic[1] * 0.001f)
               m100_Going("Down", 1);
           }
        }
        else{
          while(ultrasonic->ultrasonic[1] * 0.001f !=5)
            m100_Going("Top", 1);
        }
        if(ultrasonic->ultrasonic[2] * 0.001f > ultrasonic->ultrasonic[4] *
0.001f) {
          while(ultrasonic->ultrasonic[2] * 0.001f > ultrasonic->ultrasonic[4]
* 0.001f)
            m100_Going("Right", 1);
        }
        else  if(ultrasonic->ultrasonic[2] * 0.001f < ultrasonic->ultrasonic[4]
* 0.001f) {
          while(ultrasonic->ultrasonic[2] * 0.001f < ultrasonic->ultrasonic[4]
* 0.001f)
            m100_Going("Left", 1);
        }
        if(0 > ultrasonic->ultrasonic[3] * 0.001f) {
          while(0 > ultrasonic->ultrasonic[3] * 0.001f)
            m100_Going("Front", 1);
        }
      }
      bool m100_Going(string task, int speed) {
         obtain_control();
         Arm_Control();
         sensor_msgs::Joy controlPosYaw;
         float Xvel=0,Yvel=0,Zvel=0,Rvel=0;
         if(task=="right") {
            Yvel=speed; //FOR Y coordinate
         }
         else if(task=="left"){
            Yvel=-1*speed;
         }
         if(task=="front") {
            Xvel=speed; //FOR X coordinate
         }
         else if(task=="back"){
            Xvel=-1*speed;
         }
         if(task=="Top")  {
            Zvel=1*speed; //FOR Z coordinate
         }
         else if(task=="Down"){
            Zvel=-1*speed;
         }
         if(task=="R90") {
            Rvel=$pi*speed*(-1); //FOR Radial Rotate
         }
         else if(task=="L90"){
            Rvel=$pi*speed*(1);
         }
         controlPosYaw.axes.push_back(Yvel);
         controlPosYaw.axes.push_back(Xvel);
         controlPosYaw.axes.push_back(Zvel);
         controlPosYaw.axes.push_back(Rvel);
         Kontrolcu.publish(controlPosYaw);
      }
   };
}
int main(int argc, char** argv) {
  ros::init(argc, argv, "m100_control");
  DJI_m100::Teleop teleop;
  ros::NodeHandle n;
  ultrasonic_pub = n.subscribe("/guidance/ultrasonic",10,
UltrasonicCallback);
  return 0;
}
```

Figure 10

The autonomous control software of M100 Robot

A repeating "UltrasonicCallback()" function at 10 Hz was created (The reading frequency of the DJI Guidance sensor from the ultrasonic sensors is 10 Hz [20]) for the control of the ultrasonic sensor located on the right, left, front, under and top of the M100 Robot and for calling the "M100_going()" function with direction and speed parameters in accordance with the distance information coming from the sensors.

The autonomous control software for M100 Robot in Figure 10 was created with possible scenarios in which the M100 Robot may encounter in closed spaces. Among these scenarios are;

- Raising and lowering of M100 Robot in order to balance the top-bottom distance if the distance information from the ultrasonic sensors on the top and bottom of the M100 Robot is less than 5 m,
- Keeping the altitude of the M100 Robot at 5 m and keeping this altitude if the distance information from the ultrasonic sensors on the top and bottom of the M100 Robot is not more than 5 m,
- Ensuring the balance of M100 Robot by moving to the right if the information from the ultrasonic sensors from the right side are less from the left side of the M100 Robot or to the left direction for a vice versa situation is valid,
- The advancement of the M100 Robot if there is no obstacle ahead, according to the distance information coming from the ultrasonic sensors.

The coordinate axes in Figure 11 are used for X, Y, Z direction movement of M100 Robot.

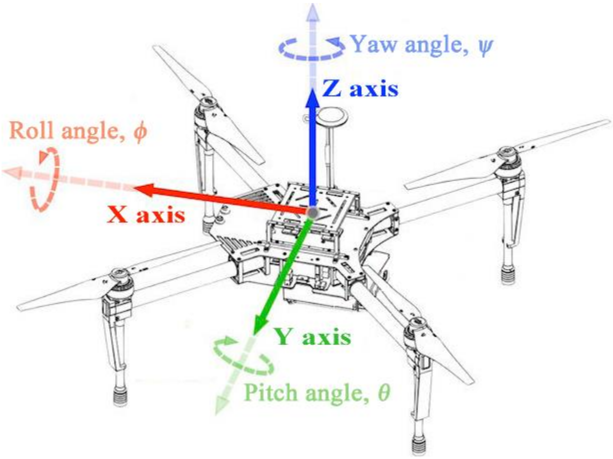


Figure 11
Coordinate axes of the M100 Robot

4 Practical Applications

The mapping studies were carried out in caves in Turkey, Isparta, Atabey district at coordinates of 37°53'41.8"N 30°32'58.5"E and 37°53'39.0"N 30°32'42.3"E.

1. Mapping Study: The entrance of the cave at the coordinates of 37°53'41.8"N 30°32'58.5"E is as in Figure 12.

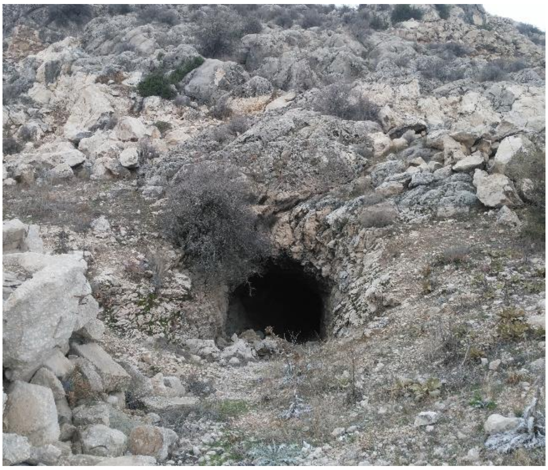


Figure 12
The entrance of the cave at the coordinates of 37°53'41.8"N 30°32'58.5"E

This cave in Figure 12 opens to 400-500 m inwards from the foot of the mountain. The height of the cave provides a space where a person of 175 cm height can walk upright (Figure 13).

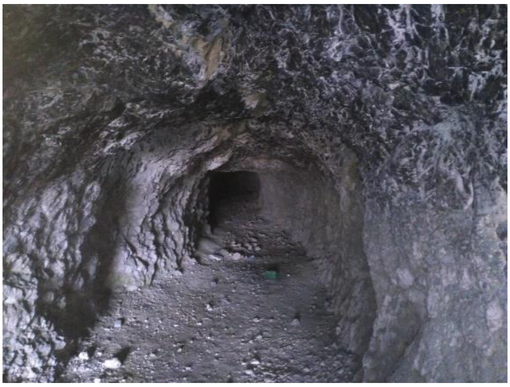


Figure 13
Inside of the cave at the coordinates of 37°53'41.8"N 30°32'58.5"E

The interior of the cave in Figure 13 is very dark and dangerous. In this study, carried out to map such dark and closed spaces, the map of this cave, at the coordinates of 37°53'41.8"N 30°32'58.5"E, was created, as shown in Figure 14.

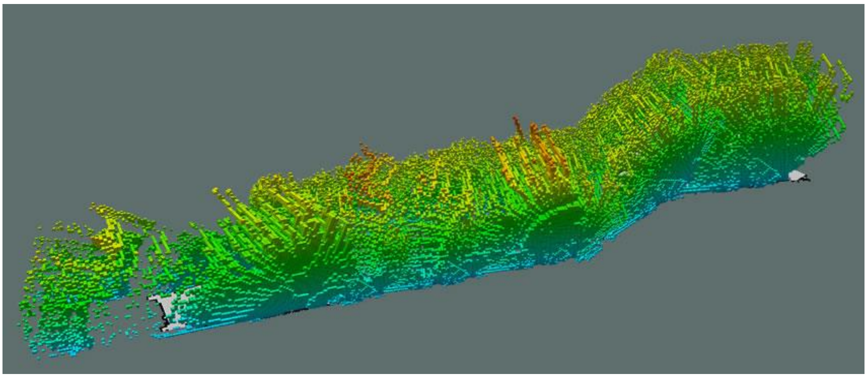


Figure 14
The map of the cave at the coordinates of 37°53'41.8"N 30°32'58.5"E

The M100 Robot moved autonomously starting from the entrance of the cave. Since the M100 Robot moves slowly during the rotation around its own axis and due to its slow movement during direction changes on z-axis (movement in the -z and +z direction), as determined in the control software of the robot, there are some accumulations on some point clouds as seen in Figure 14. At the last stage of the mapping process, due to the fact that the right and left side of the M100 Robot was closed, and thus, had to make a 180° rotation on its axis, there are some missing point clouds, at the last section of the map in Figure 14. Yet, it can be seen that a map identical to the cave could be obtained when the map was studied in general terms.

2. Mapping Study: The entrance of the cave at the coordinates of 37°53'39.0"N 30°32'42.3"E is as in Figure 15.

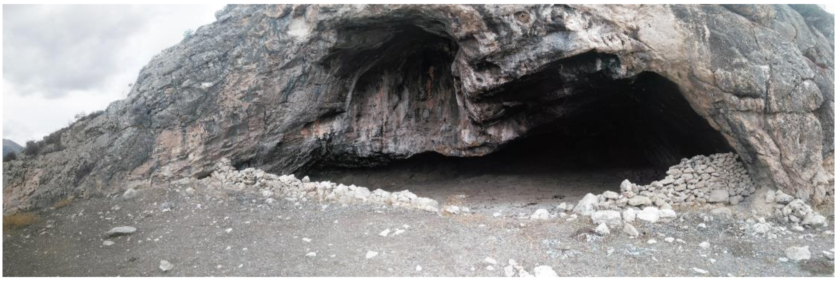


Figure 15
The entrance of the cave at the coordinates of 37°53'39.0"N 30°32'42.3"E

The location of this cave in Figure 15, is approximately 100 m above the foot of the mountain. It has an area extending to the right, after the entrance gate of the cave. The interior of the cave has a rectangular prism structure as in Figure 16.

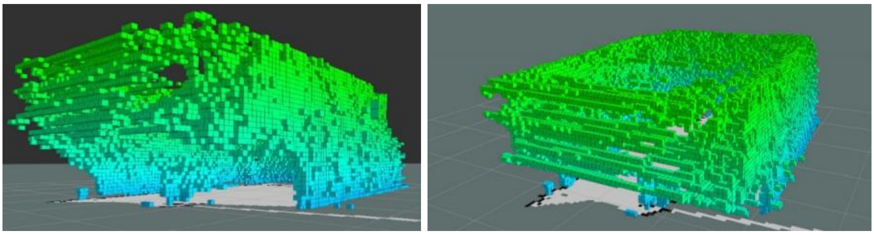


Figure 16
The map of the cave at the coordinates of 37°53'39.0"N 30°32'42.3"E

When the map in Figure 16 was studied, it was seen that a map identical to the cave at the coordinates of 37°53'39.0"N 30°32'42.3"E could be obtained.

Conclusions

In this study, an autonomous, high-precision, fast and low-cost system that has the ability to map in 3D was created. Dark, rugged and dangerous closed spaces, such as caves, underground cities and mines were mapped in 3D, with this system. The system has no limit on movement (bumpy, hollowed or sloping spaces).

In order to obtain the obstacle information, during autonomous movement, the developed system used within the scope of this study, 5 ultrasonic sensors – right, left, front, top, bottom – were used. For the purpose of data collection from Hokuyo UST-20LX laser sensor that can collect 2D data (X and Y axis) along the Z axis, a servomotor driven electromechanical equipment was developed so that the developed system can map its environment in 3D during its autonomous movement. The control of the hardware developed and used in this study was carried out by ROS nodes written in C++ programming language.

The developed system can not only be actively used in mining operations, where changes must be continuously monitored, especially these days when occupational health and safety are at the forefront, but also in the discovery of underground cities and caves, which are important sources of economy and tourism in countries, like Turkey. The system can also provide literature contributions to National and International Academic Studies.

Acknowledgments

This research was brought to life through the Project 116E013, which is supported by the Scientific and Technical Research Council of Turkey (TÜBİTAK). We would like to thank them for their contributions.

References

[1] Soma_Mine_Disaster (2014, 30.12.2015) Soma Mine Disaster [Online]. Available: https://en.wikipedia.org/wiki/Soma_mine_disaster

[2] T. Hinzmann, J. L. Schönberger, M. Pollefeys, and R. Siegwart, "Mapping on the fly: Real-time 3D dense reconstruction, digital surface map and incremental orthomosaic generation for unmanned aerial vehicles," in Field and Service Robotics, 2018: Springer, Cham, pp. 383-396

[3] L. Teixeira, I. Alzugaray, and M. Chli, "Autonomous aerial inspection using visual-inertial robust localization and mapping," in Field and Service Robotics, 2018: Springer, Cham, pp. 191-204

[4] M. Iacono and A. Sgorbissa, "Path following and obstacle avoidance for an autonomous UAV using a depth camera," Robotics and Autonomous Systems, Vol. 106, pp. 38-46, 2018

[5] E. Kaufman, K. Takami, Z. Ai, and T. Lee, "Autonomous Quadrotor 3D Mapping and Exploration Using Exact Occupancy Probabilities," in 2018 Second IEEE International Conference on Robotic Computing (IRC), 2018: IEEE, pp. 49-55

[6] P. D. Nguyen, C. T. Recchiuto, and A. Sgorbissa, "Real-time path generation and obstacle avoidance for multirotors: a novel approach," Journal of Intelligent & Robotic Systems, Vol. 89, No. 1-2, pp. 27-49, 2018

[7] Y. Yu, Y. Piao, Y. Ni, and T. Si, "Research on Accurate Positioning of Indoor Objects Based on ROS and 3D Point Cloud," in Journal of Physics: Conference Series, 2019, Vol. 1229, No. 1: IOP Publishing, p. 012005

[8] F. Yao, C. Tao, and S. Li, "Indoor positioning system research and implementation based on phase of arrival," Electronic Measurement Technology, Vol. 39, No. 11, pp. 30-35, 2016

[9] M. W. Liu, T. J. Liu, Y. Ye, and L. Wu, "The Application Research of Indoor Positioning Based on Low-power Bluetooth Technology," Wireless Communication Technology, Vol. 24, No. 3, pp. 19-23, 2015

[10] J. Zheng, H. Qian, and L. Wang, "An improved DV-Hop positioning algorithm for wireless sensor network," in 2015 IEEE International Conference on Progress in Informatics and Computing (PIC), Nanjing, China, 2015, Nanjing, China: IEEE

[11] A. K. Nellithimaru and G. A. Kantor, "ROLS: Robust Object-Level SLAM for Grape Counting," in The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, California, 2019: IEEE Xplore, p. 9

[12] I. Peitzsch, B. Liboy, S. Biaz, and R. Chapman, "3D Mapping Using a UAV, an IMU, and a 2D LiDAR," 2019

[13] A. Molnár, "Surveying Archaeological Sites and Architectural Monuments with Aerial Drone Photos," Acta Polytechnica Hungarica, Vol. 16, No. 7, 2019

[14] B. Guo, H. Dai, Z. Li, and W. Huang, "Efficient Planar Surface-Based 3D Mapping Method for Mobile Robots Using Stereo Vision," IEEE Access, Vol. 7, pp. 73593-73601, 2019

[15] T. Pozderac, J. Velagić, and D. Osmanković, "3D Mapping Based on Fusion of 2D Laser and IMU Data Acquired by Unmanned Aerial Vehicle," in 2019 6th International Conference on Control, Decision and Information Technologies (CoDIT), 2019: IEEE, pp. 1533-1538

[16] D. Backes, G. Schumann, F. Teferele, and J. Boehm, "Towards a high-resolution drone-based 3D mapping dataset to optimise flood hazard modelling," The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. 42, No. W13, pp. 181-187, 2019

[17] Z. Jiang, J. Zhu, Z. Lin, Z. Li, and R. Guo, "3D mapping of outdoor environments by scan matching and motion averaging," Neurocomputing, 2019

[18] UST-20LX (2014, 20.11.2016) Distance Data Output/UST-10/20LX [Online] Available: https://www.hokuyo-aut.jp/search/single.php?%20serial=167

[19] M. Maestro. Pololu Maestro Servo Controller User's Guide [Online] Available: https://www.pololu.com/docs/pdf/0J40/maestro.pdf

[20] Guidance_User_Manual_EN. DJI Guidance User Manual [Online] Available: http://download.dji-innovations.com/downloads/dev/Guidance/en/Guidance_User_Manual_en_V1.6.pdf

FPGA HW Accelerator of the First Step of Systematic Two-Level Minimization of Single-Output Boolean Function

Branislav Madoš, Norbert Ádám, Zuzana Bilanová, Martin Chovanec

Department of Computers and Informatics, Faculty of Electrical Engineering and Informatics, Technical University of Košice
Letná 9, 042 00 Košice, Slovak Republic
e-mail: {branislav.mados, norbert.adam, zuzana.bilanova,
martin.chovanec}@tuke.sk

Abstract: Boolean function minimization is an area important not only in the development and optimization of digital logic, but also in other research and development areas, such as, the optimization of control systems, simplifying program logic, artificial intelligence, etc. The aim of this paper is to present a hardware accelerated first step of the systematic minimization of single-output Boolean functions – the generation of a set of prime implicants for both the disjunctive normal form (DNF) and the conjunctive normal form (CNF), having defined the OFF and ON sets and – alternatively – also the DC ("don't care") set. The proposed hardware accelerator is designed as combinational logic, described in VHDL. Its advantages include an extremely short prime-implicant-generation time in the order of ns and/or tens of ns – in case of Boolean functions with small amount of input variables – and the possibility to generate the valid-prime-implicant set of Boolean functions having a defined number of input variables at a constant time, regardless of the cardinality of the ON or, eventually, the DC sets. However, these advantages come with a large spatial complexity – the number of utilized implementation elements – of the respective combinational module, generating the prime-implicant set. The authors verified the proposed design using Field Programmable Gate Array (FPGA) technology, implementing the hardware using a Xilinx Kintex-7 KC-705 Evaluation Kit development board.

Keywords: Boolean function minimization; prime implicant generation; combinational logic; FPGA; disjunctive normal form; DNF; conjunctive normal form; CNF; hardware accelerator; systematic minimization; heuristic minimization

1 Introduction

Boolean function minimization is a significant problem not only in academia and scientific research, but also in many research and development areas. This includes, for example, the development of logic designs, such as Programmable Logic Array (PLA) technology, Field Programmable Gate Array (FPGA) technology, Application Specific Integrated Circuit (ASIC) technology, as well as the design and development of control systems including the vast and important domain of controlling intelligent buildings and houses [1], software engineering (to optimize logic used in software), artificial intelligence, security of computer systems [2] and many others. Boolean function minimization is a significant challenge mainly if the input variables are numerous (counting hundreds or thousands), rendering many minimization approaches impractical, since these cannot provide minimization using available hardware in considerable, practical time.

Today, logic design optimization may be classified by various criteria, such as design characteristics (i.e. combinational logic or sequential logic), the amount of levels (two-level or multi-level minimization), or the implementation method (algebraic, table-based or graphic minimization). Some algorithms are based on using human expertise in finding patterns, thus these are implemented “manually”. Further approaches are algorithmic – these implement the respective algorithms in software running on the CPUs and GPGPUs of traditional computers [3] [4] [5].

Another way of classifying optimization is to take the minimality of the solution into account – in this case, the categories are systematic and heuristic minimization. Systematic minimization will always find the minimum solution for the specified minimization criteria. The most famed approaches of systematic minimization include graphical minimization using Karnaugh maps (KM) and a tabular method using the Quine-McCluskey (Q-M) algorithm. On the contrary, heuristic minimization often yields a near-minimal solution, with an advantage: a cut (often radical) in processing time and resources. Thus, the aim of using heuristic minimization is to utilize it even in case of Boolean functions having high amounts of input variables, in case of which systematic minimization would not be of any practical use. The most famed solutions of heuristic Boolean function minimization include Espresso – the de-facto industry standard in Boolean function minimization – and its derivatives, as well as the BOOM and BOOM-II algorithms, respectively. For a discussion of both systematic and heuristic minimizations see section 2 herein.

In this paper, the authors focused on the field of systematic, two-level minimization of single-output Boolean functions – when implementing the algorithm, instead of using the CPU and/or the GPGPU to write software, they chose to implement the algorithm in a hardware-accelerated form, using Field Programmable Gate Array (FPGA) technology. The proposed solution is based on previous development – in [6], the authors presented a hardware-accelerated generator of prime implicants for single-output Boolean functions, based on combinational logic. In this paper, the authors describe an enhancement of the aforementioned solution, allowing the generation of prime implicants both the disjunctive (DNF) and conjunctive (CNF) normal form; compared to the previous version, the solution proposed herein allows the definition of not only of the OFF and ON sets, but also of the DC (don't care) set. The aim was to create a circuit that would significantly minimize the prime-implicant-generation time to the order of ns and/or tens of ns in case of Boolean functions with small amount of input variables.

The contribution hereof lies in the following:

- Design of a hardware-accelerated implementation of the first step of the systematic two-level minimization of single-output Boolean functions, based on a combinational logic module, to generate prime implicants; the proposed hardware accelerator allows processing of Boolean functions with output values defined not only by means of OFF and ON sets, but also the DC set.

The structure of the paper is as follows:

Section 2 deals with the related work in the field of systematic and heuristic Boolean function minimization. Due to the abundance of papers published in this field, the authors resorted to a selection of the fundamental works.

Section 3 contains a detailed description of the proposed hardware accelerated Boolean function minimizer. In the introductory part of this section, the authors describe the encoding of the hardware accelerator's input and output vectors. In the last part of the section, the authors describe the structure of the hardware accelerator itself, split into three submodules: the prime-implicant-generation mode selection module; the prime-implicant-generation module (implemented as a combinational logic circuit); and the invalid-prime-implicant-exclusion module.

Section 4 summarizes the testing results of the hardware-accelerator (implemented using a Xilinx Kintex-7 KC-705 Evaluation Kit evaluation board) for various numbers (2 to 8) of input variables of single-output Boolean functions.

Section 5 contains the conclusions, distilled from the results of the implemented tests, described in the previous section.

2 Related Work

Due to the large amount of work published in the field of systematic and heuristic Boolean functions minimization, this section of the paper contains only the selection of papers that are representing fundamental works related to the solution designed as the part of this work and presented in this paper.

Systematic minimization. In 1881, Allan Marquand presented his diagrams, which allowed the simplification of the graphical presentation of Venn diagrams for a larger number of variables [7]. In 1951, the Harvard minimizing charts were presented by Howard H. Aitken, described in detail in [8]. In 1952, Edward Westbrook Veitch developed a Boolean function minimization method [9], along with the corresponding diagrams, often called Marquand-Veitch diagrams. This method was later perfected by Maurice Karnaugh in 1953 [10] – today, it is known as Karnaugh maps (KM or K-maps). In 1956, Svoboda created graphical aids for systematic Boolean function minimization [11].

Karnaugh maps, sometimes referred to as Karnaugh-Veitch (KV) maps. These are not only a graphical notation for Boolean functions, but mainly serve for minimization purposes. These utilize human expertise in finding patterns within the graphical representation of the Boolean function depicted as a diagram, instead of minimizing the particular Boolean function using a computer program. to represent Boolean functions, Karnaugh-maps use a two-dimensional grid containing 2^n fields, n being the number of input variables. The fields are organised as a $2^k \times 2^l$ grid, where $k + l = n$ and k differs from l by at most 1. Each field of the Karnaugh-map contains information about the particular Boolean function's output value. A limitation of this method is that visual pattern matching and the subsequent simplification in K-maps is practical only for a very small number of input variables, while the limit amount is stated to be 5–6 input variables. A further drawback is the human factor, which may introduce errors into the process.

The Quine-McCluskey method, also referred to as the Q–M method, is a tabular method of systematic Boolean function minimization, which is, in terms of the achieved results, analogous to the K-map method. It was developed in 1952 by Willard Quine and Edward McCluskey [12] [13] as a two-step method. In the first step, the algorithm generates the prime implicants of the Boolean function, while in the second step, it solves the issue of covering the Boolean function by the prime implicants. Compared to the K-map method, the advantage of this method is that it does not rely on the capacity of a human to find patterns, but rather it introduces an algorithm ready to be implemented in a computer, thus it may be used to process Boolean functions with significantly more variables. The systematic approach of this method prevents its practical use in case of high amounts (i.e. hundreds or thousands) of input variables – this method is time and resource hungry.

Heuristic minimization. The MINI heuristic minimizer was presented by Hong et al. in 1974 [14]. It generates a solution without the necessity to generate all prime implicants of the Boolean function to be minimized. The Espresso logic minimizer was presented by R. K. Brayton et al. with the goal to minimize logic circuits using heuristic methods [15]. The Espresso-MV (Multi-valued) method is a derivative of the Espresso method; it was developed in 1986 by Richard L. Rudell. Both the heuristic and systematic minimization approaches were described in [16].

The C language source code of the Espresso algorithm is available in [17]. Further improvements to the Espresso method include the Espresso-Exact and Espresso Signature methods [18].

The two-level Boolean minimization tool called BOOlean Minimizer (BOOM), developed by Hlavička and Fišer, is based on the new paradigm of implicant generation: unlike other minimization methods, generating implicants using the bottom-up approach, the BOOM method uses a top-down approach. A further advantage is also in the reduction of the amount of prime implicants. The proposed algorithm is well suited for Boolean functions with the large number of variables (up to thousands), when other algorithms are not able to yield results in reasonable time [19] [20] [21] [22] [23]. The FC-Min Boolean minimizer was introduced by Fišer and Kubátová in [24]; later, it was combined with the BOOM algorithm as the BOOM-II Boolean minimizer [25] [26].

3 Proposed HW Accelerator

The hardware accelerator proposed herein uses a combinational logic circuit as its most important part, aimed at the generation of prime implicants of the Boolean function. The circuit design is described using the VHDL language. In the phase of testing the design, its practical implementation was performed using Field Programmable Gate Array (FPGA) technology.

The aim of this section is to describe the encoding of two binary input vectors – containing the OFF set and the ON set – and/or the DC set of the single-output Boolean function. Then, the description of the encoding of the output vector – representing the prime implicant set of the particular Boolean function – follows. The last part of this section contains the description of the three modules of the hardware accelerator itself: the prime-implicant-generation mode selection module; the prime-implicant-generation module (implemented as a combinational logic circuit); and the invalid-prime-implicant-exclusion module.

3.1 Boolean Function Truth Table Encoding

The input of the hardware accelerator is the representation of the truth table of the single-output Boolean function of n input variables, as 2^n-sized binary vectors.

The size of vectors results from the line-count of the Boolean function truth table. Each such line of the truth table has a binary code assigned pursuant to the input variable configuration. If the input variable is in complementary form, 0 is used, while for variables in true form, 1 is used. This binary code may be transformed to a decadic equivalent (DE), as shown in Table 1. On its input, the hardware accelerator accepts a Boolean function output value from the set $\{0, 1, \times\}$, where $\times$ is the „don't care” value, i.e. the output value has no importance.

Table 1

Generation of decadic equivalents (DE) assignment of the respective minterms and maxterms of two input variable Boolean function

DE	Binary code	Minterm	Maxterm
0	00	$\overline{x_0}\overline{x_1}$	$x_0 + x_1$
1	01	$\overline{x_0}x_1$	$x_0 + \overline{x_1}$
2	10	$x_0\overline{x_1}$	$\overline{x_0} + x_1$
3	11	x_0x_1	$\overline{x_0} + \overline{x_1}$

3.1.1 Input Vector Encoding

The hardware accelerator input is encoded using two binary vectors, A and X, where A consists of 2^n bits, $A(2^n - 1{:}\,0)$

$$A = (a_{2^n-1}, a_{2^n-2}, a_{2^n-3}, \dots, a_2, a_1, a_0) \tag{1}$$

To $\forall a_p \in A : p \in < 0; 2^n - 1 >$ it applies that $a_p \in \{0,1\}$

The order of bits in the A input binary vector is selected so that the bit in position p represents the output value of the Boolean function with a decadic equivalent equal to p. If the particular Boolean function output value is set to 1, the corresponding bit of the A input binary vector is set to the same value – 1. If the particular Boolean function output value having a decadic equivalent p is set to 0 or ×, the corresponding bit with the p position in the A input binary vector is set to 0.

The X vector also consists of 2^n bits: $X(2^n - 1{:}\,0)$

$$X = (x_{2^n-1}, x_{2^n-2}, x_{2^n-3}, \dots, x_2, x_1, x_0) \tag{2}$$

To $\forall x_p \in X : p \in < 0; 2^n - 1 >$ it applies that $x_p \in \{0,1\}$

The order of bits in the X input binary vector is selected so that the bit in position p represents the output value of the Boolean function with a decadic equivalent equal to p. If the particular Boolean function output value is set to ×, the corresponding bit of the X input binary vector is set to the value 1. If the particular Boolean function output value having a decadic equivalent of p is set to 0 or 1, the corresponding bit with the p position in the X input binary vector is set to 0.

If the truth table of the Boolean function defines outputs only from the {0, 1} set, only the A binary input vector creates input to the hardware accelerator input and the X binary input vector bits have to be set to 0.

3.2 Prime-Implicant-Set Encoding

The output of the hardware accelerator allows us to generate the set of prime implicants of the particular n input variable single-output Boolean function. The truth table of this Boolean function, encoded in vectors A and X, acts at the input of the hardware accelerator. The output of the circuit shows the prime implicants set for DNF or CNF form, depending on the values of the corresponding control signals.

The hardware accelerator output is encoded as the O output binary vector, consisting of $3^n + 1$ bits: $O(3^n : 0)$

$$O = (o_{3^n}, o_{3^n-1}, o_{3^n-2}, \dots , o_2, o_1, o_0) \qquad (3)$$

To $\forall o_p \in O : p \in < 0; 3^n >$ it applies that $o_p \in \{0,1\}$

The corresponding bit of the O output binary vector at its position p in the aforementioned vector shows whether the prime implicant having the p value of its decadic equivalent is or is not a prime implicant of the particular Boolean function. If the corresponding bit of the vector is set to 1, it is a prime implicant of the particular Boolean function. It is not a prime implicant of the particular Boolean function, this bit is set to 0. The decadic equivalent of 0 and 3^n is dedicated for the single-output Boolean functions producing a constant output 0 and 1 respectively, as shown in Table 3.

For the remaining decadic equivalents, one may find out the corresponding prime implicants by converting the specific decadic equivalent to a ternary code of n ternary digits (n is the number of input variables of the particular Boolean function). Then, each such ternary digit is encoded to the corresponding variable pursuant to Table 2, i.e. in DNF form, the variable in the prime implicant description is not used (if the ternary digit is set to 0), the variable is in complementary form (the digit is a 1), or the variable is in true form (the digit is a 2).

Table 2

Encoding variables in disjunctive normal form (DNF) and conjunctive normal form (CNF), depending on the digit of ternary equivalent (TE)

Ternary digit	DNF	CNF
0	□	□
1	$\overline{x_i}$	x_i
2	x_i	$\overline{x_i}$

In CNF form, the variable in the prime implicant description is not used (if the ternary digit is set to 0), the variable is in true form (if the digit is a 1), or the variable is in complementary form (if the digit is a 2). Assigning the variables to the respective digits of the ternary equivalent of the particular prime implicant to encode its description respects the order of the input variables in the truth table of the Boolean function.

A list of all decadic equivalents, their corresponding ternary equivalents and prime implicants for the DNF and CNF forms for a two-input Boolean function is specified in Table 3.

Table 3

Equivalence of decadic equivalents (DE), ternary equivalents (TE) and prime implicants (PI) for both the DNF and CNF forms

DE	TE	DNF PI description	DNF PI	CNF PI description	CNF PI
0	00	$\_\_\_\_$	0	$\_\_ + \_\_$	1
1	01	$\_\_\overline{x_1}$	$\overline{x_1}$	$\_\_ + x_1$	x_1
2	02	$\_\_x_1$	x_1	$\_\_ + \overline{x_1}$	$\overline{x_1}$
3	10	$\overline{x_0}\_\_$	$\overline{x_0}$	$x_0 + \_\_$	x_0
4	11	$\overline{x_0 x_1}$	$\overline{x_0 x_1}$	$x_0 + x_1$	$x_0 + x_1$
5	12	$\overline{x_0}x_1$	$\overline{x_0}x_1$	$x_0 + \overline{x_1}$	$x_0 + \overline{x_1}$
6	20	$x_0\_\_$	x_0	$\overline{x_0} + \_\_$	$\overline{x_0}$
7	21	$x_0\overline{x_1}$	$x_0\overline{x_1}$	$\overline{x_0} + x_1$	$\overline{x_0} + x_1$
8	22	x_0x_1	x_0x_1	$\overline{x_0} + \overline{x_1}$	$\overline{x_0} + \overline{x_1}$
9	100		1		0

3.3 Proposed Hardware Accelerator Module Design

The prime-implicant-generation module of the hardware accelerator for n variable single-output Boolean functions consists of 2^n prime-implicant-generation mode selection modules, the prime-implicant-generation module itself and $3^n + 1$ modules to exclude invalid prime implicants.

3.3.1 Prime-Implicant-Generation Mode Selection Module

The prime-implicant-generation mode selection module allows the user to select, whether to generate prime implicants consisting of the Boolean function outputs, where the function output is a member of the $\{0,\times\}$ set (for CNF) or the $\{1,\times\}$ set (for DNF) or the $\{\times\}$ set (to identify invalid prime implicants consisting exclusively of DC output values). Generation mode selection is possible using the $m0$ and f control signals, their effects are stated in Table 4.

Table 4

Accepted output values of the Boolean function for different $m0$ and f control signal settings when generating prime implicants for CNF, DNF forms and for identifying invalid prime implicants (IPI)

$m0$	f	Mode	Accepted output values of the Boolean function
0	0	IPI	$\{\times\}$
0	1	IPI	$\{\times\}$
1	0	CNF	$\{0,\times\}$
1	1	DNF	$\{1,\times\}$

If the f control signal is set to 1, DNF prime implicants are generated; a 0 setting of this signal indicates generation of CNF prime applicants. If the $m0$ control signal is set to 0, the accelerator shall generate information only concerning prime implicants for which the Boolean function output has always an × value (flagged as invalid prime implicants). If the $m0$ control signal is set to 1, the accelerator shall generate information only concerning prime implicants for which the Boolean function output is a 1 or an × (for DNF); or a 0 or an × (for CNF). The output of the OP module is set to 1 if the particular output value of Boolean function belongs to the accepted set of output values, as specified in Table 4.

For the hardware accelerator of an n variable single-output Boolean function the authors used 2^n of these modules to select the mode of prime implicant generation. Every pair of $a_p \in A$ and $x_p \in X$ bits of the accelerator input binary vectors (where $p \in < 0;\ 2^n - 1 >$ represents their position in the vector) is the input of the corresponding prime-implicant-generation selector module; in Fig. 1, particular inputs are denoted as the IA and IX. Further inputs of the module include the $m0$ and f control signals. The module output, denoted as OP, is a bit of the $P(2^n - 1{:}\,0)$ vector having the p position in the vector; this shows whether the particular Boolean function output shall be included in the prime-implicant-generation in the particular generation mode (the OP value is set to 1) or it will be excluded from the generation (if the OP value is set to 0).

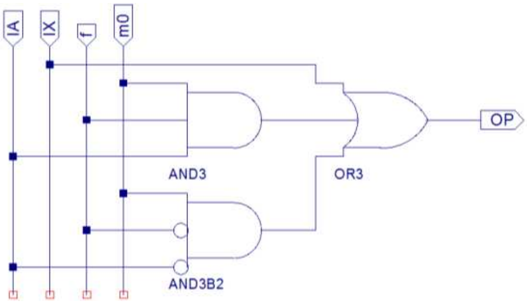


Figure 1
Combinational module for the selection of prime implicant generation mode

3.3.2 The Main Prime-Implicant-Generator Module

The input of the prime-implicant-generator module, a combinational logic circuit, is the vector $P(2^n - 1{:}\,0)$ – for a description of its computation, please refer to the previous subsection. The output of the module is the $R(3^n{:}\,0)$ binary output vector. The module consists of $l = n + 1$ layers of NAND gates, representing the potential prime implicants of the n variable Boolean function. The l_0 gate layer, containing two NAND gates, determines whether the Boolean function has a constant output value of 0 or 1. Layers l_1 – l_n contain gates representing the respective potential prime implicants (PPI). Particular NAND gate residing in layer $l_y : y \in < 1; n >$ represents the prime implicant described using y variables.

The total count of these gates in layer l_y equals to the number of potential prime implicants of the particular n variable Boolean function that may be described using y variables.

The gate in layer l_y receives information from the P input vector and from the $l_z : z \in < 1; y - 1 >$ gates layers, containing gates for the potential prime implicants described by a number of input variables lower than the particular prime implicant, specifically from that part of the gates that cover the particular prime implicant. The output of the gate is set to 1 if the particular potential prime implicant is not a prime implicant of the particular Boolean function, or, to 0, if the potential prime implicant is the prime implicant of the particular Boolean function. Before constructing the R output vector, the output signal of each NAND gate is inverted to ensure that the R output vector of the module contains a bit set to 1 if the particular potential prime implicant is really a prime implicant of the particular Boolean function.

To allow a potential prime implicant to be a real prime implicant of the particular Boolean function, three conditions must be met:

- **Condition 1:** Each bit of vector P representing output values of Boolean function which are relevant for particular prime implicant, must be set to 1.

  Meeting this condition may be tested using the information gained from the P input vector of the module.

- **Condition 2: T**he Boolean function must not produce a constant value at its output.

  Meeting this condition may be tested using the information generated in the l_0 gate layer.

- **Condition 3:** The potential prime implicant must not be covered by any other prime implicant (described with a lower amount of variables).

  Meeting this condition may be verified in case of a gate in layer l_y by acquiring the information from the respective gates of layers $l_z : z < y$.

The schematic representation of the prime-implicant generator module of two-variable Boolean function is stated in Fig. 2, showing the input of the module as an input layer and three levels of NAND gates in levels 0 to 2. Layer 0 contains two gates that indicate whether the Boolean function has a constant output of 0 or 1. Layer 1 contains four gates for the potential prime implicants, interpreted for the purposes of DNF as $\bar{a}$, a, $\bar{b}$, b. Layer 2 contains four gates for the potential prime implicants, interpreted for the purposes of DNF as $\bar{a}\bar{b}$, $\bar{a}b$, $a\bar{b}$, ab.

The meaning of the respective bits of the module's R output vector is analogous to the meaning of the respective bits of the accelerator's O output vector, as stated in section 3.2 above.

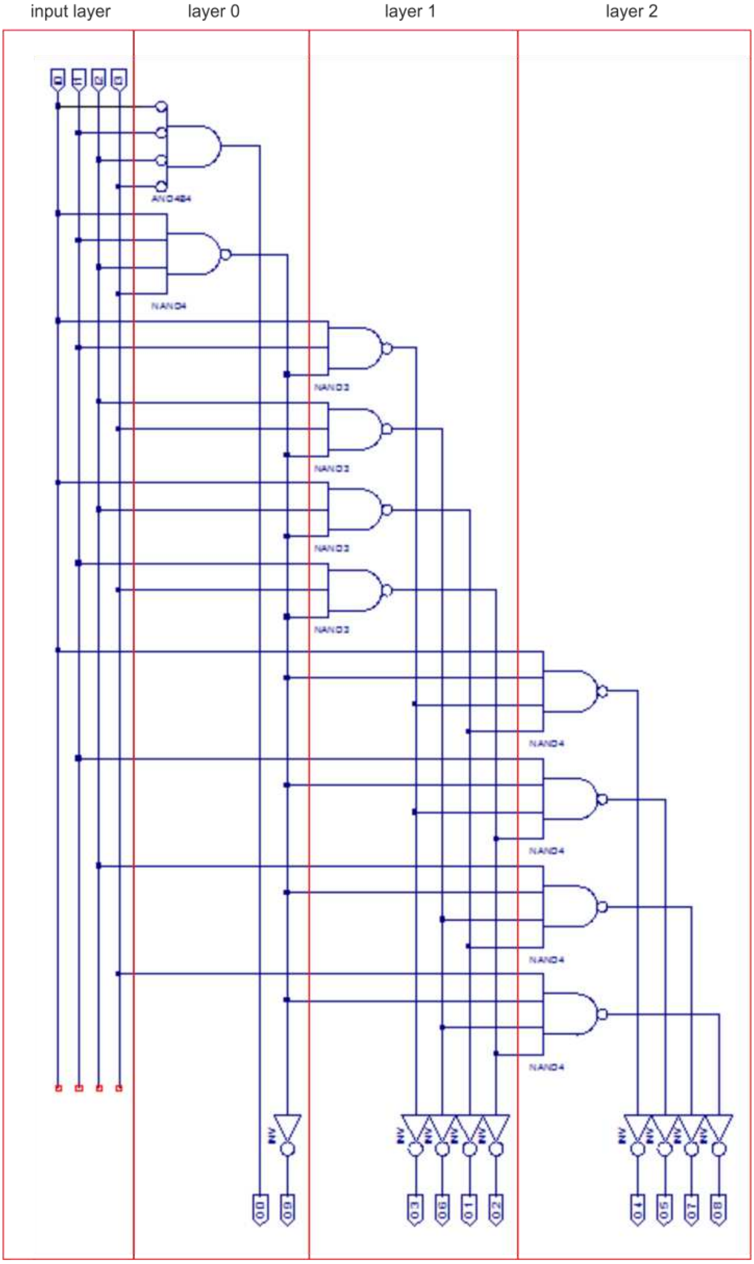


Figure 2
Gate-level schematic representation of the design of FPGA hardware accelerator module that determines prime implicants on the output of the module in the form of the R output binary vector for a two-variable Boolean function, represented on input of the module in the form of binary vector P.
Source: Madoš et al. [6]

3.3.3 Invalid-Prime-Implicant-Exclusion Module

The invalid-prime-implicant-exclusion module ensures that no prime implicant, for which the Boolean function output values belong exclusively to the DC set, is included in the final set of prime implicants of the particular Boolean function. For an n variable Boolean function, 3^n+1 modules were used to exclude invalid prime implicants. For each bit of the output vector R of the prime-implicant-generator module, such invalid-prime-implicant-exclusion module was used. The corresponding bit of the vector R at position p, is assigned to the input of the specific invalid-prime-implicant-exclusion module at the position p, brought to the input denoted as IR.

Depending on the setting of the $m0$ and $m1$ control signals, respectively, the bit of the R output vector, assigned to the IR input, is stored in the flip-flop FDE_0 (if signal $m0$ is set to 1) or in the flip-flop FDE_1 (if signal $m1$ is set to 1), as stated in the Fig. 3.

By setting signal $m0$ to 1, the circuit will generate information concerning all prime implicants, i.e. both valid and invalid. The corresponding bit at the IR input will be stored in the flip-flop FDE_0 in this case.

By setting signal $m1$ to 1, the circuit will generate information concerning invalid prime implicants, i.e. those covering the outputs of the Boolean function, in which the output is solely from the DC set. The corresponding bit at the IR input is in this case stored in the flip-flop FDE_1 and the 1 value of this bit indicates the invalidity of the prime implicant.

The module output, having the form of the O_x signal is then set to 1 only if the value if flip-flop FDE_0 is set to 1, which indicates that the potential prime implicant belongs to the set of prime implicants of the Boolean function and the FDE_0 flip-flop does not indicate the invalidity of the prime implicant. The O_x output of the module at position p is then forming the corresponding bit of the accelerator's O output vector, while the position of the bit in this vector is also p.

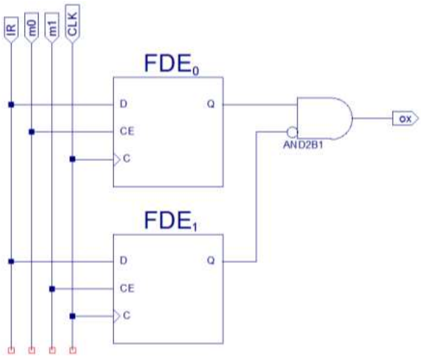


Figure 3
Schematic representation of the invalid-prime-implicant-exclusion module

If the Boolean function is defined to have an output values belonging solely to set $\{0, 1\}$, the set of prime implicants of the particular Boolean function may be acquired in a single step:

Step 1: setting the $m0$ control signal to 1, the $m1$ control signal to 0 and the f control signal to 0 for CNF and to 1 for DNF, respectively.

If the Boolean function is defined to have an output values belonging to set $\{0, 1, \times\}$, the set of prime implicants of the particular Boolean function may be acquired in two steps:

Step 1: setting the $m0$ control signal to 1, the $m1$ control signal to 0 and the f control signal to 0 for CNF and to 1 for DNF, respectively.

Step 2: setting the $m1$ control signal to 1 and the $m0$ control signal to 0.

With the first step, the accelerator finds out the set of prime implicants of a particular Boolean function containing valid and also invalid prime implicants (consisting solely of DC points). Therefore, the second step is executed, which yields a set of invalid prime implicants of the particular Boolean function, consisting solely of DC points. After the execution of the second step, the invalid-prime-implicant-exclusion modules ensure assembly of the output vector, containing only valid prime implicants of the particular Boolean function.

4 Results

The implementation language of the proposed modules is VHDL. As the target platform, the authors chose the use a Xilinx KC705 development board, using a Kintex-7 XC7K325T-2FFG900C series FPGA chip. The KC705 board used for synthesis has the speed grade -2 and a 2.5V LVDS differential 200 MHz oscillator. The output frequency could be changed within the range of 10 MHz to 810 MHz. The defined maximum clock speed limits the minimum response time, i.e. the time defined by the shortest clock period, in which any module implemented on the chip will work correctly (without violating the time constraints). With this FPGA chip, this value amounted to 1.23 ns.

The circuit synthesis was performed for Boolean functions with 2-8 input variables. A set of 7 top modules was created – these were implemented in VHDL using the Xilinx Vivado Design Suite HLx Edition 2016.2 development tool. The aim of testing the proposed module was to find out the hardware resource requirements of the synthesis and to measure the time required to generate the prime implicants.

Then, the authors compared the hardware resource requirements of the respective implementations of the particular top modules. The aim of the authors was to check if their expectations related to the resource consumption growth rate and response time growth rate were realistic. They expected that the resource consumption growth rate and time growth rate would directly correlate with the size of the input vectors defining the potential prime implicants of the Boolean function. Therefore, the authors expected the resource consumption growth rate of the implementation to be close to 3, since the number of potential prime implicants triples by adding a further input variable to the Boolean function and the response time growth rate to be much under 2. The time required to calculate the prime implicants using the particular modules was set using the minimum clock period allowing correct operation of the particular module. The authors also monitored the development of this characteristic in comparison with the input and output vector size.

As it has been stated in the previous sections, the module design was based on modules implemented as combinational logic circuits without any clock signal. Since specifying the minimum clock period using the Xilinx Vivado Design Suite HLx Edition 2016.2 tool requires using a flip-flop on both the input and the output side of the circuit, every top module had to be extended by a clock input. For testing purposes, the authors used the default circuit synthesis strategy, Vivado Synthesis Defaults 2016.

A summary of the implementation result may be found in the tables below. Figure 4 and Table 5 show the lookup table (LUT) and flip-flop consumption for the particular modules. The synthesis results confirmed a sub-linear increase in the number of consumed LUT resources, even though this was due to the increase of the AND/NAND gate input count (see also Figure 2), representing the prime implicants, the growth rate of the consumed LUT resources exceeds 3. As it is evident from Figure 4, there is a slight oscillation in the growth factor of the consumed LUT resources. The LUT resource consumption growth rate oscillation is caused by the Vivado synthesis tool, which uses an LUT-optimization technique to combine 3-input and 4-input LUTs to 5 and 6-input LUTs, implemented in Kintex-7 FPGA chips. Table 6 and the Figure 5 show a timing report of implemented modules. The authors focus their attention on the data path delay. The data path delay is the delay measured on the data path from the source to the destination. It indicates the module speed; in other words, it defines the response time. The results show that growth rate of the response time is much lower than 2, as was expected.

However, it is worth noting that the particular times define the minimum clock period on condition of implementing the computation for Boolean functions having an output from the set $\{0, 1\}$ solely; in this case, the computation time is the one stated in the table 5. Of the computation is implemented for Boolean functions with the output values from the set $\{0, 1, \times\}$, two computation steps have to be performed, as it has been stated in section 3 above and so the response time doubles.

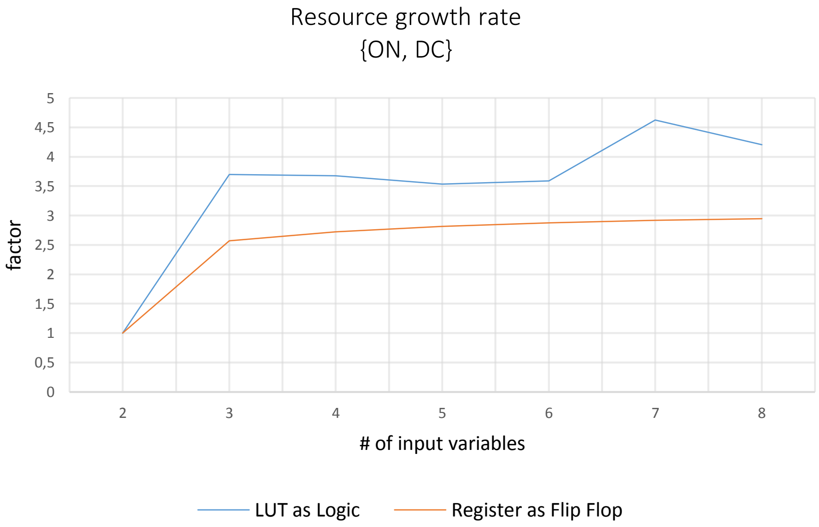


Figure 4
Resource consumption growth rate

Table 5
Summary of resource utilization and timing

Variable count	Resource Utilization		Timing Summary	
	LUT as Logic	Register as Flip Flop	Data Path Delay (Max Delay Path) [ns]	Logic Levels
2	20	28	2.818	2
3	74	72	4.396	3
4	272	196	7.77	6
5	962	552	11.336	8
6	3453	1588	14.328	10
7	15960	4632	21.88	27
8	67080	13636	30.052	39

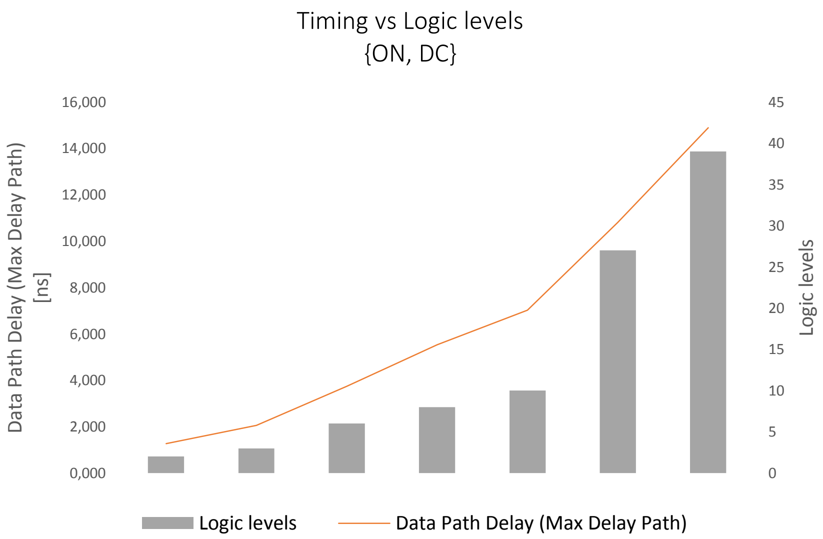


Figure 5
Timing summary

Conclusions

In this paper, the authors focused on the issue of accelerating Boolean function minimization. Systematic minimization, such as, the visual minimization method using Karnaugh map or the Quine-McCluskey algorithm implemented as a program, are two-step methods. The first step is the systematic generation of all prime implicants of a particular Boolean function. The second step is finding coverage of a particular Boolean function using the least possible prime implicants.

The work herein, is based on the previous development in this field [6], that proposed a combination logic circuit allowing the execution of the first step of systematic Boolean function minimization, i.e. allowing the generation of prime implicants of Boolean functions, on condition the particular function had fully defined output values. Defining DC output of the Boolean function was impossible. The solution proposed in this paper is an enhancement of the previous work, in which the possibility to select the mode of prime-implicant-generation for the DNF or CNF forms was added, along with the possibility to define DC outputs of the Boolean function. To generate prime implicants, the proposed hardware accelerator uses combinational logic; if the output values of the Boolean function belong only to the ON and OFF sets, it allows the generation of the prime implicant set of a particular Boolean function in a single step. If the output values of the particular Boolean function belong to the ON, OFF and DC sets, prime implicants will be generated in two steps. First, the prime implicants are generated, including those for which the Boolean function has only DC output values, i.e. they don't belong to the set of valid prime implicants of the particular Boolean function. In the second step, these invalid prime implicants are identified and then excluded from the set of prime implicants. The output of the proposed hardware accelerator is then a vector concerning valid prime implicants of the particular Boolean function.

The aim of this paper was to create a solution with exceptionally low time requirements, to generate the prime implicants of the particular Boolean function, which was achieved when the set of prime implicants could be generate for the tested Boolean functions in a matter of nanoseconds to tens of nanoseconds, the authors consider to be the main advantage of the proposed solution. Another advantage is that the time complexity depends only on the number of input variables of the Boolean functions and for the particular variable count, it is constant, regardless of the cardinality of the ON, OFF and DC sets. These advantages were achieved at the cost of spatial complexity of the proposed solution, thus, the implementation of the circuit is resource intensive, which is a disadvantage of the solution. The test of the proposed hardware accelerator, with its implementation for various amounts of input variables of the Boolean function, was performed using the Xilinx Kintex-7 FPGA KC705 Evaluation Kit evaluation board.

In future research, the authors shall focus on the possibility of decreasing the spatial complexity of the proposed solution, while maintaining the exceptionally low time requirements and allow the implementation of a further level of systematic minimization, i.e. the solution of Boolean function coverage using a FPGA hardware accelerator.

Acknowledgements

This research was supported by the Slovak Research and Development Agency, project number APVV-18-0214 and by KEGA Agency of the Ministry of Education, Science, Research and Sport of the Slovak Republic under Grant No. 003TUKE-4/2017 Implementation of Modern Methods and Education Forms in the Area of Security of Information and Communication Technologies towards Requirements of Labor Market.

References

[1] A. Zhaparova, D. Titov, A. Y. Balkanov, G. Gyorok, G. "Study of the Effectiveness of Switching-on LED Illumination Devices and the Use of Low Voltage System in Lighting", Acta Polytechnica Hungarica, Óbuda University, Budapest, Hungary, Vol. 12, Issue 5, pp. 71-80, 2015, ISSN: 1785-8860

[2] A. Baláž and R. Hlinka, "Forensic analysis of compromised systems," 2012 IEEE 10th International Conference on Emerging eLearning Technologies and Applications (ICETA), Stara Lesna, 2012, pp. 27-30

[3] V. Siládi and T. Filo, „Quine-McCluskey algorithm on GPGPU“, 3rd World Conference on Innovation and Computer Science (INSODE-2013) April 26-29, 2013, Antalya, Turkey, pp. 815-820, DOI: 10.13140/2.1.2113.1522

[4] V. Siládi, M. Povinsky, M. Povinsky, Ľ. Trajtel and M. Satymbekov, „Adapted parallel quine-McCluskey algorithm using GPGPU“, 2017 IEEE 14th International Scientific Conference on Informatics, November 14-16, 2017, Poprad, Slovakia, DOI: 10.1109/INFORMATICS.2017.8327269

[5] I. Savran and J. D. Bakos, „GPU Acceleration of Near-Minimal Logic Minimization“,Symposium on Application Accelerators in High Performance Computing, 2010

[6] B. Madoš, Z. Bilanová, E. Chovancová and N. Ádám, “Field Programmable Gate Array Hardware Accelerator of Prime Implicants Generation for Single-Output Boolean Functions Minimization”, ICETA 2019 – 17th International Conference on Emerging eLearning Technologies and Applications, November 21-22, 2019, The High Tatras, Slovakia

[7] Marquand, “XXXIII: On Logical Diagrams for n terms”, The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science. 5. 12 (75), pp. 266-270, doi:10.1080/14786448108627104

[8] H. H. Aitken, “Synthesis of electronic computing and control circuits”, Harward University Press, Cambridge, Massachusetts, 1951, p. 294

[9] E. W. Veitch, “A Chart Method for Simplifying Truth Functions”, Proceedings of the 1952 ACM Annual Meeting (Pittsburgh, Pennsylvania, USA) New York, USA: Association for Computing Machinery (ACM), pp. 127-133, doi:10.1145/609784.609801

[10] M. Karnaugh, “The Map Method for Synthesis of Combinational Logic Circuits”, Transactions of the American Institute of Electrical Engineers, Part I: Communication and Electronics. 72 (5), 1953, pp. 593-599, doi:10.1109/TCE.1953.6371932

[11] A. Svoboda, „Graficko-mechanické pomůcky užívané při analyse a synthese kontaktových obvodů” [Utilization of graphical-mechanical aids for the analysis and synthesis of contact circuits]. Stroje na zpracování informací [Symphosium IV on information processing machines] (in Czech) IV. Prague: Czechoslovak Academy of Sciences, Research Institute of Mathematical Machines. pp. 9-21

[12] M. V. Quine, “The Problem of Simplifying Truth Functions”, Amer. Math. Monthly, Vol. 59, 1952, No. 8, pp. 521-531

[13] E. J. McCluskey, “Minimization of Boolean functions”, The Bell System Technical Journal, 35, No. 5, Nov. 1956, pp. 1417-1444

[14] S. J. Hong, R. G. Cain and D. L. Ostapko, “MINI: A heuristic approach for logic minimization”, IBM Journal of Res. & Dev., Sept. 1974, pp. 443-458

[15] R. K. Brayton, G. D. Hachtel, C. T. McMullen and A. L. Sangiovanni-Vincentelli, "Logic Minimization Algorithms for VLSI Synthesis" (9th printing 2000, 1st ed.). Kluwer Academic Publishers. ISBN 0-89838-164-9

[16] R. L. Rudell, "Multiple-Valued Logic Minimization for PLA Synthesis", Memorandum No. UCB/ERL M86-65. 5th June 1986, Berkeley, p. 140

[17] Espresso source code, University of California, Berkeley, https://ptolemy.berkeley.edu/projects/embedded/pubs/downloads/espresso/

[18] P. McGeer, J. V. Sanghavi, R. K. Brayton and A. L. Sangiovanni-Vincentelli, "ESPRESSO-SIGNATURE: A new exact minimizer for logic functions", Proc. DAC'93, 1996, pp. 432-440

[19] P. Fišer and J. Hlavička, BOOM, "A Heuristic Boolean Minimizer", Computing and Informatics. Vol. 22, pp. 19-51, 25 June 2003, ISSN: 2585-8807

[20] P. Fišer and J. Hlavička, „Efficient Minimization Method for Incompletely Defined Boolean Functions", Proceedings of the 4th International Workshop on Boolean Problems, University of Mining and Technology, Freiberg, Germany), IWSBP 4, September 21-22, 2000, pp. 91-98, ISBN: 3-86012-124-3

[21] J. Hlavička and P. Fišer, A Heuristic Method of Two-Level Logic Synthesis. Proceedings of The 5th World Multiconference on Systemics, Cybernetics and Informatics ISAS-SCI'2001, Orlando, Florida (USA), July, 22-25, 2001, Vol. XII, pp. 283-288, ISBN 980-07-7541-2

[22] P. Fišer and J. Hlavička, "On the Use of Mutations in Boolean Minimization", Proceedings of the Euromicro Symposium on Digital Systems Design, Warsaw, Sep. 4-6, 01, pp. 300-307

[23] J. Hlavička and P. Fišer, BOOM — a Heuristic Boolean Minimizer. Proceedings of the 2001 IEEE/ACM International Conference on Computer-Aided Design, ICCAD 2001, San Jose, CA, USA, November 4-8, 2001, IEEE Computer Society 2001, pp. 439-442, ISBN 0-7803-7249-2

[24] P. Fišer and H. Kubátová, "Boolean Minimizer FC-Min: Coverage Finding Process", Proc. 30th Euromicro Symposium on Digital Systems Design (DSD'04), Rennes, 31.8.-3.9.04, pp. 152-159

[25] P. Fišer and H. Kubátová, Two-Level Boolean Minimizer BOOM-II, Proc. 6th International Workshop on Boolean Problems (IWSBP'04), Freiberg, Germany, 23-24.9.2004, pp. 221-228

[26] P. Fišer and H. Kubátová, Flexible Two-Level Boolean Minimizer BOOM II and Its Applications, Proc. 9th Euromicro Conference on Digital Systems Design (DSD'06), Cavtat, (Croatia), 30.8.–1.9.2006, pp. 369-376

Combined Vehicle and Driver Scheduling with Fuel Consumption and Parking Constraints: a Case Study

József Békési

University of Szeged, Juhász Gyula Faculty of Education, Department of Applied Informatics, Boldogasszony u. 6, 6720 Szeged, Hungary
bekesi@jgypk.szte.hu

Albert Nagy

Óbuda University, Applied Informatics, Bécsi út 96/b, 1034 Budapest, Hungary
albert.nagy@me.com

Abstract: Efficient operation is an important question for public transport companies, and that can most easily be achieved by reducing their operational costs. This can also be facilitated by the optimized scheduling of vehicles and the work of the drivers. Such an optimization task can be very complex. Due to the dramatically increased processing capabilities today, it can be performed using advanced optimization methods. Automation aims to reduce the time-consuming manual activities, thus increase efficiency and provide prompt opportunities of scenario planning for operational cost analysis purposes. In this paper, a case study is presented to solve the combined vehicle and driver scheduling problem. The applied mathematical model is discussed and the calculation results for practical examples are presented

Keywords: optimization; mathematical model; vehicle, crew and driver scheduling problem; public transport company

1 Introduction

Operational costs represent a large part of the expenses of public transport service companies. Their most important parts are vehicle fleet costs, fueling, and maintenance costs In addition, to driver salaries. Consequently, the budget can be improved significantly by decreasing these costs. The most commonly used technique to reduce these costs is the usage of a powerful, computer-aided information system. Due to the ICT (Information and Communication

Technology) development today almost every public transportation company has its own information system. In addition to the business applications, such as accounting, these systems may also contain modules such as

- scheduling vehicles and drivers for the company-served lines,
- monitoring the work of the vehicle fleet during the day,
- notifying the dispatcher of the unusual events (malfunctions, delays, etc.),
- tracking the status of the vehicles in the fleet,

and similar other functions. The ICT environment outlined above is often the organized backbone of efficient logistics management (see, for example, [1], but there is a next step in the process, which is to find the parts of the organization and operation that can be reduced in terms of operational costs. *Strategic planning* is used to determine the route of the buses. Elements of strategic planning are described by Desaulniers and Hickman in their review paper [2]. It discusses network planning, route planning, and passenger assignment based on expected travel needs. *Tactical planning* is used when the goal is to create an optimal schedule, and the planning can include scheduling the frequency of the trips. In addition, the fulfillment of requirements such as the capacity of the lines or the types of vehicles (e.g. on which line the low-floor vehicles should run) may also be addressed here. In the case of *operational planning*, buses and drivers are scheduled to provide the service. A number of solutions have been investigated to solve operational planning issues. Examining the problems from a theoretical point of view, most of them are NP-complete, which makes it difficult to find exact or near-exact solutions to problems that occur in practice. Even in cities with hundreds of thousands of citizens, the task is complex, if legislation, individual needs, and employee interests are taken into consideration. Such restrictive conditions may include vehicle characteristics, requirements relating to working time, driving time. and breaks for drivers, but in some cases. the constraints of the stations must also be considered. One of the major directions in the development of decision support systems over the past decades has been the development of software packages that provide a comprehensive solution to the various optimization tasks. However, practice shows that there are a number of company-specific expectations and constraints due to the specific situation of businesses on the application side, that are important to transport companies and can not be handled in a uniform way by the general systems developed. In this article, a decision support system developed for the Budapest Transport Corporation is presented. The purpose of the research was to integrate a module for solving a vehicle and driver scheduling task into the company’s information system. The method is based on an existing mathematical model, supplemented by special conditions required by the company. The article is structured as follows: In Section 2 a literature review on mathematical models is given for the driver and vehicle scheduling problem, then the problem and the solution method is introduced. In Section 3 the mathematical model is presented, then in Section 4.

the most important computational results are summarized. Finally, in Section 5 concluding remarks are given.

2 Materials and Methods

2.1 Literature Review

The scheduling problems of public transport are very complex. When looking at the problem from an operational research perspective, a global optimum is expected that minimizes the cost of both vehicle-related tasks and driver scheduling. A comprehensive survey of routing and scheduling problems of vehicles as well as crews is provided in [3]. It includes classification and categorization of routing and scheduling problems, a review of algorithmic techniques and solution methodologies. The solutions for the problems are, however mainly theoretical. Effective algorithms exist only for some of the tasks while for others the algorithmic capabilities described remain at low-level. Especially in case of bus public transport companies when the vehicle scheduling problem covers a given set of timetabled trips with consideration of practical requirements. The proposed modeling approaches in [3] are unable to solve real-world problem instances with thousands of scheduled trips by direct application of standard optimization software. The time-space network model is often used to reduce the number of variables in the exact optimization model. This model is discussed in [4] that uses a time-space-based network flow model instead of connection-based network model. It involves multiple depots for vehicles and different vehicle types for bus scheduling problem. This approach leads to size reduction of the corresponding mathematical models compared to connection-based network flow. The model size has been substantially reduced through the aggregation of incoming and outgoing arcs. The optimal solution could not be resulted by any of the above exact approaches. A combination of the methods should be able to solve the large amount of problems of practical interest in acceptable running times. Integer linear programming approach using combinatorial optimization can be seen in [5]. In the most widely used models today, the vehicle scheduling problem is formulated as an integer multicommodity network flow problem. In this model, optimal scheduling can be calculated as a solution to an integer linear programming problem. The methods apply branch and cut and branch and cut and price respectively. The column generation techniques seem indispensable for both approaches. Column generation is developed to make it possible to solve the huge linear programs with up to million integer variables. These rules for selecting new columns are based on Lagrangean relaxations and therefore called Lagrangean pricing. Other models also exist to solve this problem. The problem can also be formulated as a set partition problem with side constraints, whose continuous relaxation can be solved by column generation. (see for example [6]). A relationship is established between the bounds obtained by the assignment relaxation, the shortest path relaxation, the additive technique, Lagrangian decomposition, and column generation. It is shown that the additive bound technique cannot provide tighter bounds than those obtained by Lagrangian decomposition and not better than the linear programming bound. As in [7] the introduction of variable fixing, cutting planes, and the mixed branch-and-bound algorithm and the best-then-depth strategy leads to substantial improvements in the performance of a column generation algorithm to solve the scheduling problem. Dávid and Krész proposed a heuristic method [8]. The disadvantage of the models discussed and used in the literature in a specific, practical situation is that it only takes into account the rules relating to timetabled and overhead trips, bus types, and capacities required for them. However, it is not possible to include specific conditions that come from a real application environment. When planning the operational tasks of public transport, such typical vehicle-specific conditions are fuel consumption rules, various maintenance requirements (weekly, monthly, etc.), and parking rules. Parking rules may apply to both daytime and nighttime parking: where to park, what capacity the parking places have, etc. The length of the parking period may influence where in which geographical location it may be performed. The literature discusses several different versions of the vehicle scheduling problem, usually by the number of device types and the number of depots. The simplest version is the Single Depot Vehicle Scheduling Problem (SDVSP), where the vehicles belong to a single-vehicle type and are located in the same physical location. The first solution to the SDVSP problem was published by Saha [9]. The most commonly used model for solving a vehicle scheduling problem is the so-called Multiple Depot Vehicle Scheduling Problem (MDVSP). This case, which is more general than the single depot one, reflects the fact that in real life, different scheduled trips (their vehicles) may have different special needs. The vehicles are divided into different depots based on different vehicle types and the location of the vehicles. The MDVSP was defined by Bodin et al. and Bertossi et al. They showed that it is an NP-hard problem [11]. Tasks for vehicle-specific activities should also be taken into account, which requires a general framework for the integrated vehicle scheduling and assignment. A set partitioning-based mathematical model, where most vehicle-specific activities can be integrated based on the desired constraints is presented in [10]. This model is then solved using a column generation approach. The solution time can be reduced by the parallelization of the column generation process. If the transport company also uses alternative fuel vehicles in its fleet, the scheduling of these should take into account the number of kilometers per refueling, which may be much less than the mileage of conventional fuel vehicles. Alternative-fuel vehicles are getting more popular, and research is being done on how current infrastructure can serve them. The problem of vehicle scheduling consists of assigning a fleet of vehicles to service a given set of trips with start and end times. [12] presents the alternative-fuel multiple depot vehicle scheduling problem, a modification of the standard multiple depot vehicle scheduling problem where there is a given set of fueling stations, and a fuel capacity for the vehicles. The problem is formulated as a binary integer program, and exact column generation algorithm and a heuristic algorithm to solve the problem. [13] proposes a model for electric transit buses with either battery swapping or fast charging at a battery station, and a vehicle-scheduling model with the maximum route distance constraint for compressed natural gas, diesel, or hybrid-diesel buses. Both of these scheduling models are NP-hard. An important topic is to study the location problem of battery service stations. In [14] the developments of battery-electric buses are reviewed. A qualitative analysis on the strengths and weaknesses of each range method is conducted as well as costs and emissions of transit buses powered by different sources. Buses using alternative energy sources to reduce emissions, including some toxic air pollutants and carbon dioxide are studied in [15]. Life cycle comparison between buses fueled by different kinds of alternative energy is discussed in [16] to serve as an input to cost-benefit analysis.

Recently some other papers discussed the practical issues of vehicle scheduling. Dávid and Krész studied the handling possibilities of parking and maintenance constraints [17], and rescheduling possibilities in case of disruptions [18]. Parking and maintenance activities are handled in [19] as well.

Another important problem discussed in the literature is the Crew Scheduling Problem (CSP), also known as Driver Scheduling Problem or Duty Scheduling Problem. There are many CSP solution methods and applications in the literature. One of the best known is the so-called Generate and Select technology. The method can be summarized as follows: in the first step, generate a large number of regular shifts, and then in the selection step, look for a subset of them that is optimal in cost and covers the trips. Phase one requires significant computation time. The amount of calculation depends greatly on the number of trips and the complexity of the rules. In addition, the computational complexity of the rules greatly influences the complexity of this phase, and thus the whole problem. The problem can be defined as a set covering or set partitioning task. The partitioning model is a constrained version of the covering problem where no overlap is possible. This corresponds exactly to the real problem, but in this case, the existence of a feasible solution is not guaranteed. Note that both tasks have been shown to be NP-hard [23]. There are many ways to solve this problem. A hybrid approach incorporating a genetic algorithm is presented in [20]. It derives a small selection of good shifts to seed a greedy schedule construction heuristic. A group of shifts called a relief chain and used by the genetic algorithm for schedule construction. [21] simulates the self-adjusting process for driver scheduling. It incorporates the idea of fuzzy evaluation into a self-adjusting process, combining the features of iterative improvement and constructive perturbation, to explore solution space effectively and obtain superior schedules. A flexible system for scheduling driver applying integer programming methods are used in [22] and heuristic solution techniques are used in [24, 25].

According to the conventional approach, driver scheduling is performed after the phase of vehicle scheduling. Therefore, this is also called a sequential method. However, if the vehicle schedules are too dense, for example, there is not enough time to change drivers, then the problem can be infeasible in the driver scheduling phase. For this reason, the simultaneous optimization of the vehicle and driver scheduling may be reasonable. This problem is called Vehicle and Crew Scheduling Problem (VCSP). In 1999 Haase and Friberg published the first algorithm providing the exact solution for the single-depot case [26]. In their model, an integrated mathematical formulation was given in such a way that both sub-tasks were defined as set partition problems. The vehicle scheduling part is based on the model given by Ribeiro and Soumis [27], while the driver scheduling part uses the ideas of Desrochers and Soumis [28]. In the multi-depot case, first Gaffi et al. [29] discussed the integrated problem, using a heuristic method. In 2005, Huisman et al. [30] successfully extended the former models and algorithms of the single-depot case to the multi-depot version. This was the first general mathematical formulation of the multi-depot problem. Huisman in [31] also discusses the corresponding Lagrangian relaxations and Lagrangian heuristics. To solve the Lagrangian relaxations, column generation is applied to set partitioning type models. Haase [32] presents an exact approach for solving the simultaneous vehicle and crew scheduling problem in urban mass transit systems. This approach relies on a set partitioning formulation for the driver scheduling problem that incorporates side constraints for the bus itineraries. The proposed solution approach consists of a column generation process integrated into a branch-and-bound scheme. In 2008, Mesquita and Paias [33] also provided two mathematical formulations for this problem. In 2019 Horváth and Kis [33] proposed a novel mathematical programming formulation that combined ideas from known models and presented a solution methodology based on branch-and-price. In 2010 Steinzen et al. [34] gave another fully integrated VCSP approach, where the underlying vehicle scheduling model was based on the time-space network technique.

2.2 Problem Definition and Requirements

The problem is to automatically calculate optimal or approximately optimal vehicle and driver schedules for a given list of trips based on the master data and the company specific requirements and parameters in compliance with labor regulations. The optimality is measured by a given objective function and the aim of the optimization is to increase economy and efficiency. The developed model should take into consideration the following characteristics of the schedule planning for city buses:

- the problems are given by packages,
- the trips of a package can belong to a single line or a group of lines,
- the trips of a package can overlap 2 days,

- different day types are possible (e.g. working weekday, feast day, school day, etc.),
- deadhead trips from or to the depots can be possible from each end station,
- the maximum number of depots in a package is 5,
- the maximum number of vehicles that can be used in the solution of a package is 30,
- the maximum number of vehicle types in a package is 4,
- the type of vehicle can determine the fuel consumption, which should be taken into consideration in the schedule,
- breaks with a standoff or driver change are possible,
- standoff or parking is possible on more end stations or depots,
- parking place capacities are given for end stations in 5 minute intervals,
- driver change can be possible on given end stations,
- the labor regulations can be defined by several parameters.

The packages contain the following information:

- the lines and their parameters,
- the end stations, the depots, and their parameters,
- the trips,
- the parameters of the labor regulations and break rules.

The solution of the problem given by a package must satisfy the following requirements:

- each trip should be assigned to exactly one vehicle and driver schedule,
- the solution cannot use more vehicles than the number of available vehicles given in the package,
- the vehicle and driver schedules should be designed properly, adding the necessary activities (e.g. deadhead trips, breaks, maintenance, etc.)
- only such route can be set up for the vehicle (including deadhead trips to the depot) that can still be completed in terms of fuel consumption,
- between two passenger trips at least the required technological and compensatory time given for the line must be completed,
- drivers have rest periods while on duty, breaks with astandoff, vehicle change or both can be used based on the given parameters,
- driver change is possible if it is allowed for the given line on the given end station,
- in the case of a standoff, the parking space capacity should be checked,

- there is a possibility for divided working time, when the daily work of the driver is divided into 2 separated parts with a longer break between them,
- the generated driver schedules must comply with all work and rest regulations,
- the objective of the optimization process is to minimize the total net working time of the drivers, but if possible, the average of the drivers' net working times should be between 7 and 9 hours.

There are many conditions that are determined by the applicable legislation and the rules and regulations applicable to the employees of a given transport company. These include maximum working times, sufficient breaks after a given driving time, mandatory rest periods between two work periods, etc. In addition to scheduled and deadhead trips, there is a variety of technical and administrative tasks with well-defined timescales. These include passengers get on and get off times at end stations, vehicle pick-up, stopping, parking and various technological times, etc. In the following there is an overview of the most important working rules that are applicable to the drivers employed by the Budapest Transport Corporation. The default time values can be given as parameters.

1) The minimal length of a daily schedule (the minimal working time) is 4 hours without breaks.
2) The maximal length of a daily schedule (the maximal working time) is 10:30 hours with breaks.
3) The driving time cannot be longer than 9 hours in a schedule.
4) Vehicle oriented working times:
    a) depot release time: 25 minutes,
    b) end station release time: 20 minutes,
    c) release time of the vehicle stopped by the same driver: 10 minutes,
    d) vehicle change: 25 minutes,
    e) driver change on a vehicle: 5 minutes,
    f) stopping time at an end station: 5 minutes,
    g) stopping time at a depot: 20 minutes.
5) Rules for divided work (the driver schedule is divided into two periods, usually morning and afternoon periods).
    a) The break between the two parts cannot be shorter than 2 hours or longer than 6 hours.
    b) The total length of the two parts cannot be longer than 10:30 hours.
    c) The total length of the two parts with the break cannot be longer than 14 hours.
    d) The working time of each part cannot be shorter than 2 hours.

6) Break rules
    a) No break is necessary for a schedule shorter than 6 hours.
    b) 30 minutes break is necessary for a schedule longer than 6 hours and shorter than 9 hours.
    c) For a schedule longer than 9 hours at least 40 minutes break is necessary, which can be divided into 2 parts (20+20 minutes).
    d) A break must not be given in the first or last hour of a schedule.
    e) The continuous working time cannot be longer than 6 hours in a schedule.
    f) A break is not part of the working time.

3 The Mathematical Model

A combined vehicle and driver scheduling optimization model is used to solve the problem described in the previous chapter. Several versions of this are known in the literature. The mathematical formalization is a modified version of the model described in the paper published by Huisman et al. [30].

The most important components of the model are the set of timetabled trips and the available vehicles. Timetabled trips are all trips where vehicles carry passengers. Each such trip is determined by the departure and arrival times, the departure and arrival stations, and the distance between them. The vehicles are divided into (physical) depots based on their location. This may be the garage, parking space or site where the vehicle is parked. Vehicles can also have various important features that allow us to group them further. Based on physical (geographical) locations and features, vehicles are classified into disjoint subsets. The subsets thus formed are called depots. In addition to timetabled trips, vehicles must also carry out other types of drivings. These are called overhead trips. For example, for the first trip of the day the vehicle must leave the night parking lot and return after the last trip of the day. Such deadhead trips can occur during the day, e.g. for longer breaks. Typical additional overhead trips are when a vehicle goes to another station after completing a trip to make another trip from there. It also needs to allow these in order to get an effective schedule. For each timetabled trip, the user can specify the depots from which the trip can be served. In practice, this might mean, for example, that certain lines may be served by given type of buses or from given locations. These requirements may be determined by the location of the station and the characteristics of the traffic. It can define certain relationships between timetabled trips. Two trips are said to be compatible if after finishing the first trip the vehicle is able to arrive at the place of departure of the second trip in time. If the first trip arrives at the destination location of the second trip, the only condition is that the first trip arrives earlier as the second trip. If the arrival station of the first trip is not the same as the departure station of the second trip, the overhead time between the stations must be taken into account. There may be rules that include mandatory technological time between two trips, which should be taken into account when examining compatibility. The network is described by a directed graph. The nodes represent the trips, to which the departure and arrival depot nodes are added. Because multiple depots problem is handled, more depot nodes are possible. This corresponds to the usual technique. Two nodes of the graph are connected by a directed arc if the trips representing them are compatible. The arc length always represents the net working time of the driver, which corresponds to the objective function. It connects the depot nodes to the appropriate trip nodes, which can be accessed from that depot. Huisman et al. classifies the arcs into two main groups, namely short and long ones. Short arcs always represent the shorter events when the driver remains with the vehicle, while the long arcs represent the events when the driver stops the vehicle in the parking space and it remains unattended. Such a graph is built, which better represents the real situation, see Figure 1.

The following node types are introduced:

- source_depot: nodes representing the source depots,
- sink_depot: nodes representing the sink depots,
- trip: nodes representing timetabled trips.

There are 11 types of arcs in the model:

- start_of_schedule: arc representing the sign-on event with the first deadhead trip to the location of the first timetabled trip of a schedule,
- end_of_schedule: arc representing the last deadhead trip of a schedule to the location of the depot with the sign-off event,
- short_wait: arc representing a short wait after completing a timetabled trip, when the driver remains with the vehicle,
- short_break_endstation: arc representing a break of the driver spent at an end station after completing a timetabled trip,
- short_break_depot: arc representing a break of the driver spent in a depot after completing a timetabled trip,
- short_driverchange_endstation: arc representing a driver change event at an end station after completing a timetabled trip,
- short_driverchange_depot: arc representing a driver change event in a depot after completing a timetabled trip,
- long_stop_endstation: arc representing a driver change with a long parking of the vehicle at an end station without attendance,
- long_stop_depot: arc representing a driver change with a long parking of the vehicle in a depot without attendance,

- long_dividedstop_endstation: arc representing a long parking of the vehicle at the end station while the driver works in a divided schedule,
- long_dividedstop_depot: arc representing a long parking of the vehicle in the depot while the driver works in a divided schedule.

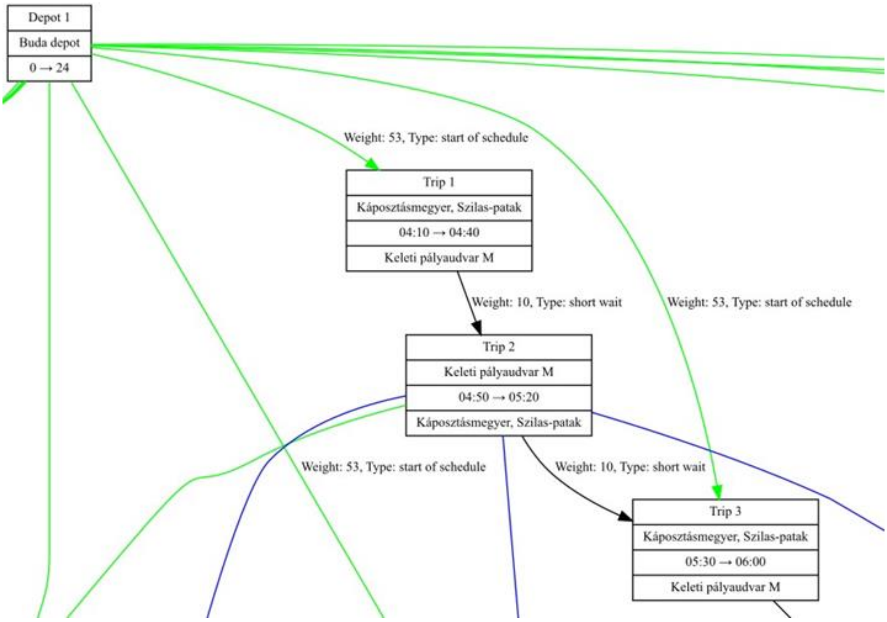

Depot 1
Buda depot
0 → 24
Weight: 53, Type: start of schedule
Trip 1
Káposztásmegyer, Szilas-patak
04:10 → 04:40
Keleti pályaudvar M
Weight: 10, Type: short wait
Weight: 53, Type: start of schedule
Trip 2
Keleti pályaudvar M
04:50 → 05:20
Káposztásmegyer, Szilas-patak
Weight: 53, Type: start of schedule
Weight: 10, Type: short wait
Trip 3
Káposztásmegyer, Szilas-patak
05:30 → 06:00
Keleti pályaudvar M

Figure 1
Part of the network

3.1 The Main Steps of the Calculation Process

1) In the first phase of the calculation, the appropriate input data and parameters are read and the graph is created. Here, the technological and compensatory times are taken into account, as well as certain rules regarding the duration of the waits or breaks. These can also be controlled by parameters. This means, for example, that if two trips are too close in time, those will not be connected, which means that those will not be executed one after the other by the same vehicle or driver. The first step in generating a graph is to create nodes based on the timetabled trips, location information, and vehicle types received. Then. add the above-listed arcs, also taking into account the parameters and labor rules. There may be different types of parallel arcs that have different weight values. For example, the driver may take a break at the end station or in a depot, depending on parking possibilities and station capabilities. However, these mean different working times, so if a model is to reflect the real situation fairly well, it needs to include both options. This

increases the number of arcs, which itself would not necessarily be a problem in the size of the optimization model, but practical experience has shown that the number of possible driver schedules can be critically large in some cases and this must be handled in some way.

2) In Phase 2, if possible, all regular driver schedules are generated. There is already a check that takes into account all the rules specified in the specification. The generating process is done systematically by traversing the base graph with depth-first search. However, every vertex is checked on the fly, and if the subschedule already formed does not conform to any rule, then this branch is cut off. When a complete schedule is done during the generation, a final check runs, which only accepts the schedule if it is found regular. The generated driver schedules are stored. If too many schedules are generated, the process is stopped and a heuristic method is executed, to decrease the size of the problem. This method will be described later. Parameters can also be used to limit the number of schedules.

3) In Phase 3, the mathematical model is constructed, which is essentially an extended version of the already mentioned VCSP model described in detail in [18]. The model includes constraints on fuel consumption, parking places and manages different vehicle types and locations.

3.2 The Formal Description of the Model

The set of timetabled trips are denoted by $U = \{u_1, u_2, \dots, u_n\}$. Let D be the set of depots, and $D_u \subseteq D$ the depot set of trip *u*: this includes those depots, from which u can be served. Note that in this case, a depot can represent a combination of physical locations and vehicle types. Denote $U_d \subseteq U$ the set of those trips that can be served from depot d. For every $d \in D$ two nodes are defined $dt(d)$ and $at(d)$ representing that a vehicle starts at depot d and goes back there. The set of N is then defined as follows

$$N = \{u \in U\} \cup \{dt(d) | d \in D\} \cup \{at(d) | d \in D\}.$$

To give the arcs of a network, the following notations are introduced.

$$B_d = \{(u, u') | \mathrm{u}, \mathrm{u}' \in \mathrm{U_d} \text{ are compatible trips}\}, \forall d \in D.$$

Other deadhead trips corresponding to depot d are the first and the last trips of the vehicle from the depot or to the depot.

$$R_d = \{(dt(d), u), (u, at(d)) | u \in U_d\}, \forall d \in D.$$

This way the set of arcs can be defined belonging to depot d in the network:

$$A_d = B_d \cup R_d, \forall d \in D,$$

and the set of all arcs of the graph is

$$A = \cup_{d \in D} A_d.$$

After these preparations, it is ready to define the VCSP problem on the network $G = (N, A)$. For this it is necessary to define an integer vector x, which can be considered as a multicommodity flow. The dimension of the vector is equal to the number of arcs in the network. If the arc e belongs to the depot d $(e \in A_d)$, then the component of the vector corresponding to the arc $e \in E$ is denoted by x_e^d. The value of x_e^d will be 1 if the given arc is included in the schedule, otherwise, it will be 0. The first condition ensures that every timetabled trip is scheduled exactly once.

$$\sum_{d \in D_u, e \in u_d^+} x_e^d = 1, \quad \forall u \in U. \tag{1}$$

where u_d^+ denotes the set of those outgoing arcs of the vertex u which belong to A_d. It ensures that all vehicles return to a depot by the end of the scheduling period. In other words, if a vehicle belonging to a particular depot arrives at an end station, it must leave it.

$$\sum_{e \in u_d^+} x_e^d - \sum_{e \in u_d^-} x_e^d = 0, \quad \forall u \in U_d, \forall d \in D, \tag{2}$$

where u_d^- denotes the set of those incoming arcs of the vertex u which belong to A_d. When defining fuel conditions, it considers the distance in km that can be covered by a single refueling. This requires to sum up the distance traveled by the vehicles and then limit it by the values given in the parameters. First, assign new variables to each vertex of N, excluding the set $\{at(d) | d \in D\}$. It is denoted this vector by t. The component of t belonging to vertex v will be denoted by t_v. The inequality that calculates the running distance up to a given node can be given as follows:

$$t_{v'} \geq x_e^d (t_v + \delta_e), \quad \forall v \in N \backslash \{at(d) | d \in D\}, \forall e \in v^+,$$

where δ_e is the distance represented by arc e, v' is its tail vertex and v^+ is the set of all outgoing arcs of vertex v. Note that these constraints are nonlinear, but those can be easily linearized using the following form:

$$t_{v'} \geq t_v + \delta_e - (1 - x_e^d)L, \quad \forall v \in N \backslash \{at(d) | d \in D\}, \forall e \in v^+, \tag{3}$$

where L is a constant larger than the longest possible distance in the system. Further constraints are necessary to check if a running distance of a vehicle remains under its maximal possible value.

$$t_{v'} + \delta_e + x_e^d L \leq r_d + L \quad \forall v \in \{at(d) | d \in D\}, \forall e \in v^-, \tag{4}$$

where L is a constant larger than the longest possible distance in the system, δ_e is the distance represented by arc e, v' is its head vertex, r_d is the maximal possible running distance allowed for the vehicles in depot d and v^- is the set of all incoming arcs of vertex v. The next constraints ensure that the capacity of the depots is not violated.

$$\sum_{e \in at(d)^-} x_e^d \leq k_d, \quad \forall d \in D, \tag{5}$$

where k_d is the number of vehicles available in depot d. In the following, it presents the conditions for the driver schedules. The set of valid driver schedules generated in Phase 2 for depot d is denoted by S_d. It assigns a vector of variables y to the schedules, where the component $y_s^d = 1$ if $s \in S_d$ is included in the solution and $y_s^d = 0$ otherwise. $S_d(u) \subseteq S_d$ denotes the set of those driver schedules which contain the vertex $u \in U_d$. Similarly $S_d(e) \subseteq S_d$ denote the set of those driver schedules that contain the arc $e \in A_d$. Similarly to the vehicle part, the first equality ensures that each trip is included in exactly one driver schedule.

$$\sum_{s \in S_d(u)} y_s^d - \sum_{e \in u_d^+} x_e^d = 0, \quad \forall u \in U_d, \forall d \in D. \tag{6}$$

During the daily work, certain events require a driver to attend the vehicle, while others do not require this. This can be modeled in such a way that certain types of arcs should be covered by both vehicle and driver schedules, while others are covered only by vehicle schedule.

It is denoted by A_q^d the set of arcs of type short_wait, short_break_endstation, short_break_depot, long_dividedstop_endstation, and long_dividedstop_depot.

The arcs of A_q^d should be covered by both vehicle and driver schedules. This can be expressed by the following equality:

$$\sum_{s \in S_d(e)} y_s^d - x_e^d = 0, \quad \forall e \in A_q^d, \forall d \in D. \tag{7}$$

Note that if short breaks with vehicle changes are allowed, then these conditions can be changed to $0 \leq \sum_{s \in S_d(e)} y_s^d - x_e^d \leq 1$ for short break type arcs e. Similar links can be given between the deadhead trips from or to the depots and the driver schedules containing them.

$$0 \leq \sum_{s \in S_d(e)} y_s^d - x_e^d \leq 1, \quad \forall e \in dt(d)^+, \forall d \in D, \tag{8}$$

$$0 \leq \sum_{s \in S_d(e)} y_s^d - x_e^d \leq 1, \quad \forall e \in at(d)^-, \forall d \in D. \tag{9}$$

Let $T_1, \dots, T_m$ be the possible time slots for which parking capacities should be checked. Furthermore, denote P the set of those stations and parking locations that can be used by the vehicles and $A_l^d(T)$ the set of those arcs of A_d that covers time slot T at location l. Denote the parking capacity of location l in time slot T for vehicles of depot d by $c_l^d(T)$. The parking constraints look like this

$$\sum_{e \in A_l^d(T)} x_e^d \leq c_l^d(T), \quad T = T_1, \dots, T_m, \forall l \in P, \forall d \in D. \tag{10}$$

The constraints for the average working time can be given as follows:

$$\sum_{d \in D} \sum_{s \in S_d} (w_s^d - w_{max}) y_s^d \leq 0, \tag{11}$$

$$\sum_{d \in D} \sum_{s \in S_d} (w_s^d - w_{min}) y_s^d \geq 0, \tag{12}$$

where w_s^d is the net working time of schedule $s \in S_d$ and w_{max} and w_{min} are the given maximal and minimal average net working times resp.

Furthermore, assume the followings:

$$x_e^d, y_s^d \in \{0,1\}, t_i \geq 0, \quad \forall e \in A_d, \forall s \in S_d, \forall d \in D, \forall i \in N \backslash \{at(d) | d \in D\}. \quad (13)$$

After these preparations, it is are ready to give the mathematical programming formulation of the problem.

$$\begin{array}{ll} \text{minimize} & \sum_{d \in D} \sum_{s \in S_d} w_s^d y_s^d \\ \text{subject to} & (1) \ldots (13) \end{array}$$

As mentioned in the description of Phase 2 the main difficulty of the model is that in some cases the number of possible driver schedules can be extremely large. Usually, the theoretical models in the literature use column generation to find the exact optimum. However the company's main aim was not getting an optimal solution, rather a feasible solution that can be calculated in a relatively short time and which is good enough to be used in practice. This means that it should satisfy all the constraints and it should not be far from the optimum. The experiments showed that in the case of practical instances it is not always easy to find a feasible initial solution by a simple heuristic. A trip contraction procedure is applied to handle this situation. This method decreases the number of vertices and arcs of the graph, so the number of possible driver schedules is also decreased. The following greedy trip grouper algorithm is developed, see Figure 2. The number of trips that will be collected in a group is given by a parameter n. The procedure will create a new set containing the trip groups.

Algorithm 1 Greedy Trip Grouper

```
procedure groupTrips($n$:Integer, $S$ : Set of Trips)
    for all $T \in S$ do
        Next($T$) ← The closest compatible trip to $T$
    for all $T \in S$ do
        if Merged($T$) = false then
            actTrip ← $T$
            MList ← ∅
            for $i$ ← 1,$n$ do
                Merged(actTrip) ← true
                MList ← MList ∪ actTrip
                if Next(actTrip) = null then
                    actTrip ← Next(actTrip)
                else
                    Exit for
            Add MList to the output
```

Figure 2
Greedy Trip Grouper

4 Discussion

Several real problems were solved by the method. There was parts of the tasks that arose during the daily work of the company. The problems were selected by experts to represent the various cases that can happen in practice. The mathematical models were generated by the system and these were solved by an optimization solver. The following tables present the most important characteristics of the inputs and the results of the computations. Table 1 shows the most important characteristics of the problems, the number of trips, vehicles, vehicle types, and depots. Table 2 presents information about the models, the sizes of the graphs, and IP models and the number of valid driver schedules. Finally, Table 3 gives the details of the solutions. In some cases, the trip grouper heuristic was used to get a solution. The number of trips in a group is given in the table. There were two possibilities to stop the optimization process. If optimal solution had been received, then the solver finished normally. If the solution was not improving for a longer duration, then the solution process was stopped. The running times are also displayed in the table. These change on a large scale, from a couple of seconds to hours depending on the problems.

Based on the results, this method can automate manual vehicle and driver scheduling in large part. The size of the problem highly depends on the characteristics of the input, such as the number of trips, vehicle types and depots, and the average length of trips. In some cases; the number of valid driver schedules was too large to solve the original problem. The trip grouper heuristic was able to handle this situation in most of the cases. The running times are very diverse, but 14 of the 20 problems were solved in 30 minutes and 5 of them in 1 minute. These times include all phases, i.e. the graph and the driver schedule generation and the solution of the mathematical program. 12 problems were solved to optimality, without using the heuristic method, or optimization process stopping.

4.1 Computational Results

Table 1

Properties of the problems solved

Number	Number of trips	Number of vehicles	Vehicle types	Depots
1	162	10	1	1
2	193	10	2	2
3	87	6	2	1
4	444	11	1	1
5	95	3	1	1
6	201	7	2	2
7	329	21	5	2

8	134	6	2	2
9	229	6	2	1
10	116	5	1	1
11	134	6	2	2
12	167	9	5	1
13	196	15	2	4
14	98	7	1	2
15	245	9	1	1
16	811	20	2	2
17	123	5	1	1
18	811	20	1	1
19	162	10	1	1
20	149	7	1	1

Table 2

Properties of the models (* = Trip grouper is used)

	Graph			IP Model	
Problem	Vertices	Arcs	Driver schedules	Columns	Rows
1	164	3698	6448	10394	5720
2	108*	11023	127432	138947	27977
3	91	4164	46608	51038	9595
4	157*	4427	1053086	1057822	10600
5	99	1604	46947	48701	3978
6	209	22850	510926	534540	57263
7	349	10459	85869	97909	26709
8	142	5808	11495	17712	14249
9	233	8330	502702	511724	20834
10	118	781	91756	92720	1997
11	142	5808	11495	17712	14249
12	177	4309	57875	62791	10771
13	212	22612	19337	42940	48478
14	102	9836	119306	129443	24910
15	247	4374	160352	165221	9488
16	216*	12734	264298	278084	29791
17	133	9032	279244	288527	21386
18	216*	12499	369327	382253	31219
19	166	3763	27973	32065	9869
20	153	6440	91561	98304	16010

Table 3
Properties of the solutions

Problem	Solution status	Running time (sec)	Trip grouper
1	Optimal	8	No
2	Optimal	998	Yes (2)
3	Optimal	138	No
4	Optimal	11364	Yes (3)
5	Optimal	543	No
6	Optimal	5199	No
7	Optimal	49	No
8	Optimal	134	No
9	Optimal	4467	No
10	Optimal	42	No
11	Optimal	27	No
12	Optimal	35	No
13	Optimal	679	No
14	Optimal	1212	No
15	Stopped	6432	No
16	Optimal	17570	Yes (4)
17	Stopped	1506	No
18	Stopped	25170	Yes (4)
19	Stopped	1512	No
20	Stopped	1531	No

Conclusion

In this paper, the use of a combined vehicle and driver scheduling model is studied for practical problems. First, a literature review on mathematical models for the vehicle and driver scheduling problems are briefly reviewed, then the scheduling problems and the solution methodology are discussed. A real problem is presented in a case study. The mathematical model is described and the most important calculation results are summarized. Based on the experience during the case study, these kinds of methods can help the planning process of transport companies with the existing constraints. Research on advanced scheduling models for public transport management systems is proven to be a relevant area of further examinations. The expected results hold out a promise to improve the operative planning activities of public transport.

Acknowledgement

József Békési was supported by the EU-funded Hungarian grant EFOP-3.6.2-16-2017-00015 and by the scientific grant of Juhász Gyula Faculty of Education, University of Szeged.

References

[1] M. Meilton, Selecting and implementing a computer aided scheduling system for a large bus company, Algorithms: Combinatorial Analysis. In Computer-Aided Scheduling of Public Transport, (eds. S. Voss and J.R. Daduna), 203-214, Springer-Verlag, Berlin, 2001

[2] G. Desaulniers and M. D. Hickman, Public Transit, In Handbook in OR & MS, (eds. C. Barnhart and G. Laporte), Vol. 14, Chapter 2, Elsevier B. V., 2007

[3] L. Bodin, B. Golden, A. Assad and M. Ball, Routing and Scheduling of Vehicles and Crews: The State of the Art, Computers and Operations Research, 10, 63-211, 1983

[4] N. Kliewer, T. Mellouli and L. Suhl, A time-space network based exact optimization model for multi-depot bus scheduling, European Journal of Operational Research, 175, 1616-1627, 2006

[5] A. Löbel, Optimal Vehicle Scheduling in Public Transit, PhD. thesis, Technische Universitaet at Berlin, 1997

[6] C. C. Ribeiro and F. Soumis, A Column Generation Approach to the Multiple-Depot Vehicle Scheduling Problem, Operations Research, 42(1), 41-52, 1994

[7] A. Hadjar, O. Marcotte, and F. Soumis, A Branch-and-Cut Algorithm for the Multiple Depot Vehicle Scheduling Problem, Tech. Rept. G–2001–25, Les Cahiers du Gerad, Montreal, 2001

[8] B. Dávid and M. Krész, Application Oriented Variable Fixing Methods for the Multiple Depot Vehicle Scheduling Problem, Acta Cybernetica, 21(1), 53-73, 2013

[9] J. L. Saha, An algorithm for bus scheduling problems, Operational Research Quarterly, 21(4), 463-474, 1972

[10] J. Békési, B. Dávid, M. Krész, Integrated Vehicle Scheduling and Vehicle Assignment, Acta Cybernetica, 23(3), 783800, 2018

[11] A. A. Bertossi, P. Carraresi, and G. Gallo, On Some Matching Problems Arising in Vehicle Scheduling Models, Networks, 17, 271-281, 1987

[12] J. D. Adler and P. B. Mirchandani, The vehicle scheduling problem for fleets with alternative-fuel vehicles, Transportation Science, 51(2), 441-456, 2016

[13] J. Li, Transit bus scheduling with limited energy, Transportation Science, 48(4), 521-539, 2013

[14] J. Li, Battery-electric transit bus developments and operations: A review, International Journal of Sustainable Transportation, 10(3), 157-169, 2016

[15] J-Q. Li and K. L. Head, Sustainability provisions in the bus-scheduling problem, Transportation Research, Part D, 49, 50-60, 2009

[16] A. Rabl, Environmental benefits of natural gas for buses, Transportation Research, Part D, 7, 391-405, 2002

[17] B. Dávid and M. Krész, Multi-depot bus schedule assignment with parking and maintenance constraints for intercity transportation over a planning period, Transportation Letters, 12(1), 66-75, 2020

[18] B. Dávid and M. Krész, The dynamic vehicle rescheduling problem, Central European Journal of Operations Research, 25(4), 809-830, 2017

[19] A. Haghani and Y. Shafahi, Bus maintenance systems and maintenance scheduling: model formulations and solutions. Transportation Research Part A: Policy and Practice, 36(5), 453-482, 2002

[20] R. S. K. Kwan, A. S. K. Kwan and A. S. K. Wren, Evolutionary Driver Scheduling with Relief Chains, Evolutionary Computation, 9, 445-460, 2001

[21] J. Li. A Self-Adjusting Algorithm for Driver Scheduling, Journal of Heuristics, 11, 351-367, 2005

[22] A. Wren, S. Fores, A. S. K. Kwan, R. S. K. Kwan, M. E. Parker and L. Proll, A flexible system for scheduling drivers, Journal of Scheduling, 6(5), 437-455, 2003

[23] M. R. Garey and D. S. Johnson, Computers and Interactability: A Guide to the Theory of NP-Completness, Freeman, San Fransisco, 1979

[24] A. Tóth and M. Krész, A flexible framework for driver scheduling, In Proceedings of the 11th International Symposium on Operational Research, Slovenia, SOR'11, 341-346, 2011

[25] A. Tóth and M. Krész, An efficient solution approach for real-world scheduling problems in urban bus transportation, Central European Journal of Operations Research, 21(1), 75-94, 2013

[26] K. Haase and C. Friberg, An exact branch and cut algorithm for the vehicle and crew scheduling problem, In Computer-Aided Transit Scheduling, Lecture Notes in Economics and Mathematical Systems, 471, (ed. N.H.M. Wilson) 63-80, Springer, Berlin, 1999

[27] M. Horváth and T. Kis, Computing strong lower and upper bounds for the integrated multiple-depot vehicle and crew scheduling problem with branch-and-price, Central European Journal of Operations Research, 27(1), 39-67, 2019

[28] M. Desrochers and F. Soumis, A column generation approach to the urban transit crew scheduling problem, Transportation Science, 23(1), 1-13, 1989

[29] A. Gaffi and M. Nonato, An integrated approach to the extra-urban crew and vehicle scheduling problem, In Computer-Aided Transit Scheduling, (ed. N. H. M. Wilson), Lecture Notes in Economics and Mathematical Systems, 471, 103-128, Springer, Berlin, 1999

[30] D. Huisman, R. Freling and A. P. M. Wagelmans, Multiple-depot integrated vehicle and crew scheduling, Transportation Science, 39, 491-502, 2005

[31] R. Freling, D. Huisman, and A. P. M. Wagelmans, Models and algorithms for integration of vehicle and crew scheduling. Journal of Scheduling, 6, 63-85, 2003

[32] K. Haase, G. Desaulniers, and J. Desrosiers, Simultaneous vehicle and crew scheduling in urban mass transit systems, Transportation Science, 35(3), 286-303, 2001

[33] M. Mesquita and A. Paias, Set partitioning/covering-based approaches for the integrated vehicle and crew scheduling problem, Computers and Operations Research, 35(5), 1562-1575, 2008

[34] I. Steinzen, V. Gintner, L. Suhl and N. Kliewer, A Time-Space Network Approach for the Integrated Vehicle- and Crew-Scheduling Problem with Multiple Depots, Transportation Science, 4(3), 367-382, 2010

Innovating a Model for Measuring Competitiveness in Accordance with the Challenges of Industry 4.0

Andrea Okanović[1], Bojana Jokanović[1], Vladimir Đaković[1], Simonida Vukadinović[2], Jelena Ješić[2]

[1]University of Novi Sad, Faculty of Technical Sciences, Trg Dositeja Obradovića 6, 21000 Novi Sad, Serbia, e-mail: andrea.katic@uns.ac.rs; bojanajokanovic@uns.ac.rs; v_djakovic@uns.ac.rs

[2]Educons University, Vojvode Putnika 87, 21208 Sremska Kamenica, Serbia, e-mail: simonida.vukadinovic@educons.edu.rs; jelena.jessic@educons.edu.rs

Abstract: In the approaching time of the Fourth Industrial Revolution, our planet has undergone dramatic changes, that will leave its mark on all aspects of our life. For this reason, countries around the world have been challenged to reinstate or redefine their national strategies in order to adjust to the requirements of the new age. Policy - makers of today are expected to evaluate each country's readiness to adopt and implement the concepts underlying the Industry 4.0. Analyzing the existing models, it became apparent to the authors and other researchers that there is no suitable model that provides adequate information on the attitude of states towards the criteria of the fourth industrial revolution. For this reason, this paper proposes a new model consisting of 42 quantitative and 8 mixed indicators, 10 of which, directly relate to the characteristics of the new age that is before us. The model has been applied in 17 OECD countries, as it is currently best suited to measure the competitiveness of the most developed countries, which offer the most data within the parameters that describe the characteristics of the smart society of the future. Nevertheless, the authors of the paper believe that the presented model will, very soon, be applicable to a much wider range of countries, and above all, that it will be well suited for measuring the competitiveness of all European countries.

Keywords: Fourth Industrial Revolution; smart society; measuring competitiveness; competitiveness indices

1 Introduction

During the entire human existence on earth, technology has played one of the most crucial roles in the development of society and civilization. Today, at the beginning of the third decade of the 21st Century, the world is already considerably debating the “so-called”, Fourth Industrial Revolution, that is expected to introduce significant changes in the way people live and work on a global scale. Looking back on the past, we recall that the first industrial revolution was fueled by the invention of the steam-powered machine, while the second was marked by the use of conveyor belt in the industry and mass electrification. The automation of the production, digitization and use of information and communication technologies was brought by the achievements of the third technological or information revolution. It allowed people to own personal devices for communication and connection with a large number of people, access to information, data storage, control of bank accounts and more. The Fourth Industrial Revolution builds on these inventions and further develops the Internet of Things (IoT), artificial intelligence (AI), 3D printing, robotics, autonomous vehicles, quantum computing, nanotechnology and biotechnology, as well as, new ways to store energy. It is important to distinguish the AI applications, namely, based on the “deepness” of AI: Artificial Narrow Intelligence (ANI), Artificial General Intelligence (AGI) and Artificial Super Intelligence (ASI). Scientists are emphasizing that the Fourth Industrial Revolution, which brings together the physical, digital and biological characteristics of products, will bring dramatic changes to the world in the course of the next twenty years, such as have not been seen in the previous hundred [1]. According to a survey, as many as 65% of children who enroll in primary school today will be doing jobs that still do not exist today [2]. Authors Pereira & Romero define Industry 4.0 concept “as an umbrella term for a new industrial paradigm that embraces a set of future industrial developments regarding Cyber-Physical Systems (CPS), Internet of Things (IoT), Internet of Services (IoS), Robotics, Big Data, Cloud Manufacturing and Augmented Reality”. Industry 4.0 is being predominantly shaped by two main drivers: Cyber-Physical Systems and the Internet of Things and Services [3]. Today's highly equipped factories use autonomous robot for work in the places where human workers are restricted to work as well as to perform autonomous production method more precisely [4]. However, the purchase and use of autonomous vehicles and robots and the R&D activity of these new transportation devices might differ in countries. The same is applicable in the field of quantum computing, nanotechnology, biotechnology and energy storage.

The OECD countries brings together the most developed group of countries in the world, composed of 36 member countries today. In its strategies, like the OECD Jobs Strategy and the OECD Skills Strategy, this organization emphasizes that today's technological advancements have an impact on society, the economy and the way of life of people like never before, and that we are living in a transformative age where disruption is the new norm [5, 6]. It is important to emphasize that, in the context of dramatic changes in technology, the aforementioned strategies emphasize that sustainable development must be an integral element of the growth and achievement of high competitiveness of its member countries.

In order to trace the envisaged goals in the most appropriate way, at a time when the Fourth Industrial Revolution has already set its challenges, this paper proposes a new model for measuring competitiveness in accordance with the challenges of the Industry 4.0. The new model builds on previous research by the authors and offers an advanced selection of indicators in accordance with the requirements of the new wave of change [7, 8, 9]. Specifically, by examining existing traditional models for measuring macro-level competitiveness, it has been found that they do not sufficiently include indicators relating to the domain of industry 4.0 [7, 10, 11]. The main hypothesis of the research reads: The new model of competitiveness measurement provides a more adequate measurement of the position of today's most developed countries because it takes into consideration the challenges of the present as well as of the time that is before us, i.e. the Fourth Industrial Revolution. The subhypothesis of the paper reads: There are subindexes that have a small - scale range of variation and variance, as well as those with more pronounced differences between the worst and the best ranked economy. The mathematical and statistical research methods needed to structure and subsequently test the set model of competitiveness measurement were used in the paper. An analysis of the obtained results and discussions regarding the position of the countries included in the survey were completed, which is also seen as the outcome of this paper.

2 Conceptual Background

Industrial revolutions have brought upon the world, the economic development, growth of world wealth, increase of leisure fund as well as longer life span of people. Each new revolution brings with it many changes that represent a potential chance for the success of those who know how to manage them, but also a threat to those who do not possess the necessary skills. Today, in order to achieve high competitiveness, at all levels (micro, macro and meso), it is important to be accustomed to the world trends, as much as it is to take part in their creation, in order to secure the highest positions in the rankings. When it comes to measuring competitiveness, it is important to emphasize that sustained innovation of the existing models and rarely their reinvention is essential for being in accordance with the meet challenges of today as well as of tomorrow.

As far back as 1969, Drucker defined the most competitive society as the "Knowledge Society" or "Society of Mobility" [12]. Somewhat later, the OECD (1996) defined a knowledge-based economy as the "one in which production, distribution and use of knowledge are the main drivers of growth, wealth creation, and employment for all industries" [13]. It is during this period, but also in the coming years, that an expansion of various researches takes place, whose authors have tried to define the parameters that measure the success of states, or the indicators according to which the countries of a society based on knowledge, innovation and technological progress are ranked [13, 14, 15, 16]. In their previous studies, the authors of the paper extracted 23 composite indices containing the parameters of the *Knowledge Society* and, on the basis of further research, offered a model for measuring competitiveness at the macro level, consisting of 65 quantitative indicators [7, 8]. On this occasion, they pointed out that the most important parameters for measuring the success of a knowledge-based society are: high percentage of highly educated population, large government investments in education, science and research, promotion of lifelong learning, high quality and accessible information and communication infrastructure and services, propulsive and competitive economy, sustainable technological development, wide availability of information and easy access to them.

Further advancements of Science and Technology generates new changes to the world, as well as, the need to improve models for measuring competitiveness. The concept underlying the onset of a new revolution wave originated in Germany under the name "Industry 4.0", and the whirlwind soon spread to other highly developed countries such as the United Kingdom, which recognized it as the "Fourth Industrial Revolution" [17, 18, 19]. Industry 4.0 is also associated with terms like "smart factory", "smart manufacturing", "advanced manufacturing" and the like [10, 20]. The issue of competitive advantage of nations, regions and companies is a topic of crucial interest for policy makers, scientists and managers worldwide. Professor Klaus Schwab, Founder and Executive Chairman of the World Economic Forum, has published a book entitled The Fourth Industrial Revolution, outlining three groups of interconnected megatrends that will mark the future. These include physical, biological, and digital megatrends. Physically they include advanced robotics, autonomous vehicles, 3D printing and the development of new materials. Biological megatrends include biotechnology and genome projects. Digital megatrends refer to artificial intelligence, the Internet of Things, blockchain technology, cloud memory and virtual reality [1]. If the predictions are true, the consequences of the changes described will be multifaceted, and will primarily affect the global economy, demographics, education, quality of life and work, etc. Futurologists tell us that one third of today's children will live longer than 100 years because they will have better options for preventing and treating the diseases [21]. Furthermore, research shows that one has to be very careful about choosing a profession today, because, for example, the job of a journalist will be partially jeopardized by the possibility of popularizing news writing programs, which could replace more than 90% of practitioners, by 2025, by writing newspaper articles. Such changes would have implications for working life, the pension system, as well as, individual life planning [22]. According to Vacek, “the deep impact of Industry 4.0 on socio-economic issues can be called Society 4.0” [23].

Through a literature review several new models were discovered that rely on the latest developments in the technique (Table 1) [24, 25, 26, 27, 28]. However, although there are published models relating to evaluation of competitiveness in the context of the fourth industrial revolution at the macro level, it can be noted that they rely largely on qualitative data whose objectivity is difficult to verify. In previous research done by the authors it was shown that the qualitative indicators can be subject of manipulative influences of experts so it was suggested that quantitative parameters are more reliable measures for competitiveness models in general [29]. Other authors have reached a similar conclusion. Specifically, Batchkova et al. conclude that in the models they have analyzed, which refer to the competitiveness indices of Industry 4.0, there are no quantitative indicators describing the main concepts, and that they are used instead of qualitative ones, and that there is a high degree of unpredictability in the information on which this evaluation is based [30].

We can conclude, from the aforementioned, that with the advent of a new, fourth industrial revolution, the models proposed to measure the competitiveness of certain entities must be re-examined and improved. The paper below proposes a new model for measuring competitiveness based on the requirements of the fourth industrial revolution. However, the usage or development of Extended Reality (Virtual, Augmented and Mixed Reality) as a significant parameter, is not mentioned in the model, due to a lack of data.

Table 1
Industry 4.0 competitiveness index overview

Index name	Authors and year of publishing	Level	Data
The Singapore Smart Industry Readiness Index	Economic Development Board, 2017.	Micro level	qualitative data; 3 subindexes: Process, Technology, Organization; 16 indicators
Metamodel for Evaluating Enterprise Readiness in the Context of Industry 4.0	Basl, J., & Doucek, P, 2019.	Micro level	7 subindexes: Society, Area of society, Branch of area of society, Enterprise, Area of enterprise, Dimension of enterprise area, Subdimension of enterprise area
RB Industry 4.0 Readiness Index	Rolland Berger, 2014.	Macro level, 22 courtiers	qualitative data; 2 subindexes: Industrial excellence, Value network
Readiness for the Future of Production	World Economic Forum, 2018.	Macro level, 100 countries	qualitative and quantitative data; 6 subindexes: Technology & Innovation, Human Capital, Global Trade & Investment, Institutional Framework, Sustainable Resources, Demand Environment; 32 qualitative indicators and 27 quantitative indicators

Industry 4.0 Readiness Index	Danish Institute of Industry 4.0, 2017.	Macro level, 120 courtiers	quantitative data; 7 subindexes: Innovation aptitude, Demand factors, Driving forces, Enterprise excellence, Basic enablers, Technological sophistication, Industry 4.0 specific enablers; 24 indicators

Source: the authors

3 Methodology

The paper used a model based on the proposed methodology of well-known authors [14]. The survey involves the collection of secondary data obtained mainly from official statistical reports or from representative institutions. Further steps are related to the formation of thematic indicators, weighting, the calculation of average values, the processing of time series, and the use of regression and correlation analysis [14]. All in all, to form a composite index it is necessary to follow the following steps: development of a thematic framework; selecting indicators, adjusting irreversible data and replacing missing data; selecting a sample of countries; formation of thematic subindicators; standardization and weighting of indicators; aggregation and ranking of countries by subindicators; subindicator weighting; aggregation and formation of composite index; composite Index evaluation [14]. Finally, it is important to stress that the model does not necessarily adhere sequentially to all steps above, they are rather undertaken simultaneously, in many cases [31].

4 Data and Results

4.1 Construction of a Thematic Framework

The first step in the construction of a composite index is to define a theoretical framework that describes the phenomenon to be quantified. For this purpose, it is necessary to carry out a detailed literature review so that indicators that accurately measure macro-level competitiveness can be extracted later [14, 32].

The choice of indicators used in the research reflects the challenges of today and the future, that is, the fourth technological revolution. The presented model contains a set of indicators, within the subindex called *Smart Society*, representing world trends such as those measuring the use of autonomous vehicles, artificial intelligence, the use of robots, 3D printing, as well as IoT [10, 33, 34, 35].

Alongside them, traditional indicators are being used today to measure the use of information and communication technologies among the population and in enterprises, followed by indicators of a knowledge economy, innovation and R&D, as well as, indicators of sustainable development.

4.2 Selecting Indicators, Adjusting Irreversible Data and Replacing Missing Data

In order to make a relevant choice of indicators, several of their key features, such as validity, measurability and availability of data, need to be taken into consideration. These characteristics are very important because in practice it often happens that the reliability of the data itself is called into question, i.e., it is not known how certain organizations and institutions collected them. For this reason, care should be taken to use only the data published by the relevant authorities. When it comes to measurability, the problem arises with certain research-relevant phenomena for which there are no statistical data or quantitative indicators [29]. For this reason, many researchers resort to the use of qualitative data based on the opinions of a narrow circle of evaluation experts. However, as shown in the paper, their use has many drawbacks and can lead to erroneous evaluations, results and conclusions [29]. Third crucial feature related to data is their availability. The importance of this feature stems from the fact that certain data are very difficult to obtain i.e. not being publicly visible, institutions that evaluate them do not display them clearly or ask for large sums of money for their use.

The model presented in the paper consists mainly of quantitative indicators, while mixed indicators are far less used, i.e. data in the form of previously measured composite indices. In order to obtain an objective comparison between countries of different sizes, it is ultimately important to adjust the data according to population, income, land size, etc.

In order to present national competitiveness in the best possible way in light of the fourth industrial revolution, the proposed model contains a set of 50 indicators, of which 42 are quantitative indicators and 8 are mixed (previously measured composite indices) (Table 2). The paper uses official statistics from relevant institutions that are published in their statistical yearbooks or websites. The majority of data was obtained from institutions such as: the World Bank, the International Telecommunication Union (ITU), the UNECO portal - UIS.Stat and the OECD statistical portal, referring to calendar years 2017 and 2018. The major issue with the survey was the choice of parameters for the *Smart Society* subindicator, as many new trends are not yet covered by the statistical measurements of the relevant institutions.

Indicators numbered 12, 14, 23, 43, 45, 48 and 49 stand for irreversible measures in which lower values indicate a higher level of development. For this reason, it is necessary to transform them according to the following formula:

$$Xtrans = 2 * (Xmax - Xmin) - Xi \quad (1)$$

As has been pointed out on many occasions before, one of the main problems with the selection of adequate indicators is the lack of available data [8, 14]. Namely, it is often the case that individual statistical databases do not have complete data for all countries described in the survey, and in such a situation it is necessary to use the "nearest neighbor" method, which means that values are estimated on the most similar basis. Of course, this rule should be used as scarcely as possible.

4.3 Country Sample Selection

The new composite index model has been implemented in selected countries by The Organization for Economic Co-operation and Development. The following countries were selected for the survey: Sweden (SE), Norway (NO), Finland (FI), Germany (DE), China (CN), South Korea (KR), United States (USA), Italy (IT), France (FR), Poland (PL), Russia (RU), United Kingdom (UK), Spain (ES), Netherlands (NL), Japan (JP), Austria (AT) and Czech Republic (CZ).

The countries selected to apply the new competitiveness index are represented by 17 representative OECD members, established in 1961 with the aim of boosting the global economy and trade. Today, the OECD brings together 36 member countries, most of which are developed countries, recording achievements in all areas and showing high results according to numerous rankings and measurements of their competitiveness. These 36 countries are responsible for as much as, 42.8% of world GDP [58].

4.4 Creation of Thematic Subindices

The new composite index model presented consists of 50 indices that are classified into five thematic subindices under the following names: *Smart Society, Society of Good Chances, Networked Society, Knowledge Society and Sustainable Society* (Table 2). *Smart Society* subindicator measures the impact the latest industry 4.0 technologies have on today's smart society. They occupy the positions from 1 to 10 in the List of Indicators. *Society of Good Chances* subindicator refers to the economic and entrepreneurial conditions that companies face in doing business. In the list of indicators, they occupy the 11th to the 17th positions. *Networked Society* subindicator represents a measure of the extent of communication between people and companies. This subindicator assesses the basic conditions for establishing communication, as well as its frequency. In the list of indicators, they occupy the 18th to the 25th positions. *Knowledge Society* subindicator refers to the development of an effective innovation climate in companies, universities and other research institutions. These measures also describe the population situation in higher education, employment in the technology sector, as well as government and private sector allocations for R&D.

In the list of indicators, they occupy the 26th to the 42nd positions. *Sustainable Society* subindicator describes a measure of the environmental impact of society's development, as well as ways in which people can contribute to a greater degree of sustainable development. In the list of indicators, they occupy from the 43rd to the 50th position.

Table 2
List of indicators

No.	Name of subindicator	Name of indicator
1	Smart Society	IoT *(The Internet of Things)* devices online (per 100 inhabitants) [36, 37]
2		Artificial Intelligence Index [38]
3		Government Artificial Intelligence Readiness Index [39]
4		Autonomous Vehicles Readiness Index [40]
5		Electric vehicles charging stations (per million inhabitants) [41]
6		The Automation Readiness Index [42]
7		Use of cloud computing (% enterprises) [43]
8		3D Printing Country Index [44]
9		Estimated annual shipments of multipurpose industrial robots (per million inhabitants) [45]
10		Robots in manufacturing industry (per 10,000 employees) [45]
11	Society of Good Chances	GDP per capita (current US$) [46]
12		Ease of doing business index [46]
13		New business density (new registrations per 1,000 inhabitants ages 15–64) [46]
14		Time required to start a business (days) [46]
15		Foreign direct investment, net outflows (% of GDP) [46]
16		Foreign direct investment, net inflows (% of GDP) [46]
17		Logistics performance index [46]
18	Networked Society	Individuals using the Internet (% of population) [46]
19		Fixed broadband subscriptions (per 100 inhabitants) [46]
20		Number of active mobile–broadband subscriptions (per 100 inhabitants) [47]
21		Secure Internet servers (per million inhabitants) [46]
22		E–Participation Index [48]
23		Rates for broadband internet in PPP $/monthly [49]
24		Countries releasing most app (per million inhabitants) [50]
25		Country distribution of active online workers (by population share) [51]
26	Knowledge Society	Highly educated population (in % of people 30-34 years old) [52]
27		Government expenditure on education, total (% of GDP) [46]
28		School enrollment, tertiary (% gross) [46]

29		Graduates in science & engineering (% gross) [53]
30		Enrollment in tertiary education – PhD students – ISCED 8 (per million inhabitants) [53]
31		Gross expenditure on R&D (% GDP) [53]
32		Gross expenditure on R&D: Performed by business enterprise (% of GDP) [53]
33		Science, technology and innovation: total R&D personnel (per million inhabitants) [53]
34		Employment in technology and knowledge–intensive sectors (% workforce) [53]
35		Labor force with advanced education (% of total working–age population with advanced education) [46]
36		ICT goods imports (% total goods imports) [46]
37		ICT goods exports (% of total goods exports) [46]
38		High–technology exports (% of manufactured exports) [46]
39		Scientific and technical journal articles (per million inhabitants) [46]
40		Patent applications (per million inhabitants) [55]
41		Patent applications per GDP [55]
42		Patent grants [55]
43	Sustainable Society	CO2 emissions (metric tons per capita) [46]
44		Alternative and nuclear energy (% of total energy use) [46]
45		Electric power consumption (kWh per capita) [46]
46		Renewable internal freshwater resources per capita (cubic meters) [46]
47		Renewable energy consumption (% of total final energy consumption) [46]
48		PM2.5 air pollution, mean annual exposure (micrograms per cubic meter) [46]
49		Municipal waste total (kilograms per capita) [52]
50		Recycling rate of municipal waste (%) [56, 57]

Source: the authors

4.5 Indicator Standardization, Weights, Aggregation and Ranking of Countries by Subindicators

During data collection, we come across different units of measure that measure different indicators. In order to avoid mixing such values, it is necessary to normalize or standardize the data obtained. In practice, there are different techniques that can be applied, each with its own advantages and disadvantages and can produce different research results [14]. The paper presents a data standardization technique that gives an average of 100 for all variables.

$$Sij = \frac{Xij}{\bar{x}} * 100 \quad (2)$$

where s_{ij} is standardized value of the j-th indicator of indicator i-th of state; x_{ij} is value of the j-th indicator of indicator of the i-th state; $\bar{x}_j$ is average value of the j-th indicator.

The next step in obtaining relevant values is the assignment of weights. Weight values or weights are assigned to emphasize the importance of individual indicators and subindicators when constructing a composite index. There are several methods used for this purpose. These include regional analysis, principal component analysis, factor analysis, etc. [14]. In addition to the aforementioned methods, weight values can be assigned based on the analysis of experts in the analyzed areas, as well as on the quality and availability of the data obtained. It is important to point out that none of the above methods is completely reliable, and that different weighting techniques give different end results in measuring the competitiveness of countries. For this reason, some authors believe that this step should not be applied and that all factors should have the same weight value [31, 59]. However, the authors of this paper are of the view that weighting should be done for the reasons already mentioned. The weighting of individual indicators was carried out in accordance with their importance, in the opinion of the authors, within each of the five subindicators presented. In final step, the aggregation or summing up of values after standardization and weighting is performed, which results in the formation of results according to thematic indicators or subindicators.

4.6 Subindicator Weighting, Aggregation and Formation of the Competitiveness Index

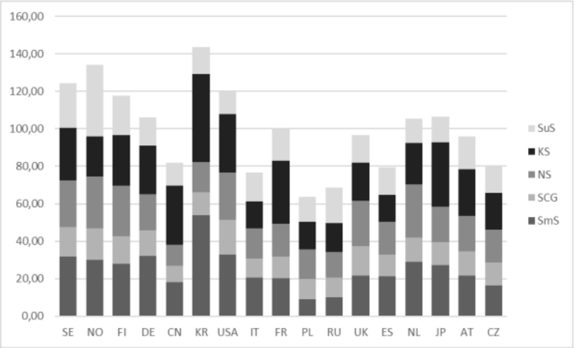
The largest weights are assigned to the *Smart Society* subindex (25%) and the *Knowledge Society* subindex (25%) because these two groups of parameters contain a large number of individual indicators but also have the greatest impact on the competitiveness of today's smart society in which we live. The subindexes *Networked Society* and *Sustainable Society* are assigned a weighting value of 17%, while the subindex *Society of Good Chances* has a weighting of 13%. As a reason for this method of assigning weights, we can state the author's estimated impact of the indicators themselves and the groups of indicators on society 4.0, as well as the number of individual indicators within the subindicators. Table 3 shows the values for each analyzed country according to the 5 subindicators. As can be seen from Table 3 and Figure 1, South Korea stands out as the leader according to the subindices *Smart Society* and *Knowledge Society*, while USA, stands out as having the best performance, within the subindex that describes the conditions for entrepreneurship, i.e. the *Society of Good Chances*. Within the *Networked Society*, Netherlands dominates but with almost the same result in the Norway and Finland. In the *Sustainable Society* subindex Norway is the undisputed leader.

The final result was obtained as the sum of 5 weighted subindicators under the names: *Smart Society* (SmS), *Society of Good Chances* (SGC), *Networked Society* (NS), *Knowledge Society* (KS) and *Sustainable Society* (SuS). Table 3 shows the weighted values of each subindex for each analyzed country, as well as the aggregation and formation of the total composite index result, or the final ranking of the countries considered in the survey.

Table 3
Composite index – assigning weight coefficients to subindices and aggregation and formation of the competitiveness index

Comp. subind.	SmS	SGC	NS	KS	SuS	Comp. index	Rank
Weight:	0.25	0.13	0.20	0.25	0.17		
SE	31.94	15.52	25.14	27.79	23.81	124.20	3
NO	29.92	16.83	27.67	21.66	37.90	133.98	2
FI	28.00	14.71	27.08	26.72	21.06	117.57	5
DE	32.21	13.79	18.95	26.20	14.95	106.09	7
CN	18.15	8.63	11.38	31.53	12.12	81.80	12
KR	53.98	12.03	16.29	46.97	14.16	143.43	1
USA	32.94	18.64	25.15	31.02	12.16	119.90	4
IT	20.66	10.08	16.21	14.27	15.30	76.52	15
FR	20.42	11.52	17.58	33.51	17.01	100.04	9
PL	9.04	10.92	15.70	14.76	13.46	63.87	17
RU	10.15	10.59	13.51	15.31	19.05	68.61	16
UK	21.81	15.79	24.06	20.14	14.65	96.46	10
ES	21.21	11.70	17.39	14.31	14.89	79.50	14
NL	29.10	12.81	28.46	22.17	12.99	105.54	8
JP	27.21	12.30	19.11	33.99	13.88	106.49	6
AT	21.83	12.76	18.98	25.00	17.54	96.10	11
CZ	16.43	12.37	17.35	19.64	14.09	79.88	13

Source: the authors



Source: the authors

Figure 1
Composite index and subindexes

The conducted research shows that South Korea is by far the first in the list according to the *Competitiveness Index of the Society 4.0* with a total of 143.43 index points. It especially stood out on indicators that measure the competitiveness of states in the smart and knowledge society. Norway is on the second place (133.98), being an unprecedented leader according to the *Sustainable Society* subindicator. Sweden (124.2), USA (119.9) and Finland (117.57) took the third, fourth and fifth place, showing exceptional performance in the *Smart Society* subindices. Japan (106.49), although the high-performing scorer in several fields, ranked 6th in the ranking of the composite index, primarily due to slightly worse results in the field of entrepreneurship and sustainable development. It is followed by several countries with small differences in points, namely: Germany (106.09) and Netherlands (105.54) which scores best according to the *Smart Society* subindex, France (100.04) which has shown exceptional performance in terms of *Knowledge Society*, United Kingdom (96.46) which stands out according to the subindicator *Networked Society* and Austria (96.10). They are followed by the China (81.8), the Czech Republic (79.88), Spain (79.5), Italy (76.52), while Russia (68.61) and Poland (63.87) are in below-average positions. Although today, China stands out as a leader in innovation and technological development, according to this composite index it did not occupy the highest positions, primarily due to the adjustment of indicators in proportion to the population. For example, if we look at the number of patent applications (residents), we can see that China is the global leader with 1,245,709 patent applications in 2017, which is more than half of the 2,161,610 patents reported in that year as a whole in the entire world. However, if adjusted according to population, countries such as South Korea (3119 patents per million population) and Japan (2041 pence per million population) have far better results than China (888 patents per million population) in 2017 [46].

4.7 Testing the Composite Indicator

Considering that the choice of statistical methods influences the end result, it is necessary to test the composite index. This step involves several tests, such as uncertainty and sensitivity tests, to understand the impact of certain variables, weights, and standardization techniques on the overall score or rank of the countries analyzed. In this way it is possible to evaluate the quality of the methods used and to improve it. Regression and variance methods were used as in [14].

4.7.1 Regression Analysis Based on the Indicators of Economic Dynamism

An economic dynamism indicator can show how, for example, GDP per capita affects the end result. To obtain such an indicator, we must first standardize the value of the composite index using the “minimum-maximum” method [14]. With this step, we get transformed values that range between zero (minimum value) and one (maximum value). That way, we will be able to get a picture of the state of a country's distance from its best and worst position. Otherwise, the standardization itself does not affect the ranking of countries according to certain indicators. The following step is to calculate the indicators of economic dynamism (*ECi*) using the following formula:

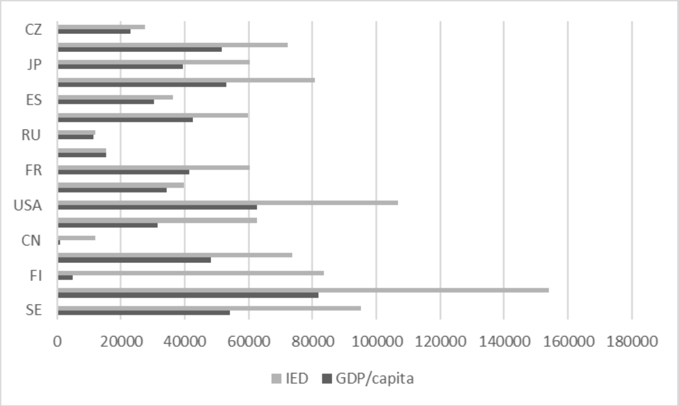
$$ECi = GDPi * (1 + yi) \tag{3}$$

where *yi* is a common composite index in relation to the difference between maximum and minimum, and *GDPi* is GDP per capita in USD thousands. The obtained results are shown in Table 4 and Figure 2, which shows a comparison between the country rankings obtained from calculations of *The Competitiveness Index of the Society 4.0* and the indicator of economic dynamism. Although, in most countries, similar results were obtained in the ranking, large oscillations are present in some countries, such as South Korea, which ranked on the 8th in this way compared to the 1st place, it was ranked as, according to *The Competitiveness Index of Society 4.0 (CIS 4.0)*. On the other hand, USA jumped from the 4th to the 2nd place, the Netherlands from the 8th to the 5th, Austria from the 11th to the 7th, and Italy from the 15th to the 12th.

Table 4
Composite *Competitiveness Index of the Society 4.0* and indicator of economic dynamism – ranking of countries

Country	***CIS 4.0***	**RANK type I**	***CIS 4.0* in relation to the difference between max. and mini.**	**GDP per capita in USD thousands**	**Indicator of Economic Dynamism**	**RANK type II**
SE	124.20	3	0.758	54,112	95,146	3
NO	133.98	2	0.881	81,807	153,903	1
FI	117.57	5	0.675	49,960	83,683	4
DE	106.09	7	0.531	48,195	73,772	6
CN	81.80	12	0.225	9,770	11,972	16
KR	143.43	1	1.000	31,362	62,724	8
USA	119.90	4	0.704	62,641	106,755	2
IT	76.52	15	0.159	34,318	39,773	12
FR	100.04	9	0.455	41,463	60,313	10
PL	63.87	17	0.000	15,424	15,424	15
RU	68.61	16	0.060	11,289	11,962	17
UK	96.46	10	0.410	42,491	59,894	11
ES	79.50	14	0.196	30,323	36,279	13
NL	105.54	8	0.524	52,978	80,725	5
JP	106.49	6	0.536	39,287	60,334	9
AT	96.10	11	0.405	51,512	72,380	7
CZ	79.88	13	0.201	22,973	27,597	14

Source: the authors



Source: the authors

Figure 2
Indicator of Economic Dynamism

4.7.2 Analysis of the Range of Variation and Variance

The analysis of the range of variation and variance (Table 5, Figure 3) shows the differences between the best and worst ranked countries by composite subindicators. Standard deviation is an indicator that shows us the average deviation from the average value, while the coefficient of variation shows the relationship between the standard deviation and the average value. From this analysis, we can see that some parameter groups have smaller ranges of variation, while others have larger ranges.

The subindices *Smart Society* (44.95) and *Networked Society* (32.70) have the biggest differences between the best and worst-ranked economies. In the *Smart Society* subindex, high variations between countries, such as highly ranked South Korea and Sweden versus lower ranked Poland and Russia, resulted mostly from Poland and Russia's lack of willingness to use robots in manufacturing industry. On the other hand, the analysis has shown that the *Good Chance Society* subindex (10.01) does not show large variations between the ranked countries, from which, we can conclude that all analyzed economies invest sufficient efforts in the development of entrepreneurship and good business climate.

Table 5
Analysis of the range of variation and variance

Comp. subindices	SmS	SGC	NS	KS	SuS
min	9.04	8.63	11.38	14.27	12.12
max	53.98	18.64	28.46	46.97	37.90
range of variation	44.95	10.01	17.09	32.70	25.78
variance	5726.66	7575.34	6425.63	5697.16	6925.67

standard deviation	312.01	358.86	330.51	311.21	343.13
average value	425	221	340	425	289
coefficient of variation	0.73	1.62	0.97	0.73	1.19

Source: the authors

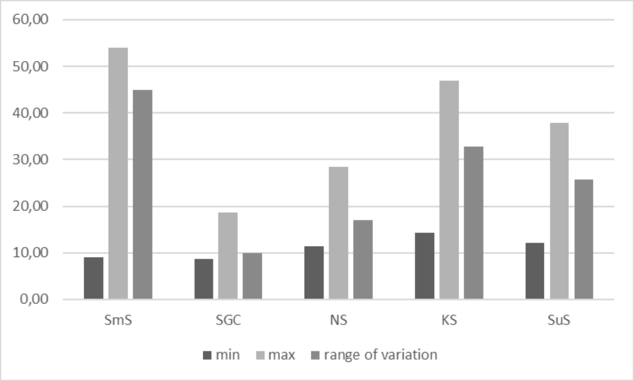


Figure 3

Range of variation

5 Dilemmas and Reflections

As the starting point of the research, the authors used the results of the previous work in which the classification of macroeconomic competitiveness indices was carried out, containing knowledge parameters in their model [7]. It can be seen that there are many models that are in some way related to the competitiveness of the knowledge society, but that they do not take into consideration the indicators of the new wave of change, i.e. the industry 4.0, and therefore are not suitable for measuring the national competitiveness of today's most developed countries, it was concluded in the previous works that qualitative indicators do not provide sufficiently objective results and that quantitative ones should be used as much as possible [7, 29]. In their further research, the authors developed a model for measuring the competitiveness of a knowledge-based society consisting of quantitative indicators. However, during development and advancement of technology, emerged the need to innovate the said model by incorporating new indicators reorienting to the demands of industry 4.0 [8]. In Table 1, an analysis of existing competitive models related to Industry 4.0 was implemented and among those three were found that measure competitiveness at the macro level. However, it has been found that they rely mainly on qualitative data and that there is a need to innovate competitiveness models in line with the requirements of Industry 4.0. From all of the above we can point out that we accept the main hypothesis of the paper which reads: The new model of competitiveness measurement presented in the paper provides a more adequate measurement of the position of today's most developed countries because it takes into consideration the challenges not only of the present but also of the future time, i.e. the fourth industrial revolution. The constructed index ranked 17 OECD countries using statistical methods of standardization, weighting and aggregation. The results showed that the Nordic countries like Norway, Sweden and Finland stood out, but also the success of the lands of the Asian Tigers i.e. South Korea, which ranked second.

Testing of the composite index was performed using the economic dynamism indicator. In this way it is shown how *The Competitiveness Index of the Society 4.0* depends on a parameter such as GDP per capita. In fact, the differences between lower-ranked countries like China and Russia and those at the top - Norway, USA, Sweden - have been found to be much greater when comparing economic indicators like GDP per capita than when ranking according to the parameters of a smart society. The differences in the ranking of some countries such as South Korea, which fell from the 1st to the 8th place upon crossing the parameters of economic dynamism, indicate the importance of using indicators that describe society in the context of industry 4.0. Namely, South Korea, although not at the top of the world in terms of economic performance, justifies its leadership by investing in the technologies and knowledge we need in the future, which is much more important today when measuring national competitiveness.

Furthermore, by conducting an analysis that determines the range in variations and variance for different composite index subindicators, it was found that there were differences between them. While, on one hand, in the subindicators *Sustainable Society* and the *Networked Society*, the ranges in variation and variance are pronounced, in the subindicators the *Society of Good Chances*, they much lower. Taking these data in consideration, as well as, the explanation given earlier in the paper, we can conclude that we accept the subhypothesis of the paper that reads: There are subindexes that have a small range of variation and variance, as well as those with more pronounced differences between the worst and the best ranked economy.

Conclusions

The study of the concept of Competitiveness, is a subject of increasing interest of Authors around the world, as the micro- and macro-environment becomes more complex and the number of factors, with multiple influences increase. That is why measuring and analyzing competitiveness today is vital for the creation of national and regional strategies and plans, as it is thus possible, to obtain guidance for tuning future development.

To more accurately measure the competitiveness of countries in the future, the authors believe that it is necessary to innovate existing models and introduce parameters such as: IoT devices online, Artificial Intelligence Index, use of industrial robots, 3D Printing Country Index, etc. This choice of indicators is particularly suitable for measuring the competitiveness of highly developed countries such as OECD members.

Although gross domestic product (GDP) per capita, was previously used as a major benchmark for success of a country, today, such an indicator can produce misinformation, due to the complexity of today's society in which we live. This paper establishes that there are several models that relate to the examination of competitiveness in the context of the Fourth Industrial Revolution at the macro level, but also that they mainly rely on qualitative data, which diminishes their objectivity. For this reason, the modeling of the composite *Competitiveness Index of the Society 4.0* was performed, which was the aim of the paper.

When it comes to the limitations of the research, it is also worth pointing out that the current availability of data regarding the parameters describing the characteristics of the Fourth Industrial Revolution is quite low, and that in this respect there is a limitation in the use of the proposed model in terms of expansion of the country sample. However, the Authors believe that this limitation is only current and that the model will be able to be applied very quickly over a much wider range. Regarding directions for further research, it can be stated that in the coming period it is very important to constantly monitor the development of science and technology and gradually introduce new indicators that reflect this development. In the future, this could be, for example, an indicator measuring the number of autonomous vehicles or domestic robots. It is also necessary to constantly re-examine the role of certain weight values in the overall score and to change it according to the need. Also, research can be conducted at the regional level, of course, if relevant data are available.

The results provided by this model can be used in many applications, and above all, it would benefit all those who wish to quantify information and compare countries in terms of competitiveness in the coming 4th Industrial Revolution. Also, the results may indicate that there are some negative trends in individual countries, which can be a good signal not to proceed with such trends. Although the survey is faced with limited availability of parameters, primarily with regard to the *Smart Society* subindex, the Authors believe that data availability will improve and the survey can be extended to a much larger range of countries. The authors believe that this paper significantly contributes to the work of all researchers engaged in the study of competitiveness, at the national level and even more so, in a broader sense, i.e. for those who can use and make use of these results in further research.

References

[1] Schwab, K.: The Fourth Industrial Revolution, Currency, 2017

[2] World Economic Forum: The Future of Jobs: Employment, Skills and Workforce Strategy for the Fourth Industrial Revolution, In Global Challenge Insight Report, World Economic Forum, Geneva, 2016

[3] Pereira, A. C., Romero, F.: A Review of the Meanings and the Implications of the Industry 4.0 Concept, Procedia Manufacturing, Vol. 13, 2017, pp. 1206-1214

[4] Vaidya, S., Ambad, P., Bhosle, S.: Industry 4.0 – A Glimpse, Procedia Manufacturing, Vol. 20, 2018, pp. 233-238

[5] OECD: OECD Skills Outlook 2019: Thriving in a Digital World, OECD Publishing, Paris, 2019

[6] OECD: OECD Employment Outlook 2019: The Future of Work, OECD Publishing, Paris, 2019

[7] Katić, A., Ćosić, I., Anđelić, G., Raletić, S.: Review of Competitiveness Indices that Use Knowledge as a Criterion, Acta Polytechnica Hungarica Vol. 9, No. 5, 2012, pp. 25-45

[8] Katić, A., Kiš T., Ćosić I., Vukadinović S., Dobrodolac Šeregelj T.: Modelling the Composite Competitiveness Index of the Knowledge-Based Society, Acta Polytechnica Hungarica, Vol. 12, No. 1, 2015, pp. 229-249

[9] Katić, A., Ćosić, I., Kupusinac, A., Vasiljević, M., Stojić, I.: Knowledge-Based Competitiveness Indices and its Connection with Energy Indices, Thermal Science, Vol. 20, 2016, pp. 451-461

[10] Porter, M. E., Heppelmann, J. E.: How Smart, Connected Products are Transforming Competition, Harvard Business Review, 2014, Vol. 92, No. 11, pp. 64-88

[11] Porter, M. E., Heppelmann, J. E.: How Smart, Connected Products are Transforming Companies, Harvard Business Review. 2015, Vol. 93, No. 10, pp. 96-114

[12] Drucker, P.: The Age of Discontinuity: Guidelines to our Changing Society, Routledge, 2017

[13] OECD: The Knowledge-Based Economy, OECD Publishing, Paris, 1996

[14] Nijkamp, P., Siedschlag, I.: Innovation, Growth and Competitiveness. Dynamic Regions in the Knowledge-Based World Economy, Springer-Verlag Berlin Heidelberg, 2011

[15] Porter, M., Stern, S.: The New Challenge to America's Prosperity: Findings from the Innovation Index. Council of Competitiveness, Washington DC, 1999

[16] Freudenberg, M.: Composite Indicators of Country Performance: a Critical Assessment, OECD Science, Technology and Industry Working Papers, 2003/26, OECD Publishing, Paris, 2003

[17] Sendler, U.: Industry 4.0 - Mastering Industrial Complexity with SysLM (in German), Springer, 2013, pp. 456

[18] Kagermann, H., Wolfgang W., Johannes H.: Securing the Future of German Manufacturing Industry, Recommendations for Implementing the Strategic Initiative Industrie 4, No. 199, 2013

[19] Morrar, R., Arman, H., Mousa, S.: The Fourth Industrial Revolution (Industry 4.0): A Social Innovation Perspective, Technology Innovation Management Review, Vol. 7, No. 11, 2017, pp. 12-20

[20] Dutton, H. W.: Putting Things to Work: Social and Policy Challenges for the Internet of Things, Info, Vol. 16, No. 3, 2014, pp. 1-21

[21] Office for National Statistics: What are the Chances of Surviving to Age 100?, https://www.ons.gov.uk, 2016, Accessed on October 2019

[22] Okanovic, A.: Management of Competitiveness, Faculty of Technical Sciences, Novi Sad, Serbia, 2018

[23] Vacek, J.: On The Road: From Industry 4.0 to Society 4.0., Trendy v Podnikání, Vol. 7, No. 4, 2017, pp. 43-49

[24] Dujin, A., Geissler, C., Horstkötter, D.: Industry 4.0 The New Industrial Revolution - How Europe Will Succeed. Rol. Berger Strateg. Consult, 2014, pp. 1-24

[25] Basl, J., Doucek, P.: A Metamodel for Evaluating Enterprise Readiness in the Context of Industry 4.0., Information, 2019, Vol. 10, No. 89, doi:10.3390/info10030089

[26] Economic Development Board: The Singapore Smart Industry Readiness Index, Catalyzing the Transformation of Manufacturing, Economic Development Board, 2017

[27] Kearney, A. T.: Readiness for the Future of Production Report, The World Economic Forum, 2018

[28] Faarup J., Faarup A.: Global Industry 4.0 Readiness Report 2016, Danish Institute of Industry 4.0 (DII 4.0), 2017

[29] Katić, A., Ćosić, I., Anđelić, G.: Knowledge Based Competitiveness Indices and Position of Serbia, PSU-UNS International Conference on Engineering and Technology – ICET, Novi Sad, 2013

[30] Batchkova, A., Popov T., Ivanova A., Belev A.: Assessment of Readiness for Industry 4.0, International Scientific Journal "Industry 4.0", Vol. 3, No. 6, 2018, pp. 288-291

[31] Booysen, F.: An Overview and Evaluation of Composite Indices of Development, Social Indicators Research, Vol. 59, No. 2, 2002, pp. 115-151

[32] Wang, Y., Ma, H. S., Yang, J. H., Wang, K. S.: Industry 4.0: A Way from Mass Customization to Mass Personalization Production, Advances in Manufacturing, Vol. 5, No. 4, 2017, pp. 311-320

[33] Liu, C.: International Competitiveness and the 4. Industrial Revolution, Entrepreneurial Business and Economics Review, Vol. 5, No. 4, 2017, pp. 111-133

[34] Xu, M., David, J. M., Kim, S. H.: The Fourth Industrial Revolution: Opportunities and Challenges, International Journal of Financial Research, Vol. 9, No. 2, 2018, pp. 90-95

[35] Guoping L., Yun H., Aizhi W.: Fourth Industrial Revolution: Technological Drivers, Impacts and Coping Methods, Chinese Geographical Science, Vol. 27, No. 4, 2017, pp. 626-637

[36] Peña-López, I.: OECD Digital Economy Outlook 2015, OECD Publishing, Paris, 2015

[37] OECD: Key emerging technologies, OECD Publishing, Paris, 2017

[38] Shoham, Y., Perrault, R., Brynjolfsson, E., Clark, J., LeGassick, C.: Artificial Intelligence Index 2017 Annual Report, Artificial Intelligence Index, 2017

[39] Miller, H., Stirling, R., Chung, Y., Lokanathan, S., Martinho-Truswell, E., New, J., Rutenberg, I., Scrollini, F.: Government Artificial Intelligence Readiness Index 2019, Oxford Insights, 2019

[40] Threlfall, R.: Autonomous Vehicles Readiness Index, Klynveld Peat Marwick Goerdeler (KPMG) International, 2019

[41] Open Charge Map: The Global Public Registry of Electric Vehicle Charging Locations, https://openchargemap.org/site/, Accessed on September 2019

[42] Unit EI. The Automation Readiness Index: Who is Ready for the Coming Wave of Automation?. London: Economist Intelligence Unit. 2018

[43] OECD: ICT Access and Usage by Businesses (database), http://oe.cd/bus, Accessed on September 2019

[44] HP: 3D Printing: Ensuring Manufacturing Leadership in the 21st Century, HP, 2018

[45] International Federation of Robotics: Robot Density by Country 2016, https://ifr.org/, Accessed on September 2019

[46] The World Bank Group: https://data.worldbank.org/, Accessed on September 2019

[47] OECD: Telecommunications and Internet Statistics: Broadband database, https://data.oecd.org/, Accessed on September 2019

[48] Zhenmin L.: United Nations E-Government Surveys: Gearing E-Government to Support Transformation Towards Sustainable and Resilient Societies, United Nations, New York, 2018

[49] Zavazava, C.: ICT Prices 2017, International Telecommunication Union, Geneva Switzerland, 2017

[50] Appfigures: Indicator: Countries releasing most app, https://appfigures.com/explorer/datasets, Accessed on September 2019

[51] Kässi, O., Lehdonvirta, V.: Online Labour Index: Measuring the Online Gig Economy for Policy and Research, Technological Forecasting and Social Change, Vol. 137, 2018, pp. 241-248

[52] OECD: https://data.oecd.org/, Accessed on September 2019

[53] UNESCO Institute for Statistics: http://data.uis.unesco.org/, Accessed on September 2019

[54] Eurostat: Indicator: Employment in technology and knowledge-intensive sectors at the national level, http://appsso.eurostat.ec.europa.eu/, Accessed on September 2019

[55] World Intellectual Property Organization: World Intellectual Property Indicators 2018, World Intellectual Property Organization, 2018

[56] Eurostat: Indicator: Recycling rate of municipal waste in %, https://ec.europa.eu/, Accessed on September 2019

[57] Pariona, A.: OECD Recycling Statistics, World Atlas, https://www.worldatlas.com/, Accessed on September 2019

[58] OECD: List of OECD Member countries, https://www.oecd.org/, Accessed on September 2019

[59] Porter, M., Stern, S.: The New Challenge to America's Prosperity: Findings from the Innovation Index, Council of Competitiveness, Washington DC, 1999

Quality Evaluation of Audio and Video Signals in Videoconferences

Jana Filanová, Iveta Ondrášová, Anikó Töröková

University of Economics in Bratislava
Dolnozemská cesta 1, 852 35 Bratislava, Slovak Republic
jana.filanova@euba.sk, iveta.ondrasova@euba.sk, aniko.torokova@euba.sk

Abstract: Videoconferencing represents a technology of the future, in modern education. A combination of audio and video information serves in understanding the content of lectures or presentations, in the form of videoconferencing. The evaluation of the quality of videoconferencing is difficult, as the image and sound affects the final quality. In general, occasional image disturbance has less impact on the perception of quality in comparison to the disturbances in an audio track. In this research, we simulated a real packet network environment and tested video sequences that present different teaching content. We artificially degraded the quality of video sequences by packet loss and jitter. Our test aimed to compare subjective methods of video quality evaluation with objective methods and to evaluate the impact of audio quality on the overall video sequence quality. This paper describes a novel process of evaluating the quality of audio and video signals. Time-consuming subjective measurements were supported by models and programs that simplified the preparation, testing, and processing of results. The contribution of this article is to present and evaluate the results of video sequence quality testing with an emphasis on semantics, which has a significant impact on viewers' sensitivity to video sequence quality.

Keywords: videoconferencing; virtual reality; quality evaluation of video and audio; packet loss; latency; objective assessment; subjective assessment; MOS scores; semantics

1 Introduction

Videoconferences represent a form of synchronous communication based on audio and video transmission with the possibility to integrate text and other forms of presentation of information at a distance. The quality of this communication is influenced by the used communication technologies and transmission characteristics of communication networks [1]. Videoconferencing is one of the most appropriate ways of online transmission information to participants. The videoconferences could be recorded and it is possible to view the records even in the off-line mode [2].

At the primary and secondary education levels, videoconferencing can be used for teaching pupils without access to regular education. This might be due to the students' physical isolation (e.g., students living in remote areas, disabled students or students quarantining at home) or for various economic and social reasons. In addition, videoconferencing can be applied to the teaching of gifted students who can benefit from more intense learning or choice of subjects not available at their school. Teaching and learning are complex activities realized by many methods [3].

Universities, High Schools, and various Higher Education Institutions are trying to meet the needs of growing numbers of external students, whose, other commitments, do not allow them to attend regular lectures and exercises. The modern trend in education is virtual reality. It represents a modern form of education, which brings education content from the classical education room to an online environment. Students and teachers have then access to education and information from anywhere [4].

The visual perception of people is a highly complex matter that involves several mechanisms. It is influenced by their expectations and their previous experience. The view of the quality is linked to their mechanisms of imagination. The quality of the presentation through videoconferencing will depend not only on the technical quality of the videoconferencing but also on other factors such as lecture content [5]. In [6] the authors show that semantics has a significant impact on viewers' sensitivity to the quality of a video sequence for spatially separated parts of the sequence and, more importantly, that this difference in sensitivity can be changed by the presence of an audio signal. This result is important for any testing of subjects' responses to visual material. One example is the subjective assessment of the quality of video in an audio-visual communications system (such as television or videoconferencing) [6].

Videoconferencing quality testing is very specific. In the real world, we usually perceive information simultaneously from two or more sources and then process them into the resulting form. A good example is the reading from lips where, besides the speaker's voice, we also observe the movement of his/her lips. From the perspective of subjective evaluation of videoconferencing quality, it is true that some parts captured by the camera are more important than others. Such areas are known as "Foregrounds". For example, during a videoconferencing, the most important areas are the head and shoulders of the person being captured, while the rest in the background is not important [7].

The human eye is the most important organ in sensory perception. Human beings acquires about 80% of the world's information using their eyes. But one must realize that the eye does not give the brain a definite picture of the outside world. The image of the outside world consists of a combination of information from the eye and the observer's experience [8]. The transition from the stimulus in the eye to the central nervous system analysis is not immediate but has a delay of approximately 20 ms (on average and differs from person to person). This means that patterns changing at a rate greater than 50 Hz are perceived as continuous movements [9]. For instance, the television works on the same principle.

The sound is defined as every longitudinal mechanical oscillation in a medium that is capable of creating a hearing perception in the human ear. The sensitivity threshold_of the auditory organ in a healthy human is about I0 = 10-12 Wm-2 at a frequency of 1000 Hz. This amount is referred to as the zero volume level or the conventional listening threshold (0 dB) at a frequency of 1000 Hz [10]. At this threshold sound intensity, the amplitudes of the movement of the eardrum are of the order of the atom diameter. The basilar membrane oscillations show approximately the same amplitudes. According to current knowledge, it is difficult to explain the mechanism by which these slight deflections can cause irritation of the nerve endings [11].

The results of the research [12] have shown that the presence or absence of audio has a significant impact on the overall subjective perception of the videoconferencing quality. It has also been found that the viewer is more sensitive to the quality of the image in the foreground of the speaking person than to the quality of the image in the background. If there are multiple people in the scene, even not speaking right now, the viewer is likewise more sensitive to the quality of the image of the captured people than to the quality of the image in the background [12].

Digital image data stored in image databases and distributed over communication networks are subject to various types of distortions during data acquisition, compression, processing, transmission, and reproduction. e.g., lossy video compression methods that are almost always used to reduce the bandwidth needed to store or transmit video data may degrade video quality during the quantization process. In fact, digital video streams transmitted over error-prone channels (e.g., wireless channels) may be received as incomplete due to the deterioration encountered during the transmission. Packet communication channels (Internet) can cause loss or delay of received packets, depending on network status and QoS (Quality of Service) used [2]. The effects of time delay can be reduced with various control methods designed for latency-tolerance [13]. Transmission errors can result in a deterioration of the received image information. Therefore, it is desired that systems designed for video services are able to realize and quantify the degradation of video quality that occurs in the system. This is especially important in order to maintain, manage, and at best, improve the quality of image data. Effective metrics of quality of static image and video are essential for this purpose [14].

Image quality assessment is a challenging task that is traditionally approached by computational models. To maintain, control, and enhance the quality of images, it is important for image acquisition, management, communication, and processing systems to be able to identify and quantify image quality degradations. A great deal of effort has been made in recent years to develop objective image quality metrics that correlate with perceived quality measurement [15, 16].

The aim of developing new methods for evaluating video quality objectively is to design metrics that can independently predict video quality [15]. Objective video metrics can be used to monitor image quality in quality management systems. When using objective video metrics, a network video server can monitor the quality of video transmitted by the network and manage video streaming. Objective video quality measures play important roles in various video processing applications, such as compression, communication, printing, analysis, registration, restoration, and enhancement. Experiments on the video quality experts group (VQEG) test dataset show that the new quality measure has a higher correlation with subjective quality measurement than the proposed methods in VQEG's Phase I tests for full-reference video quality assessment [17].

The most reliable way to measure video quality is subjective assessment because in most cases, a human being is the ultimate recipient of the video.

However, one of the major issues is that subjective methods are inconvenient, slow, and costly for practical use.

This article presents the process of quality evaluation of video sequences, which gives practical instructions to facilitate and accelerate subjective evaluation. Section 2 explains the methodology of our research. It describes subjective and objective methods for evaluating video and audio and finally a process model used in our research. Section 3 presents the results of the research including a comparison of the video sequences quality evaluation results. The contribution of the article is to present and evaluate the results of video sequence quality testing with an emphasis on semantics, which has a significant impact on viewers' sensitivity to video sequence quality.

2 Research Methodology

In this research, we simulated an environment of a real packet network and tested video sequences that would simulate the diverse content of teaching. We artificially degraded the quality of video sequences by packet loss and jitter. The objective of the test was to compare subjective methods with objective methods and evaluate the impact of the quality of the audio on the overall quality of the video sequence. We also wanted to show that semantics has a significant impact on viewers' sensitivity to the quality of the video sequence [6].

Subjective quality cannot be represented by an exact figure. Due to its inherent subjectivity, it can only be described statistically. Even in psychophysical threshold experiments, where the task of the observer is just to give a yes/no answer, there is a significant variation in contrast sensitivity functions and other critical low-level visual parameters between 50 different video quality observers. When the artifacts become supra-threshold, the observers are bound to apply different weightings to each of them [18].

International recommendations for subjective methods of quality testing include specifications on how to implement different types of subjective tests. Some of these test methods are known as "double stimulus" methods where an observer evaluates quality or quality change between two (reference and test) video sequences. There are also "single stimulus" methods where the observer evaluates the quality of just one (test) video sequence [19, 20, 21]. The following subsections 2.1 to 2.3 describe three subjective methods: two “double stimulus” methods DSCQS and DSIS and one “single stimulus” ACR method. Subsections 2.4 and 2.5 introduce the metrics MSE and PSNR and SSIM index used in objective evaluation methods. Finally, subsection 2.6 presents the structural process model we created for evaluating video sequences.

2.1 DSCQS Method

The Double Stimulus Continuous Quality Scale (DSCQS) method is suitable for measuring the quality of the system that is related to the reference value as the observer is not familiar with the reference sequence order [19]. DSCQS is quite sensitive to small differences in quality and is thus the preferred method when the quality of the test sequence and reference sequence are similar [18].

2.2 DSIS Method

The Double Stimulus Impairment Scale (DSIS) method is suitable for assessing the extent of degradation of the test sequence as compared to the reference one, especially in case of visible/significant degradation. For example, it is used to evaluate the degradation of the sequence during transport. This method is faster than DSCQS since the sequences are displayed only once [19]. Subjects rate the amount of impairment in the test sequence on a discrete five-level scale ranging from “very annoying” to “imperceptible”. The DSIS method is well suited for evaluating clearly visible impairments such as artifacts caused by transmission errors [18].

2.3 ACR Method

The Absolute Category Rating (ACR) method is a single stimulus method; viewers only see the video under test, without the reference. They give one rating for its overall quality using a discrete five-level scale from “bad” to “excellent”. The fact that the reference is not shown with every test clip makes ACR a very efficient method compared to DSIS or DSCQS, which take almost 2 to 4 times longer, respectively [18, 20].

2.4 MSE and PSNR

The best-known methods for objective evaluation of signal quality include metrics based on pixel comparisons, such as MSE (Mean Squared Error) and PSNR (Peak Signal to Noise Ratio). An advantage of these methods is the speed and ease of calculation. A disadvantage is that they do not accurately capture the perception of quality and distortion by the human visual system [22].

The MSE is the mean of the squared differences between the gray-level values of pixels in two pictures or sequences *I* and *I'*:

$$\mathrm{MSE} = \frac{1}{\mathrm{TXY}} \sum_t \sum_x \sum_y \left[\mathrm{I(t,x,y)} \text{-} \tilde{\mathrm{I}}\mathrm{(t,x,y)} \right]^2 \tag{1}$$

for pictures of size *X* x *Y* pixels and *T* frames in the sequence [22].

The PSNR in decibels is defined as:

$$\mathrm{PSNR} = \ 10 \log \frac{\mathrm{m}^2}{\mathrm{MSE}} \tag{2}$$

where *m* is the maximum value that a pixel can take [22].

2.5 SSIM Index

Newer methods for objective evaluation of the signal quality include the SSIM Index (Structural Similarity Index). The SSIM metric measures three components: the luminance similarity, the contrast similarity, and the structural similarity and combines them into one final value that determines the quality of the test sequence (Figure 1). This method differs from the above-described error-based methods described by using the structural distortion measurement instead of the error one [23]. It is due to the human visual system that is highly specialized in extracting structural information from the viewing field and it is not specialized in extracting the errors. Owing to this factor, the SSIM metric achieves a good correlation to subjective impression [24, 25]. The results are in the interval [0,1], where 0 and 1 denote the worst and the best quality, respectively.

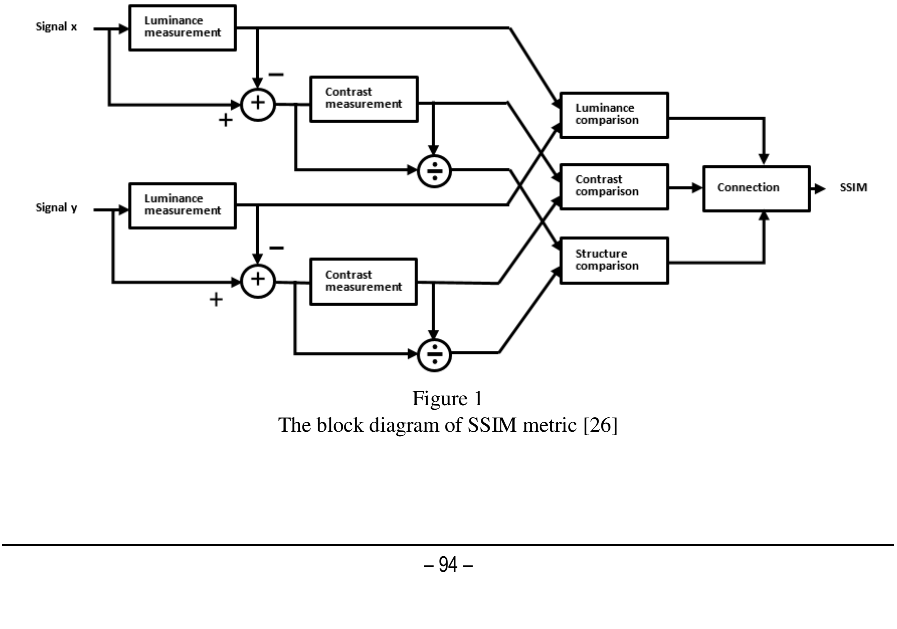


Figure 1
The block diagram of SSIM metric [26]

2.6 Process of Subjective Evaluation of Video Sequences

There is no single and ideal method to measure video quality. It is very important to choose the right method to meet our needs. Subjective methods provide more reliable results but objective methods are not influenced by the viewer's opinions or experiences [27]. Subjective video quality testing is difficult not only because of the time-consuming nature of testing itself but also due to the complexity of the steps that precede the actual testing. Figure 2 describes five steps of the process model we have designed for subjective evaluation of the quality of video sequences. It is based on the process model that we presented in the article [21].

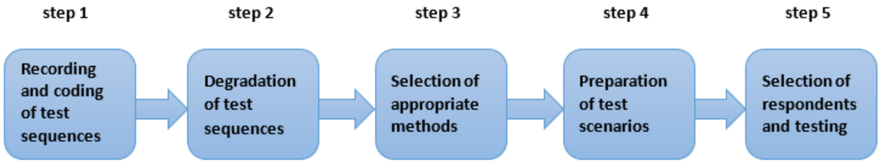

step 1
step 2
step 3
step 4
step 5
Recording and coding of test sequences
Degradation of test sequences
Selection of appropriate methods
Preparation of test scenarios
Selection of respondents and testing

Figure 2
The structural process model of video sequences quality evaluation [21]

2.6.1 Recording and Coding of Test Sequences

Reference video sequences were created based on real video calls. These sequences were recorded using the Logitech C270 web camera with HD resolution of 1270 x 720 pixels, utilizing the Logitech Webcam Software shipped with the web camera. Due to the purpose of the testing, it was important to create diverse demonstrations with a different emphasis on content, the importance of video or audio capture. Four types of reference video sequences are described below.

In the first test sequence (video sequence No. 1), the intention was to create a preview where the emphasis would be on the picture detail. The lecturer in this video preview informs students that if they have any questions, they can contact him at his e-mail address. The person in the preview does not pronounce this e-mail address but writes it on the board (Figure 3). So the only way this e-mail address information gets to the user of the videoconference is assuring that the image quality will be sufficient, to recognize it without difficulty.

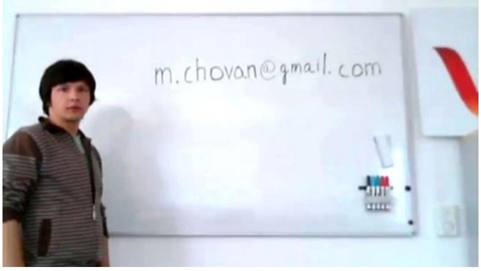


Figure 3
Photo from the test video sequence No. 1

In the second and third test video sequences, the aim was to create a demonstration where an emphasis would be placed on the quality of the audio during static image transfer. In the second example (video sequence No. 2), a woman asks the recipient to contact someone by phone. She dictates her name and phone number. In the third test sequence (video sequence No. 3), the student asks a classmate to provide him with the lecture notes he missed. He uses several shortcuts, so passing the information takes a short time. Unlike in the first demonstration, in the video sequences two and three, the information is provided only in the form of sound. Therefore, to interpret it correctly, the audio must be captured completely and correctly.

In the fourth test sequence (video sequence No. 4), the teacher explains the formula for calculating electrical efficiency. The formula is written on the board, while the teacher simultaneously talks about individual variables in the formula. Since the information is provided through both image and sound at the same time, minor audio outages can be compensated for by the visual clarity of information or vice versa minor video outages can be compensated for by the audio clarity.

Each video sequence was encoded, because the video and audio formats used, as well as bit rates, do not match those used in videoconferencing. Recording and coding technical parameters of reference video sequences are described in Tab. 1.

Table 1

Recording and coding technical parameters

parameter	recording	coding
pixel	1270 x 720	1270 x 720
frame rate	15	30
sequence length [sec]	10	10
video format	WMV2	MPEG-4 AVC
audio format	WMA	AAC
bit rate of video [kbit/sec]	3535	1024
bit rate of audio [kbit/sec]	1411	128
audio sampling rate [kHz]	48	22.05

2.6.2 Degradation of Test Sequences

An important part of the research was the selection of appropriate subjective methods for evaluation of the quality of video sequences. As we wanted to use “double stimulus” methods in testing, we had to create degraded samples in addition to reference samples. To introduce degradations into the reference videoconferencing sequences, it was necessary to emulate the transfer environment through which the sequences were transmitted (Figure 4).

Network emulation is a process by which we can control and repeatedly simulate network performance. The changes in network parameters such as latency and packet loss are provided by traffic shapers. They must be controlled according to predefined specifications to simulate the required features of the network.

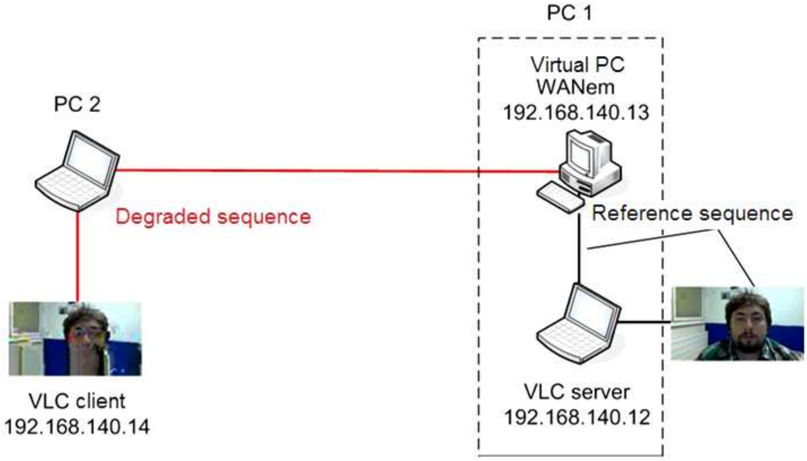


Figure 4
Network model for creation of degraded sequences [21]

PC 1 served as a video streaming server (Figure 4). We had to set the destination IP address, data transfer protocol (UDP), port, and modify the routing table to route all outgoing packets to the virtual PC. On the PC 2 side, VLC media player 0.8.6f was used as a client to receive the streamed video and also allowed to save it. Similarly, to the server, it was necessary to set the destination IP address (PC 2 IP address), data transfer protocol (UDP), port and address where the received video should be stored [21, 23].

Another program we used was WAN from TATA Consultancy Services (Figure 4). It is an open-source program used to emulate WAN networks (e.g. Internet) in a LAN environment. It allows setting many parameters such as bandwidth for transmission, latency, jitter and packet loss [28].

Each of the four reference samples was degraded by packet loss (0.5%, 1%, 3%, 5%, and 10%) and jitter (50 ms jitter at 100 ms latency).

2.6.3 Selection of Appropriate Methods

Absolute Category Rating (ACR) and Double Stimulus Impairment Scale (DSIS) methods were selected for the subjective evaluation of video samples. The ACR method has the advantage of being fast as the evaluator watches the sample only once and the length of the sample is relatively short (about 10 seconds). The DSIS method was also selected because of its time efficiency and the ability to capture more accurate differences between degraded samples, as we also have a reference sample for this method [20, 21]. The choice of suitable methods was also influenced by the fact that both the ACR‘s and DSIS‘s outputs are MOS scores with values ranging from 1 to 5, so the results can easily be compared [14].

To objectively evaluate the quality of video sequences we used the MSU Video Quality Measurement Tool. From the portfolio of available methods, we chose the PSNR method, whose advantage is the speed of calculation [29]. The second objective method we used was the SSIM method that already includes models of the human visual system, and therefore, the results should better correspond to the outcomes of subjective evaluation [26]. Both methods required a comparison of the degraded video sequence with the reference sequence.

2.6.4 Preparation of Test Scenarios and Selection of Respondents

Since the testing was performed within the VLC multimedia player environment, it was necessary to create playlists in which the individual video sequences were arranged appropriately. To prepare the scenarios and the course of the subjective measurements, a program was created in the C# programming language. To play a video sequence the program uses an open-source DmediaPalyer that is a modification of the VLC player. The program consists of two parts: test manager part and tester part (Figure 5). The Test manager part is an interface used to create structure of the test. You can choose the type of subjective method, test sequence, reference sequence (if necessary) and enable or disable sound step-by-step. We presented this program in the article [27].

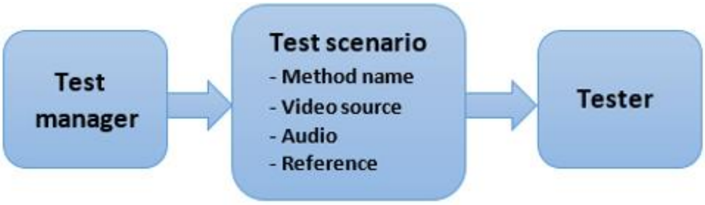


Figure 5
The block diagram of testing and test scenarios preparation program [27]

The ITU-T Recommendations specify that the number of respondents for subjective quality assessment must be greater than 4 and less than 40 [19, 20]. Based on this, we selected 20 respondents (10 women and 10 men), aged 20-51.

The fifth step of the subjective quality evaluation includes testing. The course of testing, evaluation, and comparison of the results are described in the following section.

3 Comparison of Video Sequences Quality Measurement Results

Our research aimed to compare subjective and objective methods of video sequence testing and to determine the degree of impact of audio quality on overall video quality with respect to the semantics.

Due to the time-consuming manual processing of results, two programs were created.

The first program was written in the C# programming language. There are two list data structures, one for each sequence. These data structures store individual objects whose variables have values read from individual result files. In the case of the DSIS method, the variables are the method name, respondent name, age, gender, reference and ranked sequence name, and the evaluation itself. In the case of the ACR method, the variables are the method name, respondent name, age, sex, names of the first and second sequence to be evaluated, and their evaluation itself. The program processes each file sequentially. After reading all the data, it checks whether the list contains an object with the same values of the variables. If there is no such object, the object with the loaded variables is saved. Otherwise, the object is deleted and a message about its deletion is written to the console. The algorithm then sequentially scans individual objects and writes them to the output file according to the given criteria. It also allows the results to be processed with respect to their statistical processing (performed by the second program described below). If the respondent was excluded from the DSIS method, they are also excluded from the ACR method.

The program has two outputs in the form of text files. In the first one, the results are processed according to the evaluation of the individual sequences. In the case of the DSIS method, the format is the reference sequence name and the test sequence name followed by five numbers. In the case of the ACR method, the format is the test sequence name and five numbers. The five numbers correspond to the evaluation scale of the given methods [20, 21]. If it has been chosen to take the statistical processing into account, the output is in the same file. In the second text file, the results are processed according to respondents who evaluated individual sequences. This output is needed for statistical processing of results for the DSIS method.

The second program is used for statistical processing of measured results. It was created in Matlab version R2008b. The algorithm for statistical processing of measured results was designed as follows:

The average score $\bar{u}_{jkr}$ is calculated for each test sequence

$$\bar{u}_{jkr} = \frac{1}{N}\sum_{i=1}^{N} u_{ijkr} \tag{3}$$

where u_{ijkr} is the respondent score i for test condition j, sequence k and number of repetitions r. N is the total number of respondents.

The standard deviation S_{jkr} and the peak coefficient β_{2jkr} are also calculated for each test sequence:

$$S_{jkr} = \sqrt{\sum_{i=1}^{N} \frac{(\bar{u}_{jkr} - u_{ijkr})^2}{(N-1)}} \tag{4}$$

$$\beta_{2jkr} = \frac{m_4}{(m_2)^2} \tag{5}$$

where $$m_x = \frac{\sum_{i=1}^{N} (u_{ijkr} - \bar{u}_{ijkr})^x}{N} \tag{6}$$

Then we find Q_i and P_i for each respondent i as follows:

If $2 \leq \beta_{2jkr} \leq 4$ then

if $u_{ijkr} \geq \bar{u}_{jkr} + 2\, S_{jkr}$ then $P_i = P_i + 1$ (7)

if $u_{ijkr} \leq \bar{u}_{jkr} - 2\, S_{jkr}$ then $Q_i = Q_i + 1$ (8)

If $\beta_{2jkr} < 2$ *or* $\beta_{2jkr} > 4$ then

if $u_{ijkr} \geq \bar{u}_{jkr} + \sqrt{20}\, S_{jkr}$ then $P_i = P_i + 1$ (9)

if $u_{ijkr} \leq \bar{u}_{jkr} - \sqrt{20}\, S_{jkr}$ then $Q_i = Q_i + 1$ (10)

The assessment of respondent i will not be taken into account if conditions (11) and (12) apply simultaneously:

$$\frac{P_i + Q_i}{J \cdot K \cdot R} > 0.05 \tag{11}$$

$$\left| \frac{P_i - Q_i}{P_i + Q_i} \right| < 0.3 \tag{12}$$

where J is the number of test conditions, K is the number of test sequences, and R is the number of repetitions.

The output of the program is a text file with the names of the individual respondents who were excluded based on the above algorithm.

The processing of data from objective evaluation methods consisted of a mathematical evaluation of each method for each test sequence. The MSU Video Quality Measurement Tool was used for this evaluation [29]. The program supports a large number of video formats and objective methods and allows visualization of results or their subsequent saving in text form to a file.

Table 2 lists the summary of video sequences quality measurement results. In the case of subjective evaluation, the video sequences were rated by MOS scores that range from 1 to 5 [19, 20]. A video sequence rated by the score of 4 or higher is considered to be of high quality [9, 14].

The *ACR d.a.* and *DSIS d.a.* columns show the results for video sequences degraded by packet loss (0.5%, 1%, 3%, 5%, and 10%) and jitter (50 ms jitter in 100 ms latency). The *ACR r.a.* and *DSIS r.a.* columns show the results for video sequences degraded by packet loss and in which the degraded audio track was replaced by the audio track from the reference sequence. From the obtained results, it is clear that as the sequences deteriorate, the quality of the sequences decrease, both objectively and subjectively.

Comparing the evaluations for the ACR and DSIS methods, we found that video sequences were rated by a higher score when the DSIS method was used. This difference can be explained by the fact that in the case of the DSIS method the respondent was influenced by the reference sample.

Even at 0.5% and 1% packet loss degradation, some video sequences with the reference audio track received higher ratings than those with the original disturbed audio track. The results also imply that, in general, the degradation caused by jitter (50 ms jitter in 100 ms latency) does not affect the quality ratings as much as the degradation due to packet loss.

Table 2

Results of quality evaluation of test sequences (d.a. – degraded audio, r.a. – reference audio)

Video sequence	**Degradation**	**Subjective methods**				**Objective methods**	
		ACR d.a.	**ACR r.a.**	**DSIS d.a.**	**DSIS r.a.**	**PSNR**	**SSIM**
No. 1	**Packet loss 0.5%**	4.70	4.65	4.80	4.75	42.75	0.98
	Packet loss 1%	3.55	3.90	4.15	3.90	35.34	0.97
	Packet loss 3%	2.25	3.15	3.15	3.30	30.52	0.92
	Packet loss 5%	2.15	2.65	1.95	2.85	28.03	0.86
	Packet loss 10%	1.00	1.95	1.05	2.35	25.65	0.84
	Latency 100 ms, Jitter 50 ms	3.65	x	3.75	x	44.14	0.98
No. 2	**Packet loss 0.5%**	2.45	3.25	3.25	3.70	33.41	0.95
	Packet loss 1%	2.10	3.40	2.30	3.80	33.31	0.95
	Packet loss 3%	1.55	2.65	2.05	3.20	31.15	0.91
	Packet loss 5%	1.70	2.20	1.35	2.25	24.00	0.86
	Packet loss 10%	1.10	1.70	1.00	1.85	18.99	0.78

	Latency 100 ms, Jitter 50 ms	4.15	x	3.85	x	43.14	0.98
No. 3	**Packet loss 0.5%**	3.50	3.05	4.45	4.30	34.57	0.96
	Packet loss 1%	2.85	3.20	3.65	3.60	26.89	0.87
	Packet loss 3%	1.25	2.80	2.30	3.45	25.52	0.84
	Packet loss 5%	1.20	2.75	1.35	2.65	22.97	0.77
	Packet loss 10%	1.00	2.05	1.00	1.95	18.38	0.66
	Latency 100 ms, Jitter 50 ms	3.50	x	3.85	x	35.56	0.96
No. 4	**Packet loss 0.5%**	4.20	4.25	4.15	4.05	31.03	0.95
	Packet loss 1%	3.65	3.75	3.95	3.85	30.78	0.94
	Packet loss 3%	2.25	2.50	2.10	2.65	23.06	0.84
	Packet loss 5%	1.45	2.10	1.45	2.35	20.32	0.79
	Packet loss 10%	1.00	2.00	1.00	1.70	17.40	0.74
	Latency 100 ms, Jitter 50 ms	4.35	x	3.95	x	43.99	0.98

As we have assumed, the subjective evaluation was also influenced by pictorial information. From Figure 6, showing the comparison of the evaluation of the video sequences by the ACR method, we can clearly see that the video sequences No. 2 and No. 3 were evaluated by the lowest marks. In these video sequences, the image being transmitted was static and an emphasis was placed on the content of the audio. In the case of the objective assessment (Table 2), this difference has not been proved to such an extent.

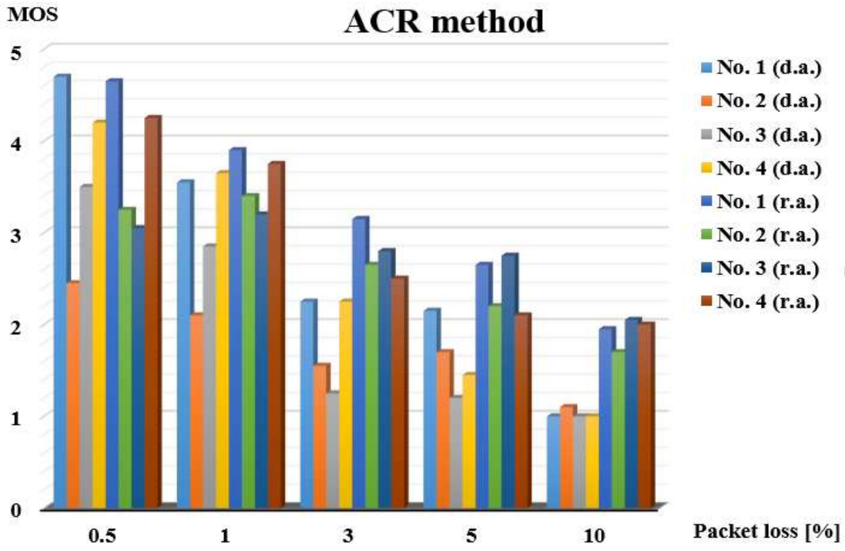


Figure 6

Comparison of the evaluation of the video sequences by the ACR method

The results of the subjective quality evaluation have shown that under the ideal conditions in the transmission network (without packet loss and latency) the quality of videoconferencing has been rated as "good" (MOS > 4). Therefore, from the perspective of the user, the video frame resolution, audio and video bitrate, and the used codecs provide the user with sufficient quality.

However, each internet protocol (IP) based transmission network will cause packet loss and latency. Their source is the non-link structure of the network. Quality codecs can at least partially compensate for the loss of information transmitted [23]. The results of the subjective quality assessment of various distorted video sequences have confirmed that packet loss of less than 1% must be achieved to obtain a very good quality videoconference.

In subjective methods (ACR, DSIS), the lowest score was evaluated for sound-related sequences (No. 2, No. 3). This confirmed that both the content of the information transmitted and the clarity of information for the evaluator play an important role in subjective quality assessment. Of course, in the videoconference that supports the learning process, the other receiving party must at least partially understand the lecture or lesson issues.

Figure 7 shows a comparison of the evaluation of the video sequences by the SSIM method. The resultant SSIM index is a decimal value between -1 and 1. The value of 1 is only reachable in the case of two identical sets of data and therefore indicates perfect structural similarity. A value of 0 indicates no structural similarity [22].

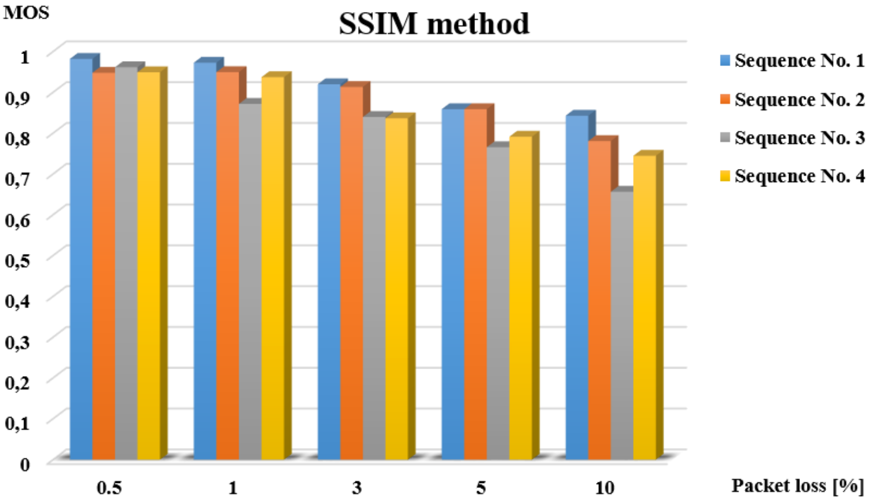


Figure 7

Comparison of the evaluation of the video sequences by the SSIM method

When evaluating video sequences using objective methods (PSNR, SSIM), video sequences No. 3 and No. 4 were scored by the lowest marks (Figure 7, Table 2).

So, we can conclude that the results of subjective and objective methods are different for our research samples. This implies that we still do not have objective methods available to replace demanding and lengthy subjective evaluation.

Based on the results of the subjective evaluation of the sequences with the original audio and the sequences in which the degraded audio was replaced by the reference, we see that for the packet loss of 3% and 5% the sequences with the reference audio are rated much higher (often by more than 1 point on the MOS scale). The difference between individual sequence evaluations is much smaller in samples with the reference audio compared to sequences with the original audio track. In our research, we have confirmed that the quality of audio has a great impact on the overall quality of videoconferencing. In future work, we can investigate whether a similar trend is observed when changing the tasks, that is, if we gradually insert different deteriorated audio tracks into the reference video sequence.

From the measured values, it also follows that in the case of 10% packet loss the respondents rated with the worst possible marks ("bad" or "poor"). In future research, the degradation with packet loss of over 10% would not make sense to test. However, it would be interesting to extend the tests with a greater number of sequences or more types of deterioration. A significant disadvantage of subjective tests is that they are time-consuming, which to a large extent limits their use. With a higher number of test sequences or a higher number of evaluators, we no longer recommend using a questionnaire for writing but a suitable software tool that would also facilitate the evaluation process.

3.1 Correlation between Objective and Subjective Methods

The correlation coefficient describes the direction and the magnitude of the relationship between two variables. It is calculated as follows:

$$r_{yx} = \frac{k_{xy}}{\sigma_x \sigma_y} \tag{13}$$

where σ_x a σ_y are standard deviations of variables x and y, respectively, and k_{xy} is their covariance calculated as:

$$k_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})$$

The value of a correlation coefficient ranges between −1 and 1. The greater the absolute value of a correlation coefficient, the stronger the linear relationship. The strongest linear relationship is indicated by a correlation coefficient of −1 or 1. The weakest linear relationship is indicated by a correlation coefficient equal to 0. A positive correlation means that if one variable gets bigger, the other variable tends to get bigger. A negative correlation means that if one variable gets bigger, the other variable tends to get smaller [26].

For each test sequence, the correlation coefficients between particular objective (SSIM, PSNR) and subjective (ACR, DSIS) methods were calculated (Table 3).

Table 3

Correlation between objective and subjective methods

	ACR d.a.	ACR r.a.	DSIS d.a.	DSIS r.a.
SSIM	0.834	0.798	0.858	0.881
PSNR	0.853	0.850	0.827	0.927

The results show that the highest correlation is between the objective metric PSNR and the subjective method DSIS with reference audio (Table 3). However, correlation results cannot be generalized based on our measurements. In general, there is no objective method by which we can completely replace the subjective perception of a person.

Conclusions

Videoconferencing technology brings vast new possibilities into the process of modern education and overcomes distance barriers. Combined with interactive computing technology, it represents the technology of the future, in the learning process.

Increasing transmission speeds in today's modern networks enable us to provide new e-learning support services such as videoconferencing, on-demand streaming, or online streaming. Both voice services (VoIP) and moving image transfer services need to be monitored to see if the service is of adequate quality to the customer. This quality monitoring must necessarily be automated because it would be impractical, financially demanding and vulnerable to errors, to employ people for these activities.

This experiment compared the subjective methods of evaluating videoconferencing quality with known objective methods and thereby contribute to the development of new objective metrics. Time-consuming subjective measurements were supported by models and programs that simplified scenario preparation, testing and results processing. These will be used in further research dealing with the measurement of video sequences quality.

The results of our comparison have confirmed that we still do not have an objective method that can fully substitute the time-consuming subjective testing. Based on the results of the subjective evaluation of sequences, with the original audio track and the sequences in which the degraded audio track was replaced by the audio track from the reference sequence, we have confirmed that the quality of the audio has a significant impact on the overall quality of videoconferencing and the ultimate understanding of its content. As a result, if any video information is supported by relevant audio information, we can compensate for the loss of video information by improving the audio quality. We can also influence the quality of videoconferencing by ensuring correct pronunciation, intelligibility and articulation.

References

[1] Azimi-Sadjadi, B. et al.: Robust Key Generation from Signal Envelopes in Wireless Networks. In CCS '07: Proceedings of the 14th ACM Conference on Computer and Communications Security, pp. 401-410, New York, NY, USA, 2007

[2] Bisták P., et al.: Utilisation of Videoconferencing for Education. 1st ed., Elfa Kosice, 2005

[3] Haffner, et al.: The multimedia as a form of modern education. In QUAERE 2018, Hradec Králové: Magnanimitas, pp. 898-907, 2018

[4] Haffner, et al.: Multimedia support for education of mechatronics. In 2018 Cybernetics & Informatics (K&I): 29th International Conference. Lazy pod Makytou, Slovakia, 2018

[5] Whittaker, S.: Video as a technology for interpersonal communications: a new perspective. Proc. SPIE 2417, Multimedia Computing and Networking, 1995, https://doi.org/10.1117/12.206055

[6] Frater, M. R., Arnold, J. F., & Vahedian, A.: Impact of audio on subjective assessment of video quality in videoconferencing applications, In *IEEE* Transactions on Circuits and Systems for Video Technology, Vol. 11, No. 9, pp. 1059-1062, Sept. 2001

[7] Heribanová, P., Polec, J., Poctavek, J. & Mordelová, A.: Intelligibility Threshold for Cued Speech in H.264 Videoconference. International Journal of Electronics and Telecommunications, 57, pp. 383-387, 2011

[8] Coudoux, F.-X., Gazalet, M. G., Derviaux, C. & Corlay, P.: Picture quality measurement based on block visibility in discrete cosine transform coded video sequences. Journal of Electronic Imaging. 2001, https://doi.org/10.1117/1.1344184

[9] Wang, Z. & Bovik, A.: Handbook of video and image processing. Academic Press. 2000

[10] Káňa, L.: Elektroakustika, ČVUT Brno, Česká republika, 2013

[11] Saadatzi, M., Saadatzi, M. N., Tavaf, V. & Banerjee, S.: Development of a PVDF based artificial basilar membrane. Proc. SPIE 10593, Bioinspiration, Biomimetics, and Bioreplication VIII, 2018

[12] Andriichenko, O. O., Denysenko, O. I.: Subjective evaluation of the clarity of the noisy language in the lecture room, Electronic and Acoustic Engineering, 2, 3, 55-60, 2019

[13] Takács, A. et al: Models for Force Control in Telesurgical Robot Systems. Acta Polytechnica Hungarica, 12, pp. 95-114, 2015

[14] Rizek, H., Brunnström, K., Wang, K., Andrén, B. & Johanson, M.: Subjective evaluation of a 3D videoconferencing system. Proc. SPIE 9011, Stereoscopic Displays and Applications XXV, 2014

[15] Mardiak, M. & Polec, J.: Novel method for objectively measuring video quality. Proceedings ELMAR-2010, Zadar, pp. 109-112, 2010

[16] Zaric, A. et al.: Image quality assessment - comparison of objective measures with results of subjective test. Proceedings ELMAR-2010, Zadar, pp. 113-118, 2010

[17] Wang, Z. & Bovic, A. C.: A universal image quality index. IEEE Signal Processing Letters, Vol. 9, pp. 81-84, March 2002

[18] Winkler, S.: Digital video quality vision model and metrics. Chichester: John Wiley & Sons Ltd, 2005

[19] ITU-R Recommendation ITU-R BT.500-11. Methodology for the subjective assessment of the quality of television pictures, 2002

[20] ITU-T Recommendation ITU-T P.910. Subjective video quality assessment methods for multimedia applications, 2008

[21] Filanová, J. & Mardiak, M.: Meranie kvality video signálu. Elektrorevue, 15, pp. 32.1-6, 2010

[22] Wang, Z., Lu, L. & Bovic, A. C.: Video quality assessment using structural distortion measurement. Signal Processing: Image Communication, special issue on "Objective video quality metrics", Vol. 19, No. 2, pp. 121-132, 2004

[23] Votruba, A. & Medvecký, M.: Evaluation of the Effectiveness of QoS Provisioning in Ethernet Networks. EE časopis pre elektrotechniku a energetiku, 17, pp. 92-96, 2011

[24] Wu, H. R. & Rao, K. R.: Digital Video Image Quality and Perceptual Coding (Signal Processing and Communications) Boca Raton: CRC Press, 2006

[25] Wu, H. R. & Rao, K. R. & Kassim, A.: Digital Video Image Quality and Perceptual Coding. Journal of Electronic Imaging 16(3), 2007

[26] Uhrina, M. & Hlubik, J. & Vaculík, M.: Correlation between Objective and Subjective Methods Used for Video Quality Evaluation. Advances in Electrical and Electronic Engineering. 11, 2012

[27] Mardiak, M. & Filanová, J.: Quality of a Video Signal. In: New Information and Multimedia Technologies. NIMT - 2008 : Brno, Czech Republic, 18.-19.9.2008. - Brno : Brno University of Technology, 2008

[28] Nambiar, M. et al.: WANem - Open Source software, Performance Engineering Research Centre, TATA Consultancy Services, Mumbai India, 2008

[29] Vatolin, D. et al.: MSU Quality Measurement Tool: Metrics information. Available: http://compression.ru/video/quality_measure/info_en.html

Applying DNN Adaptation to Reduce the Session Dependency of Ultrasound Tongue Imaging-based Silent Speech Interfaces

Gábor Gosztolya[1,2], Tamás Grósz[2,3], László Tóth[2], Alexandra Markó[4,6], Tamás Gábor Csapó[5,6]

[1] MTA-SZTE Research Group on Artificial Intelligence of the Hungarian Academy of Sciences and University of Szeged, Tisza Lajos krt. 103, H-6720 Szeged, Hungary, ggabor@inf.u-szeged.hu

[2] Institute of Informatics, University of Szeged, Árpád tér 2, H-6720 Szeged, Hungary, groszt@inf.u-szeged.hu, tothl@inf.u-szeged.hu

[3] Department of Signal Processing and Acoustics, Aalto University, Otakaari 3, FI-02150 Espoo, Finland, tamas.grosz@aalto.fi

[4] Department of Applied Linguistics and Phonetics, Eötvös Loránd University, Múzeum krt. 4/A, H-1088 Budapest, Hungary, marko.alexandra@btk.elte.hu

[5] Department of Telecommuniations and Media Informatics, Budapest University of Technology and Economics, Magyar tudósok körútja 2, H-1117 Budapest, Hungary, csapot@tmit.bme.hu

[6] MTA-ELTE Lingual Articulation Research Group, Múzeum krt. 4/A, H-1088 Budapest, Hungary

Abstract: Silent Speech Interfaces (SSI) perform articulatory-to-acoustic mapping to convert articulatory movement into synthesized speech. Its main goal is to aid the speech handicapped, or to be used as a part of a communication system operating in silence-required environments or in those with high background noise. Although many previous studies addressed the speaker-dependency of SSI models, session-dependency is also an important issue due to the possible misalignment of the recording equipment. In particular, there are currently no solutions available, in the case of tongue ultrasound recordings. In this study, we investigate the degree of session-dependency of standard feed-forward DNN-based models for ultrasound-based SSI systems. Besides examining the amount of training data required for speech synthesis parameter estimation, we also show that DNN adaptation can be useful for handling session dependency. Our results indicate that by using adaptation, less training data and training time are needed to achieve the same speech quality over training a new DNN from scratch. Our experiments also suggest that the sub-optimal cross-session behavior is caused by the misalignment of the recording equipment, as adapting just the lower, feature extractor layers of the neural network proved to be sufficient, in achieving a comparative level of performance.

Keywords: Silent speech interfaces; articulatory-to-acoustic mapping; session dependency; Deep Neural Networks; DNN adaptation

1 Introduction

Over the past few years, there has been significant interest in articulatory-to-acoustic conversion research, which is often referred to as "Silent Speech Interfaces" (SSI) [5]. The idea is to record the soundless articulatory movement, and automatically generate speech from the movement information, while the subject is not producing any sound. Such an SSI system might be very useful for the speaking impaired (e.g. after a laryngectomy), and for scenarios where regular speech is not feasible, but information should be transmitted from the speaker (e.g. extremely noisy environments and/or military situations). For this automatic conversion task, typically electromagnetic articulography (EMA, [3, 19, 20]), ultrasound tongue imaging (UTI, [4, 14, 18, 28]), permanent magnetic articulography (PMA, [10]), surface Electromyography (sEMG, [6, 16, 22]), lip video [1, 7] and multimodal approaches are used [5]. Current SSI systems mostly apply the "direct synthesis" principle, where speech is generated without an intermediate step, directly from the articulatory data. This approach has the advantage compared to Silent Speech Recognition (SSR) that there is a significantly smaller delay between articulation and speech generation, and there are fewer error possibilities than in the case of the SSR + TTS (Text-to-Speech) approach, where first the articulatory movement is translated to a phoneme or word sequence, and then it is used to generate the speech signal via standard TTS techniques.

As Deep Neural Networks (DNNs) have become dominant in more and more areas of speech technology, such as speech recognition [9, 13, 26], speech synthesis [2, 21] and language modeling [23, 24, 29], it is natural that recent studies have attempted to solve the ultrasound-to-speech conversion problem by employing deep learning, regardless of whether sEMG [17], ultrasound video [18] or PMA [10] is used as an input. Our team used DNNs to predict the spectral parameter values [4] and F0 [12] of a vocoder using UTI as articulatory input; in a later study we extended our method to include multi-task training [28].

A recent study [25] has summarized the state-of-the-art results in silent speech interfaces. Although there are lots of research findings on generating intelligible speech using EMA, UTI, PMA, sEMG, lip video and multimodal data, all the studies were conducted on relatively small databases and typically with just one or a small number of speakers [25]; while all of the articulatory tracking devices are obviously highly sensitive to the speaker. Another source of variance comes from the possible misalignment of the recording equipment. For example, for tongue-ultrasound recordings, the probe fixing headset has to be mounted onto the speaker before use, and in practice it is impossible to mount it onto exactly the same spot as before. This inevitably causes the recorded ultrasound video to become misaligned compared to a video recorded in a previous session. Therefore, such recordings are not directly comparable. In the following, by "session" it is meant that the probe fixing headset is dismounted and mounted again onto the speaker.

There have already been some studies that use multi-speaker and/or multi-session articulatory data for SSI and SSR. Kim et al. investigated speaker-independent SSR using EMA and compared Procrustes matching-based articulatory normalization, feature-space maximum likelihood linear regression and i-vector experimentally on 12 healthy and two laryngectomized English speakers [19, 20]. The best results were achieved with a combination of the normalization approaches. For EMG-based recognition, a variety of signal normalization and model adaptation methods were investigated, as experiments revealed an across-sessions deviation of up to 5 mm [22]. From the nine different normalization and adaptation procedures, sharing training data across sessions and Variance Normalization and Feature Space Adaptation proved to be the most useful [22]. Janke et al. also studied session-independent sEMG: 16 sessions of a speaker were analyzed and the results indicated that the MCD (Mel-Cepstral Distortion) in the case of cross-session conversion is only slightly worse compared to the 500 sentence session-dependent result from the same speaker, confirming that sEMG is robust even with minor changes in the electrode placement or other influence [16]. Wand et al. utilized domain-adversarial DNN training for session-independent EMG-based speech recognition [30].

Unfortunately, for ultrasound-based SSI, there are no methods currently available for the alignment / adaptation / normalization of articulatory data recorded in different sessions or with different speakers. All the above-mentioned studies [16, 20, 22, 30] used EMA or sEMG for tracking articulatory movements; and although e.g. Maier-Hein et al., state that even slight changes in electrode positions affect the myoelectric signal [22], Janke et al. found that their sEMG-based framework employing GMMs virtually behaves session-insensitively without any form of adaptation [16]. In the ultrasound-based SSI systems, however, where slight changes in probe positioning can cause shifts and rotations in the image used as input (for an example, see Fig. 1), might not turn out to be ideal.

To this end, in this study we focus on the session dependency of the ultrasound-based direct speech synthesis process. Although we also consider speaker dependency to be a significant issue, here we will just concentrate on session dependency. Notice that using recordings from different speakers inevitably means using data from different sessions as well, but without the option of identifying and analyzing the negative effect of using different speaker data (e.g. F0, speaking style, oral cavity structure) and the effect of slight changes in the position of the recording equipment. To separate the effect from the two possible error sources, in this study we shall focus on the session dependency of the ultrasound-based direct speech synthesis process. We will demonstrate experimentally, that a simple, yet, efficient, standard feed-forward DNN-based system displays clear signs of session dependency, to such an extent, that the synthesized utterances are practically unintelligible. Furthermore, we propose a simple session adaptation method, and show that it is more efficient than training a neural network from scratch using the adaptation data. We shall also examine the amount of training data required for successful DNN model adaptation. Of course, the applicability of the proposed approach for session adaptation (i.e. DNN model adaptation) is not necessarily limited to the UTI case, but it may be of interest for a broader audience as well.

2 Methods

2.1 Data Acquisition

A Hungarian female subject with normal speaking abilities was recorded while reading sentences aloud. Tongue movement was recorded in midsagittal orientation using the "Micro" ultrasound system of Articulate Instruments Ltd. at 82 fps. The speech signal was recorded with a Beyerdynamic TG H56c tan omnidirectional condenser microphone. The ultrasound data and the audio signals were synchronized using the tools provided by Articulate Instruments Ltd. (For more details, see our previous studies [4, 12, 28].) In our current experiments, the scanline data of the ultrasound recording was used. The original ultrasound images of 64×842 pixels were resized to 64×106 by bicubic interpolation, leading to 6784 features per time frame. To create the speech synthesis targets, the speech recordings (resampled to 22050 Hz) were analyzed using an MGLSA vocoder [15] at a frame shift of 1 / (82 fps), which resulted in F0, energy and 24-order spectral (MGC-LSP) features [27]. The vocoder spectral parameters (excluding F0) served as the DNN training targets.

Our data was collected in four sessions. The headset and the ultrasound probe were fitted each time using the same procedure; however, it cannot be guaranteed that the orientation of the probe remained "exactly" the same, across each session. In the first session we recorded 200 individual sentences (about 15 minutes in total), while in sessions two, three and four, we recorded 50 different sentences (less than 4 minutes each). In addition, in each session, the subject read the 9-sentence long Hungarian version of the short tale `The North Wind and the Sun'. We used the independent sentences for training purposes, while the utterances of "The North Wind and the Sun" were used as test sets. For more information about the four sessions, see Table 1.

Table 1
Key properties of the recordings used in our experiments; duration is expressed in terms of min:sec

Recording session	Individual Sentences (Train)		North Wind & Sun (Test)	
	Count	Duration	Count	Duration
Session #1	200	14:48	9	0:50
Session #2	50	3:44	9	0:49
Session #3	50	3:53	9	0:47
Session #4	50	3:41	9	0:48

Fig. 1 shows sample images taken from the four sessions with similar tongue positions. Although all four images are similar, there are visible positioning differences among them, which might lead a DNN trained on the first session to perform sub-optimally on the other sessions. We will demonstrate this sub-optimality experimentally in Section 3, and we will describe how we applied DNN adaptation to handle this issue in Section 4.

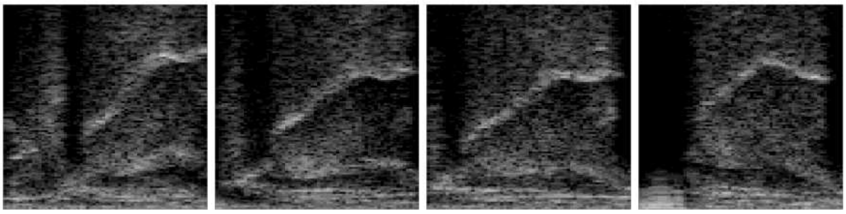


Figure 1
Sample ultrasound tongue images from the four sessions used. Note that all the images belong to the same speaker

2.2 DNN Parameters

We trained feed-forward, fully-connected DNNs with 5 hidden layers, each hidden layer consisting of 1000 ReLU neurons. The input neurons corresponded to the image pixels, while the output layer contained one linear neuron for each MGC-LSP feature and one for the gain (25 output parameters overall). To assist prediction, we presented a time slice of the ultrasound video (five consecutive frames) as input to the DNN, since in our previous studies [4, 12, 28] we found this technique to be beneficial. The input images consisted of 6784 pixels, meaning that the network had a total of 33920 input neurons.

2.3 Evaluation

As estimating the parameters of the synthesizer is a simple regression problem, the most suitable evaluation metric is the Pearson correlation; or, in our case, as we have 25 speech synthesis parameters to predict, we will take the mean of the 25 correlation values. In our earlier studies, [28], we also used this evaluation metric. In our last experiments, however, in order to determine which proposed system is closer to natural speech, we also conducted an online MUSHRA (MUlti-Stimulus test with Hidden Reference and Anchor) listening test [31]. The advantage of MUSHRA is that it allows the evaluation of multiple samples in a single trial without breaking the task into many pairwise comparisons. Our aim was to compare the natural sentences with the synthesized sentences of the baseline, the proposed approaches (various session adaptation variants) and a benchmark system (the latter being cross-session synthesis without adaptation). In the test, the listeners had to rate the naturalness of each stimulus in a randomized order relative to the reference (which was the natural sentence), from 0 (very unnatural) to 100 (very natural). We chose sentences from 4-layer adaptation and full training, and tested two adaptation data sizes (20 and 50 sentences). Altogether 96 utterances were included in the test (12 sentences x 8 variants). In the MUSHRA evaluation, each configuration was evaluated by 12 native Hungarian speakers with normal hearing abilities.

3 Results with Single-Session DNN Training

3.1 The Effect of the Amount of Training Data

In our first experiments, we examined how the amount of training data affects the performance of the DNN model. For this, we trained our neural network on the recordings of the same session that we used for testing. We used $N = 1, 5, 10, 20$ and 50 sentences for training, and evaluated our models on the 9 sentences of `The North Wind and the Sun' from the same session. Since for Session #1 we had more utterances in the training data, there we also experimented with $N = 100$, 150 and 200.

The mean correlation values obtained this way have been plotted in Fig. 2. Clearly, the correlation scores vary to a great extent among the different sessions, though at this point we did not perform any cross-session experiments: DNN training and evaluation were performed by using recordings taken from the same session. We can also see that, by increasing the number of training sentences, the correlation values increased, as expected. Also note that, when we used more than $N = 100$ sentences (roughly 7 minutes of recordings), there is a slight improvement only, although we had only one session with enough training data to confirm this.

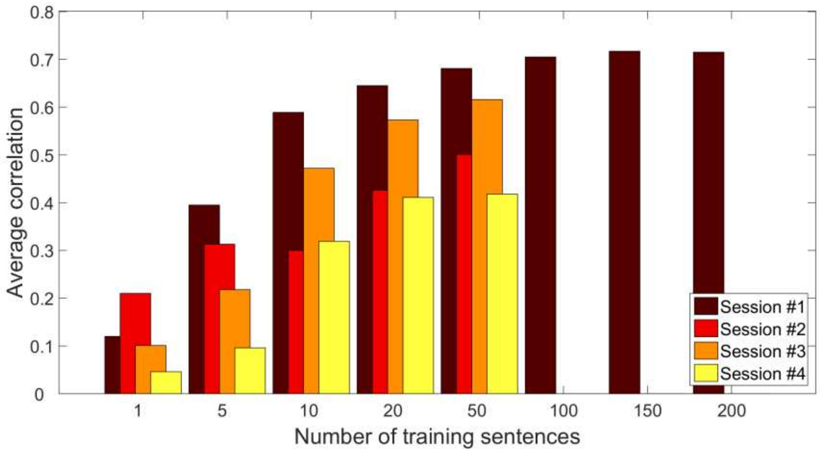


Figure 2

Average correlation values obtained for the four sessions as a function of the number of sentences used for training

Examining our sample images (see Fig. 1), it is hard to see any difference among the sessions which might explain the significant difference in the average correlation scores observed in Fig. 2. Perhaps the only exception is the large dark area in the posterior region (on the left hand side of the image) in session #4, where not only the hyoid bone blocked the ultrasound waves (as it did on the other images), but also there was probably insufficient amount of gel between the transducer and the skin, limiting the visibility in that particular direction. However, for session #2 we got similarly low correlation scores, while the ultrasound video contained no such artifact. Since we fitted the recording equipment following the same procedure for each session, these results alone, in our opinion, indicate that UTI-based SSI systems are session-sensitive even without using data taken from multiple speakers.

3.2 Cross-Session Results

In our next experiment, we sought to examine how the misalignment of input images affects the performance of the neural network. To this end, we trained our DNN on all the 200 sentences of the first session, and evaluated it on the utterances of `The North Wind and the Sun' recorded in the remaining three sessions.

Table 2

Average correlation scores obtained for the recordings of `The North Wind and the Sun' depending on the DNN training data

Training Data		Average correlation for sessions			
Session	Size	#2	#3	#4	Avg.
Session #1	200	0.075	0.100	0.143	0.106
Same as test	50	0.501	0.616	0.418	0.512

The first row of Table 2 shows the average correlation values obtained this way. We can see that the DNN predictions are practically worthless, as the average Pearson's correlation values fall between 0.075 and 0.143. (We also confirmed the low quality of these predictions by listening tests, and found the synthesized 'utterances' unintelligible.) In contrast (see the second row), using just 50 sentences for DNN training, but from the same session, we get average correlation scores in the range 0.418-0.616. This huge difference, in our opinion, also demonstrates that ultrasound-based DNN SSI approaches are quite sensitive to misalignments of the ultrasound images, even if these come from the same speaker, and this issue has to be handled if we intend to develop SSI systems for practical use.

4 DNN Adaptation

In the previous section we showed experimentally that DNN models trained on the recordings of one session cannot be utilized to predict speech synthesis parameters in another session, even when both sessions were recorded with the same speaker. Next, we will show that the issue of session-dependency can be handled effectively via the adaptation of the DNN model trained on data from a different session. In practice, adaptation means that we train the DNN further, using recordings taken from the actual session. For the general scheme of the proposed approach, see Fig. 3. Of course, to ease the use of our SSI equipment, this adaptation material has to be as short as possible, hence we simultaneously aim for high-quality spectral parameter estimation while keeping the amount of adaptation data to a minimum. To this end, we performed DNN adaptation experiments using $N = 1, 5, 10, 20$ and 50 sentences from each session; we used once again the 9 sentences of `The North Wind and the Sun' of the actual session for evaluation purposes.

It is well known (e.g. [8, 11]) that the lower layers of a deep neural network are responsible for low-level feature extraction, while the higher layers perform more abstract and more task-dependent functions. As in our case session dependency appears as a change in the input image, while the task remains the same (i.e. to predict the spectral representation of the speech of the same speaker), it seems reasonable to expect that it might be sufficient to train just the lower layers of the network instead of adapting all the weights. This way, we might achieve the same level of accuracy with faster training, or obtain better estimates [11]. Since in our experiments we employed DNNs with five hidden layers, we have six choices of which layers to adapt (i.e. only the weights between the input layer and the first hidden layer, adapt the weights among the input layer and the first two hidden layers, etc.). To test this, we also experimented with adapting just the first two and first four layers of the network. Furthermore, as a comparison, we also tried training a DNN from scratch using $N = 1, 5, \ldots, 50$ sentences on data taken from the same session as our baselines.

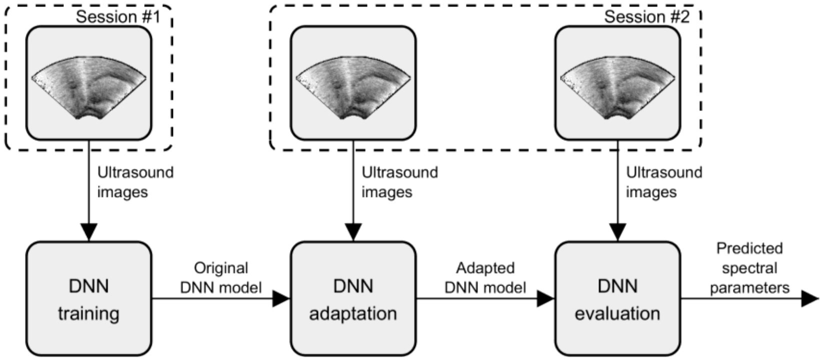


Figure 3
The general workflow of the proposed DNN SSI model adaptation procedure

4.1 DNN Adaptation Results

4.1.1 Correlation Values

Fig. 4 shows the average correlation value**s** measured, as a function of the number of training sentences. The scores are averaged out for the three sessions (i.e. Session #2, #3 and #4); the error bars represent minimal and maximal values. We can see that, in general, if we used more sentences either for DNN training or for adaptation, the accuracy of the predictions improved. It is also quite apparent that when we have only a few sentences taken from the current session, adaptation leads to more accurate predictions than training a randomly initialized DNN. For the $N = 20$ and $N = 50$ cases, however, full DNN training resulted only in slightly lower correlation values than adaptation did. Still, even when we have a higher number of sentences, we can state that by using DNN adaptation, fewer sentences are needed to achieve the same performance as with full DNN training. For example, adapting 3 layers with 10 utterances (about 20-25 seconds) of training data from the given session leads to roughly the same averaged correlation score that can be achieved by using 20 sentences and full DNN training.

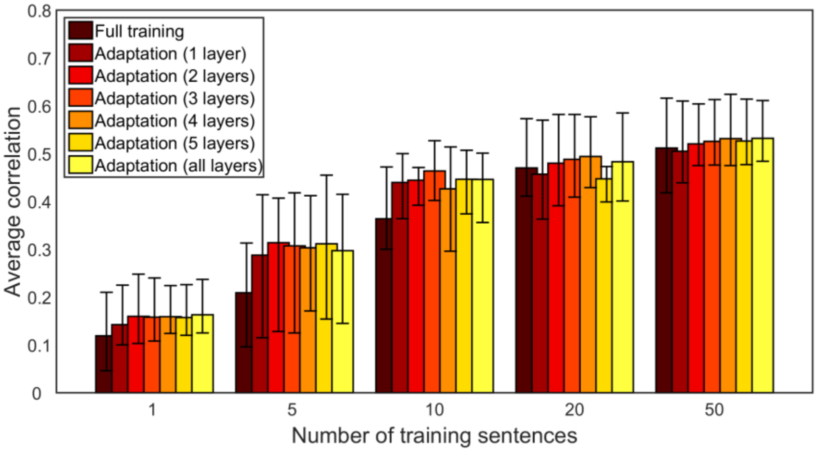


Figure 4
Average correlation scores via full DNN training and DNN adaptation as a function of the number of sentences used

Regarding the number of layers adapted, there are only slight differences in DNN performance. Although adapting only one layer (i.e. the weights between the input and the first hidden layer) led to the lowest correlation value in each case, the remaining five variations proved to be quite similar, and usually adapting the first four layers (for $N = 10$, three layers), proved to be optimal.

Inspecting the minimal and maximal correlation scores for each configuration, these values usually behaved just like the mean correlation scores did: adapting only one layer resulted in a suboptimal performance, but when we adapted at least two layers, there were no large differences. However, it is quite apparent that for the case $N = 50$ and adapting at least two layers, the minimal correlation value greatly exceeded that of full training, while the maximal scores appeared to be roughly the same. For an SSI system used in everyday practice, where we have no guarantee of the precision of the current equipment positioning, the minimal performance of the (adapted or newly trained) DNN model might be just as important as the average one; and in this respect, DNN adaptation performed much better than full DNN training did.

Table 3 lists the notable correlation scores for all three sessions and their average. These numeric values confirm our previous findings; namely, the average performance of full DNN training always falls closer to the best correlation score of DNN adaptation using fewer sentences than using the same amount of training data. Furthermore, for the case $N = 50$, full DNN training led to a correlation value of 0.418 as the worst score, while for adaptation it is never lower than 0.475.

Table 3
Average correlation scores obtained for `The North Wind and the Sun' depending on the amount of DNN adaptation data

No. of Train Sentences	Adapted Layers	Average correlation for sessions			
		#2	#3	#4	Avg.
10	Full training	0.300	0.472	0.319	0.364
	Input to 2nd	0.471	0.470	0.392	0.444
	Input to 3rd	0.462	0.527	0.402	0.464
	All layers	0.481	0.501	0.356	0.446
20	Full training	0.426	0.573	0.411	0.470
	Input to 2nd	0.467	0.582	0.391	0.480
	Input to 3rd	0.476	0.577	0.429	0.494
	All layers	0.463	0.585	0.401	0.483
50	Full training	0.501	0.616	0.418	0.512
	Input to 2nd	0.475	0.604	0.482	0.520
	Input to 3rd	0.475	0.624	0.495	0.531
	All layers	0.484	0.611	0.501	0.532

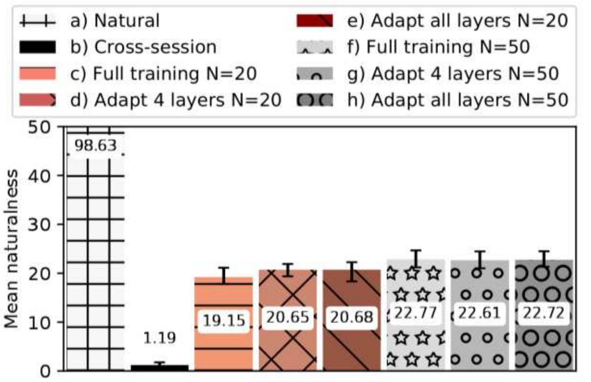


Figure 5
Mean naturalness scores of the MUSHRA listening test; error bars show the 95% confidence intervals

4.1.2 MUSHRA Listening Tests

Fig. 5 shows the results obtained from the MUSHRA listening tests. (The samples used in the test can be found at http://smartlab.tmit.bme.hu/actapol2019_ssi_session.) The naturalness of the synthesized utterances turned out to be somewhat low in each case, probably due to the small size of the training data (i.e. 20 or 50 sentences overall, equivalent to about 90 seconds and less than 4 minutes of duration, respectively). Still, the effect of the number of sentences used for training or adaptation is clearly visible: using no adaptation led to unintelligible speech (a mean naturalness score of only 1.19), while using 20 sentences resulted in naturalness scores between 19.15 and 20.68, which increased to 22.61-22.77 for the case *N* = 50. The listening tests also reinforced our previous findings that for *N* = 20, DNN adaptation is a better approach, while for *N* = 50 there is no observable difference among the output of the full DNN training and the DNN adaptation techniques. According to the Mann-Whitney-Wilcoxon ranksum test with a 95% confidence level, differences between variants c) to h) (i.e. the tested models with *N* = 20 and *N* = 50) were not statistically significant.

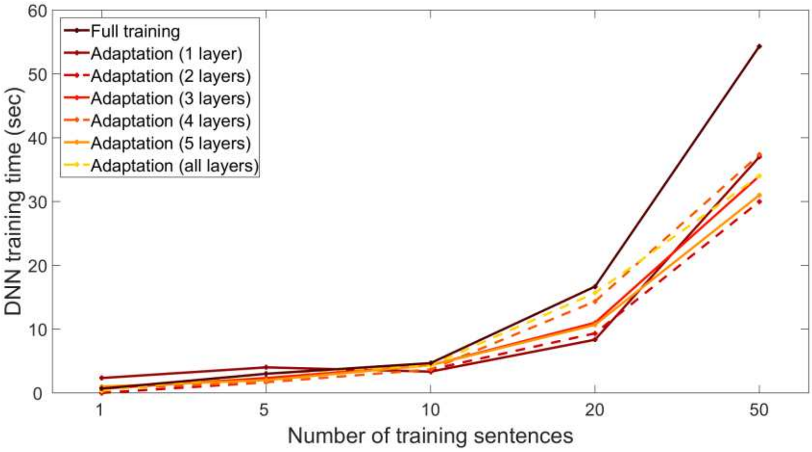


Figure 6

Average wall clock training times as a function of the number of sentences used for training

4.1.3 DNN Training Times

Fig. 6 shows the (wall clock) DNN training and DNN adaptation times expressed in seconds (averaged out for the three sessions), measured on an Intel i7 4.2 GHz PC with 32 GB RAM and an NVidia Titan X video card. From these values, it is clear that the DNN adaptation time is primarily affected by the size of the adaptation data: for *N* = 10, the average values fell between 3 and 5 seconds, which increased to 8-15 seconds for *N* = 20 and to 30-37 seconds for *N* = 50. In contrast, full DNN training took 17 seconds for *N* = 20 and 54 seconds for *N* = 50. From these values, however, we cannot confirm that adapting fewer layers leads to lower execution times; in our experience, DNN adaptation time is primarily affected by the size of the adaptation data. Full DNN training led to by far the highest training time in the *N* = 50 case, while for *N* = 20 its training time is much higher than those of most adaptation configurations. This indicates that DNN adaptation has a further advantage: it allows quicker convergence than training a DNN with random initial weights. Specifically, for the case *N* = 50, DNN adaptation required about two-thirds the time compared to DNN training from scratch did; and adapting a DNN with 20 sentences needed far less training time (17-29%) to achieve the same performance as full DNN training did with *N* = 50.

Overall, from these results, DNN adaptation with 20 sentences seems to be the best approach, since it requires significantly less training material than full DNN training in the case $N = 50$, and it was also much quicker to train. Furthermore, it led to a higher minimal correlation value, while the average correlation and MUSHRA naturalness scores appeared to be quite similar, and the difference was not statistically significant.

Conclusions

In this study, we focused on the session dependency of the ultrasound-based direct speech synthesis process, during articulatory-to-acoustic mapping. Similarly to studies using sEMG [16, 22] and EMA [19, 20], we investigated how the reattachment of the articulatory equipment affects the final output. For the first time in the scientific community, we used ultrasound tongue imaging for this purpose, building on our earlier single-session studies [4, 12, 28]. We expected that reattaching the probe would greatly diminish the accuracy of a previously trained system.

We found that our hypothesis was supported by the following results:

1) The synthesized speech was unintelligible if the network was trained on one session and evaluated on another session as-is (without the adaptation of the network weights)
2) We found large differences even among the performance of DNN models used within the same session, depending on the actual session
3) To create a DNN model for the actual session, DNN adaptation performed better than full DNN training did during UTI-to-spectral feature conversion

Furthermore, DNN adaptation had the advantage of allowing quicker convergence than random DNN weight initialization did.

The findings of our experiments are an important step within the articulatory-to-acoustic research area, as the simple-yet-effective adaptation method proposed herein, should contribute to the development of practical and efficient Silent Speech Interfaces. For example, a DNN adaptation with 20 sentences takes roughly 15 seconds on a current computer (such as the Intel i7 4.2 GHz PC used in our experiments), after which, speech can be synthesized directly from ultrasound-based articulatory data. However, the current study was conducted on regular speech and it is a future task to experiment with real silent (mouthed) speech. In the future we also plan to investigate the speaker-dependency of the ultrasound tongue imaging.

Acknowledgement

The authors were partially funded by the Ministry of Human Capacities, Hungary (grant no. 20391-3/2018/FEKUSTRAT), the National Research, Development and Innovation Office of Hungary (grants no. FK 124584 and PD 127915) and by the

MTA “Lendület” grant. Gábor Gosztolya and László Tóth were supported by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences and by the ÚNKP-19-4 New National Excellence Program of the Ministry of Human Capacities.

References

[1] H. Akbari, H. Arora, L. Cao, and N. Mesgarani, “LIP2AUDSPEC : Speech reconstruction from silent lip movements video,” in Proceedings of ICASSP, Calgary, Canada, 2018, pp. 2516-2520

[2] M. S. Al-Radhi, T. G. Csapó, and G. Németh, “Deep Recurrent Neural Networks in speech synthesis using a continuous vocoder,” in Proceedings of SPECOM, Hatfield, Hertfordshire, UK, 2017, pp. 282-291

[3] B. Cao, M. Kim, J. R. Wang, G. Van Santen, T. Mau, and J. Wang, “Articulation-to-speech synthesis using articulatory flesh point sensors' orientation information,” in Proceedings of Interspeech, Hyderabad, India, 2018, pp. 3152-3156

[4] T. G. Csapó, T. Grósz, G. Gosztolya, L. Tóth, and A. Markó, “DNN-based ultrasound-to-speech conversion for a silent speech interface,” in Proceedings of Interspeech, Stockholm, Sweden, 2017, pp. 3672-3676

[5] B. Denby, T. Schultz, K. Honda, T. Hueber, J. M. Gilbert, and J. S. Brumberg, “Silent speech interfaces,” Speech Communication, Vol. 52, No. 4, pp. 270-287, 2010

[6] L. Diener, and T. Schultz, “Investigating objective intelligibility in real-time EMG-to-Speech Conversion,” in Proceedings of Interspeech, Hyderabad, India, 2018, pp. 3162-3166

[7] A. Ephrat and S. Peleg, “Vid2speech: Speech reconstruction from silent video,” in Proceedings of ICASSP, New Orleans, LA, USA, 2017, pp. 5095-5099

[8] T. Gao, J. Du, L.-R. Dai, and C.-H. Lee, “Joint training of frontend and back-end Deep Neural Networks for robust speech recognition,” in Proceedings of ICASSP, Brisbane, Australia, 2015, pp. 4375-4379

[9] X. Glorot, A. Bordes, and Y. Bengio, “Deep sparse rectifier networks,” in Proceedings of AISTATS, Fort Lauderdale, FL, USA, 2011, pp. 315-323

[10] J. A. Gonzalez, L. A. Cheah, A. M. Gomez, P. D. Green, J. M. Gilbert, S. R. Ell, R. K. Moore, and E. Holdsworth, “Direct speech reconstruction from articulatory sensor data by machine learning,” IEEE/ACM Transactions on Audio, Speech, and Language Processing, Vol. 25, No. 12, pp. 2362-2374, 2017

[11] G. Gosztolya and T. Grósz, “Domain adaptation of Deep Neural Networks for Automatic Speech Recognition via wireless sensors,” Journal of Electrical Engineering, Vol. 67, No. 2, pp. 124-130, 2016

[12] T. Grósz, G. Gosztolya, L. Tóth, T. G. Csapó, and A. Markó, "F0 estimation for DNN-based ultrasound silent speech interfaces," in Proceedings of ICASSP, Calgary, Alberta, Canada, 2018, pp. 291-295

[13] G. Hinton, L. Deng, D. Yu, G. Dahl, A.-r. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. Sainath, and B. Kingsbury, "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," IEEE Signal Processing Magazine, Vol. 29, No. 6, pp. 82-97, 2012

[14] T. Hueber, E.-l. Benaroya, B. Denby, and G. Chollet, "Statistical mapping between articulatory and acoustic data for an ultrasoundbased silent speech interface," in Proceedings of Interspeech, Florence, Italy, 2011, pp. 593-596

[15] S. Imai, K. Sumita, and C. Furuichi, "Mel log spectrum approximation (MLSA) filter for speech synthesis," Electronics and Communications in Japan (Part I: Communications) Vol. 66, No. 2, pp. 10-18, 1983

[16] M. Janke, M. Wand, K. Nakamura, and T. Schultz, "Further investigations on EMG-to-speech conversion," in Proceedings of ICASSP, Kyoto, Japan, 2012, pp. 365-368

[17] M. Janke and L. Diener, "EMG-to-speech: Direct generation of speech from facial electromyographic signals," IEEE/ACM Transactions on Audio, Speech, and Language Processing, Vol. 25, No. 12, pp. 2375-2385, 2017

[18] A. Jaumard-Hakoun, K. Xu, C. Leboullenger, P. Roussel-Ragot, and B. Denby, "An articulatory-based singing voice synthesis using tongue and lips imaging," in Proceedings of Interspeech, San Francisco, CA, USA, 2016, pp. 1467-1471

[19] M. Kim, B. Cao, T. Mau, and J. Wang, "Multiview representation learning via deep CCA for silent speech recognition," in Proceedings of Interspeech, Stockholm, Sweden, 2017, pp. 2769-2773

[20] M. Kim, B. Cao, T. Mau, and J. Wang, "Speaker-Independent Silent Speech Recognition From Flesh-Point Articulatory Movements Using an LSTMNeural Network," IEEE/ACM Transactions on Audio, Speech, and Language Processing, Vol. 25, No. 12, pp. 2323-2336, 2017

[21] H.-Z. Ling, S.-Y. Kang, H. Zen, A. Senior, M. Schuster, X.-J. Qian, H. Meng, and L. Deng, "Deep learning for acoustic modeling in parametric speech generation," IEEE Signal Processing Magazine, Vol. 32, No. 3, pp. 35-52, 2015

[22] L. Maier-Hein, F. Metze, T. Schultz, and A. Waibel, "Session independent non-audible speech recognition using surface electromyography," in Proceedings of ASRU, San Juan, Puerto Rico, 2005, pp. 331-336

[23] G. Melis, C. Dyer, and P. Blumsom, "On the state of the art of evaluation in neural language models," Proceedings of ICLR, Vancouver, BC, Canada, 2018

[24] T. Mikolov, M. Karafiát, L. Burget, J. Černocký, and Sanjeev Khudanpur, "Recurrent neural network based language model," in Proceedings of Interspeech, Makuhari, Japan, 2010, pp. 1045-1048

[25] T. Schultz, M. Wand, T. Hueber, D. J. Krusienski, C. Herff, and J. S. Brumberg, "Biosignal-based spoken communication: A survey," IEEE/ACM Transactions on Audio, Speech, and Language Processing, Vol. 25, No. 12, pp. 2257-2271, 2017

[26] F. Seide, G. Li, X. Chen, and D. Yu, "Feature engineering in context-dependent deep neural networks for conversational speech transcription," in Proceedings of ASRU, Big Island, HI, USA, 2011, pp. 24-29

[27] K. Tokuda, T. Kobayashi, T. Masuko, and S. Imai, "Melgeneralized cepstral analysis – a unified approach to speech spectral estimation," in Proceedings of ICSLP, Yokohama, Japan, 1994, pp. 1043-1046

[28] L. Tóth, G. Gosztolya, T. Grósz, A. Markó, and T. G. Csapó, "Multi-task learning of phonetic labels and speech synthesis parameters for ultrasound-based silent speech interfaces," in Proceedings of Interspeech, Hyderabad, India, 2018, pp. 3172-3176

[29] Z. Tüske, R. Schlüter, and H. Ney, "Investigation on LSTM Recurrent N-gram Language Models for Speech Recognition," in Proceedings of Interspeech, Hyderabad, India, pp. 3358-3362, 2018

[30] M. Wand, T. Schultz, and J. Schmidhuber, "Domain-adversarial training for session independent EMG-based speechrecognition," in Proceedings of Interspeech, Hyderabad, India, 2018, pp. 3167-3171

[31] "ITU-R recommendation BS.1534: Method for the subjective assessment of intermediate audio quality," 2001

Enhanced Adaptive Random Test Case Prioritization for Model-based Test Suites

Tomas Pospisil, Jan Sobotka and Jiri Novak

Czech Technical University in Prague, Faculty of Electrical Engineering,
Technicka 2, 166 27 Prague, Czech Republic, E-mail: pospito7@fel.cvut.cz,
jan.sobotka@fel.cvut.cz, jnovak@fel.cvut.cz

Abstract: Adaptive Random Prioritization is a Test Case Prioritization technique which orders test cases within a test suite with a goal of earlier fault detection using semi-random heuristics. Compared to other Test Case Prioritization methods, Adaptive Random Prioritization has only, an "average fault detection performance. However, it is less sensitive to some test suite features which negatively affect fault detection performance than other TCP techniques due to its semi-random nature. The article proposes an improved version of Adaptive Random Prioritization technique. The key idea behind the presented enhancement is to extend the test case selection process with additional information about control flow and change of test statements coverage, of a test suite. The enhancement replaces the original Test set distance function with a Multi-Criteria Decision-Making method. Validity of the proposed method is evaluated on data from six embedded systems. The evaluation criterion is fault detection performance expressed by Average Percentage of Faults Detection metric and Â12 statistic. The proposed improvement achieved better fault detection performance for all of the examined systems.

Keywords: Adaptive Random Testing; Model-based testing; Multi-Criteria Decision-Making

1 Introduction

Testing of modern electronic/software systems is an essential activity in the quality assurance process. Many approaches for how to design and execute a test suite currently exist. A test suite consists of individual test cases. Since testing resources are always limited, a research of optimization techniques focused on cost reduction is needed [1]. One way of testing automation is Model-Based Testing (MBT) [2, 3]. In MBT world, test cases are generated by a software tool from a model. E.g., System-under-Test (SUT) behavior is modeled by Finite State Machine and a test is an oriented path through the automaton. A test suite is produced by a graph traversal algorithm like a Breadth-First search. Due to the automatic nature, a Model-Based Test Suite can contain a tremendous number of generated test cases. Optimization by ordering of these test cases according to given (structural model coverage, random and stochastic, data coverage ...) criteria is required. This problem is dealt with a category of techniques referred to as Test Case Prioritization (TCP) [4].

Recently, TCP techniques were explored mainly in code-based and regression testing areas [5, 6]. On the other hand, TCP techniques for MBT approach are not so well investigated. The main difference between these areas is that general TCP techniques in MBT do not use any external information, as historical information [6] or expert knowledge (Risk-based TCP [7]). Since this context is not so profoundly explored, further research in this area is needed.

Results of a study on TCP techniques in the MBT area [8] show that no TCP technique has superior performance. Examined techniques performance varies for different scenarios. The study indicates TCP techniques based on Path Complexity and additional coverage of statements achieve good overall performance, but they are affected by the size of the test cases that fail. In contrast, Adaptive Random Prioritization (ARP) is less sensitive to the size of the test cases that fail, but they have only moderate fault detection results.

Based on these results, an enhanced ARP technique is proposed. The presented technique replaces a standard *Test set distance* function (see Section 2.1) with a Multi-Criteria Decision-Making (MCDM) method [9]. The method combines criteria based not only on the distance metric but also on other test case features. These features include additional information to TCP process from various sources, which can be code-based, model-based, can use previous tests results, etc. The presented technique is mainly designed for testing of the embedded systems, described with high-level behavioral models and for cases where other information as source code, or fault history is not available. Therefore, the proposed technique uses only the MBT metrics with high fault detection performance. The primary goal is to improve the fault detection performance of adaptive random based techniques and simultaneously preserve their low sensitivity to the size of the test cases that fail.

2 Background

This section contains a summary of the ARP technique and other methods that are incorporated in proposed MCDM ARP enhancement, specifically Path Complexity and Additional coverage techniques. Besides that, a short overview of TCP in MBT area is outlined at the end of the section.

2.1 Adaptive Random Prioritization

Adaptive Random Prioritization technique [10] belongs to a family of Adaptive Random Strategy (ARS) techniques. Chen et al. [11] initially proposed the ARS as an online test case generation method for software with numerical inputs. ARS was proposed as an alternative to a pure random test input generation strategy. The idea behind ARS is an input space estimation that selects areas that cause failures and spreads the test cases more efficiently over the input space.

The ARP is an application of ARS to the TCP problem. The technique is described in the Algorithm – Part 1. Where the *UTS* input is an unsorted test suite, and the output of the function is a prioritized test suite (*PTS*). In the first step (line 3), the algorithm randomly selects a test case. It is saved as the first prioritized test case in *PTS* and subsequently removed from the *UTS* set (lines 4 and 5).

Algorithm – Part 1: The main procedure

```
1: function Prioritize(UTS)
2:      PTS ← Ø
3:      first ← randomChoice(UTS)
4:      PTS.add(first)
5:      UTS.remove(first)
6:      while UTS ≠ Ø do
7:          cand ← genCandSet(UTS)
8:          nextTC ← selectNext(PTS, cand)
9:          PTS.add(nextTC)
10:         UTS.remove(nextTC)
11:     end while
12: return PTS
13: end
```

The prioritization of the suite is executed within the main loop on lines 6 to 11. In the first part of the loop (line 7), a set of candidates is generated according to Part 2 of the Algorithm.

Algorithm – Part 2: Candidates generation

- *UTS* is the test suite
- *cand* is the candidate set
- *candMax* is the maximal size of *cand* set
- *S*: $\{s_1, s_2, ...\}$ is set of statements
- *S'*: $\{s'_1, s'_2, ...\}$ is set of statements

```
1: function genCandSet (UTS)
2:      S ← Ø
3:      S' ← Ø
4:      cand ← Ø
5:      TC ← randomSelect(UTS)
6:      S ← statements covered by TC
```

```
7:      if $S' \cup S = S'$ then return cand
8:      $S'$.add($S$)
9:      cand.add($TC$)
10:     if cand.size == candMax then return cand
11:     goto 5
12: end
```

The candidate set *cand* is iteratively selected by the candidate generation function. In each iteration step, the generation process randomly selects a not-yet-prioritized test case (Part 2 – line 5). The selection process continues until newly selected test cases increase coverage of candidate set (Part 2 – line 5 to 10), or until the maximum number of candidates is reached.

The next step (Part 1 – Line 8) is the *selectNextTestCase* function, which selects a test case from the candidate set; Part 3 describes this function in detail. The *selectNextTestCase* function is based on functions f_1 and f_2. Function f_1 (Part 3 – line 5) is *Test case distance* function that calculates distances between candidates and the test cases that were already prioritized and saves them into the distance matrix *d*. Jiang et al. [10] used Jaccard distance function [12] instead of Euclidean distance function, which was proposed for the original ARS. Zhou [13] proposed an application of modified Manhattan distance function. Lately, Zhou et al. compared both distance functions in an empirical study [14]. The study shows that the Manhattan function provides better fault detection performance than Jaccard in the code-based context. Further improvement was proposed by Coutinho et al. [15], where Similarity function was implemented instead of distance functions.

```
Algorithm – Part 3: Next test case selection
        • PTS is the already prioritized test sequence
        • cand is the candidate set
1: function selectNextTestCase(PTS, cand)
2:      $d \leftarrow$ array[PTS.size][cand.size]
3:      for $i = 0$ to PTS.size - 1 do
4:          for $j = 0$ to cand.size - 1 do
5:              $d[i,j] \leftarrow f_1(PTS[i], cand[j])$
6:          end for
7:      end for
8:      $index \leftarrow f_2(d)$
9:      $nextTestCase \leftarrow$ cand.get(index)
10: return nextTestCase
11: end
```

Function f_2 is *Test set distance* function that returns the index of the selected test case that is farthest away from the prioritized set (Part 3 – line 8). This function can select test cases according to several prioritization rules. For example, the rule MaxMin – maximum of the minimum distances (similarities) first, determines the smallest distance of each candidate to all already prioritized test cases and then finds a candidate with the lagest value of those minimal distances. If this rule is applied to the distance matrix in Figure 1, then the candidate TC_2 would be selected. Jiang et al. [10] also examined MaxMax and MaxAvg variants. In the presented enhanced version of ARP technique, the function f_2 is replaced with an MCDM method (see Section 3.1).

	Prioritized TC$_1$	Prioritized TC$_2$	Prioritized TC$_3$
Candidate TC$_1$	0.2	0.5	0.3
Candidate TC$_2$	0.8	0.4	0.1
Candidate TC$_3$	0.5	0.6	0.2

Figure 1
Distance matrix

In the last steps, the selected test case is removed from *UTS* and added into the prioritized test sequence *PTS*. (Part 1 – lines 9 and 10). This process is repeated until the prioritization process is finished (all the test cases are sorted).

2.2 Path Complexity

Kaur et al. [16] originally designed Path Complexity (PC) TCP technique for systems modeled as UML activity diagrams. Each test case is represented as a path through a Control Flow Graph (CFG), which is converted from a UML nested activity diagram. For each path *P* from a generated set, several properties are calculated:

- N_p - the number of nodes traversed by *P*
- W_p - weight of test path *P*
- P_p - number of predicate nodes traversed by *P*
- C_p - number of logical conditions traversed by *P*
- complexity *C* of *P* by using the formula $C = N_p + W_p + P_p + C_p$

Where the weight of the path is based on Information Flow metric (*IF*). The *IF* metric was designed by Sharma et al. [17] and originally applied to the components of system design. In our case, it is used on each node *N* in the CFG model, and it is calculated as:

$$IF(N) = FANIN(N) \times FANOUT(N) \qquad (1)$$

where, *N* is a node from CFG, *FANIN*(*N*) is a number of incoming flows to *N*, and *FANOUT*(*N*) is a number of outgoing flows from N. The weight of the path is a sum of IF from all nodes in the path.

$$W_p = \sum_{i=1}^{n} w_i \forall P \in T_P \quad (2)$$

Where W_p is weight of the path *P*, w_i is weight of i^{th} node (*IF(N)*), *n* is a count of nodes in the current path *P,* and T_p is set of generated paths. When the complexity for all paths is calculated, the test cases (paths) are executed in order of complexity from highest to lowest.

2.3 Additional Coverage

The additional coverage technique is based on greedy reasoning applied to TCP [4]. The technique progresses iteratively through a test suite. In each step, a test case that yields the highest coverage of not yet covered statements, is selected.

2.4 Related Work

Code-based and regression testing are the most investigated TCP approaches [5, 6, 18, 19]. In the MBT area, proposed TCP techniques are frequently connected with UML (Activity) diagrams models [14, 20-22]. These techniques implement a wide range of strategies. Some of the strategies are modified variants from code-based context; they can be relatively straightforward as Path Complexity, or more advanced ones that include historical data and data mining techniques.

The proposed technique extends ARP technique, which is a general black-box technique. In this context, an interesting article was presented by Hemmati et al. [23]. The article compares three different black-box TCP (code-based) approaches: topic coverage, text diversity, and risk-driven heuristic. In the topic coverage TCP, topics are extracted from a textual description of test cases and their expected results by a text mining algorithm. In the subsequent step, the technique tries to prioritize test cases in a way to maximize coverage of those topics. The text diversity approach prioritizes test cases using a string distance between two text representations of the test cases. The risk-based approach uses information about detected faults in their previous executions for test case prioritization. If historical information is available, the results show that the risk-based approach is superior. However, none of the approaches significantly outperform the others, if history is not available.

Empirical study [24] performed by João Felipe Silva Ouriques et al. investigates the effect of the model structure and characteristics of test cases that fail on the fault detection capability of several TCP techniques. Study results show that the characteristics of failed test cases affect the investigated techniques more significantly than the model layout. Due to the study being performed on synthetic data that may not exactly correspond to real systems, authors publish a replication study [8] on data from real industrial projects. In this study, multiple general TCP techniques were evaluated with similar results as in previous work. The results show that none of the compared techniques is the best, concerning fault detection ability. Besides, the study investigates a dependency between fault detection performance of particular techniques and different properties of test cases. Specifically, the effect of varying sizes of the test cases that fail is examined. Authors conclude that the examined techniques have different sensitivity to this test case feature. The presented article builds on these studies and presents a new enhanced version of ARP technique, which improves the fault detection performance of the standard version and preserves limited effect of this negative feature.

3 Novel Enhanced ARP Technique

The main objective of the novel enhanced ARP technique is to improve fault detection performance over the original ARP method. The ARP demonstrates consistent results in different scenarios thanks to its semi-random nature. The study results [8] show there are techniques with better fault detection performance, but their performance is also more affected by various test suite properties. One of those properties is an inconsistent fault detection performance between cases where failed test cases are shorter or longer than the average test case size in a test suite (see Section 4.1). The development of the proposed TCP technique was focused on fault detection performance enhancement while keeping relatively low sensitivity on the size of the test cases that fail.

The proposed modification replaces *Test set distance* function (function f_2 Algorithm 1 Part 3 – Line 8) with Weighted Product Model (MCDM) method [9]. The method compares candidates (test cases) among themselves, using distance, path complexity and additional coverage criteria. The novel enhanced TCP technique based on the MCDM method was named Multi-Criteria Adaptive Random Prioritization technique (MC-ARP). The advantage of MC-ARP is that the criteria can be extended/exchanged in a situation when a new source of information becomes available (e.g., fault history during testing of an updated system version).

The introduced MC-ARP technique partially decreases the importance of test case distances and increases the chance of selection for test cases which cover more not yet covered statements, or test cases with higher path complexity. The method can be tuned by weights, which determine the strength of these properties and thus also results of the prioritization. The novel MC-ARP technique is described in the following subsection.

3.1 Multi-Criteria ARP Technique

The proposed application of Weighted Product Model method has *m* alternatives (not-yet prioritized test cases) and *n* decision criteria. The method compares alternatives among themselves, using these decision criteria, and determines the one that is better than others. The decision criteria are benefit criteria (higher values are better), and they are divided into three main groups:

- Distance criteria – performance value of a distance criterion corresponds to distance between a particular candidate and already sorted test cases (distance matrix – see Section 2.1)
- Path Complexity – path complexity value of the candidate (see Section 2.2)
- Additional coverage – represents the count of newly covered statements, if the candidate is selected (see Section 2.3)

Variable *w*, which determines the relative weight of importance of the criteria, is assigned to each of these criteria groups. Due to the fact that the count of prioritized test cases changes during the TCP process (each time when a new sorted test case is added), the weight of the individual criterion can be calculated as an equal share of initial weight for distance criteria w_{DC}.

$$w_i = \frac{w_{DC}}{n_{DC}} \tag{3}$$

Where w_i is weight for a specific distance criterion (prioritized test case), and n_{DC} is the number of prioritized test cases. Weights for path complexity w_{PC} and additional cover w_{AC} do not change during algorithm iterations. Thus, the ratio between all weights is always the same; only the values of w_i are iteratively changed.

The performance value of candidate test case TC_i during evaluation by criterion *j* is denoted as pv_{ij}. The following function $P(TC_K/TC_L)$ compares two candidates TC_K and TC_L. In case the result is higher than value 1, the first candidate is superior to the second. The newly selected candidate should be better than or at least equal to all other candidates.

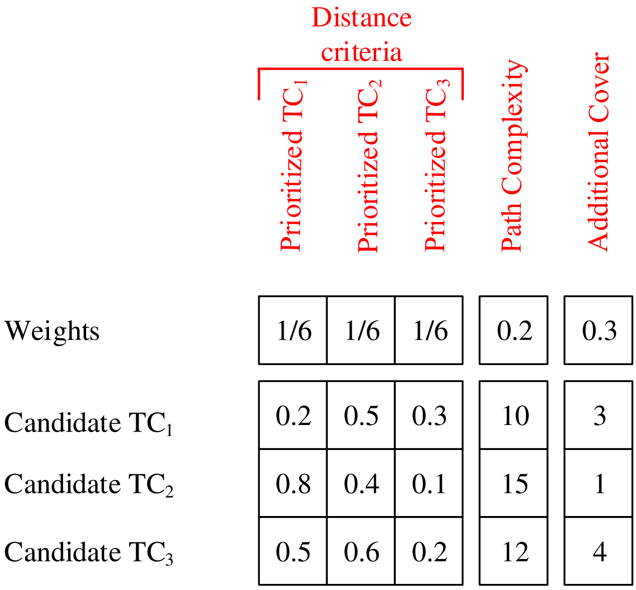
$$P\left(TC_K/TC_L\right) = \prod_{j=1}^{n}\left(\frac{pv_{Kj}}{pv_{Lj}}\right)^{w_j}, \text{ for } K, L = 1, 2, 3\ldots\; m; \text{ and } K \neq L \tag{4}$$

The example shows the method on model data. Let us say, there are three candidates and three already prioritized test cases, weights are set to $w_{DC} = 0.5$, $w_{PC} = 0.2$ and $w_{AC} = 0.3$, and the performance values of candidates are depicted in Figure 2, where columns represent particular criteria, the first row shows weights for a specific criterion, and other rows are connected to candidate test cases (alternatives). In our case, distance criteria performance values match to the distance matrix d from the original ARP example. However, other criteria are now considered in the test case selection.

The comparison of candidate TC_1 and TC_2 would be as follows:

$$P(TC_1/TC_2) = (0.2/0.8)^{\frac{1}{6}} * (0.5/0.4)^{\frac{1}{6}} * (0.3/0.1)^{\frac{1}{6}} * (10/15)^{0.2} * (3/1)^{0.3}$$
$$\cong 1.33$$

Similarly, $P(TC_1/TC_3) \cong 0.79$ and $P(TC_2/TC_3) \cong 0.62$. Therefore, the selected candidate is TC_3 (instead of TC_2 from the ARP example), since it is better than all of the other candidates.



	Distance criteria				
	Prioritized TC$_1$	Prioritized TC$_2$	Prioritized TC$_3$	Path Complexity	Additional Cover
Weights	1/6	1/6	1/6	0.2	0.3
Candidate TC$_1$	0.2	0.5	0.3	10	3
Candidate TC$_2$	0.8	0.4	0.1	15	1
Candidate TC$_3$	0.5	0.6	0.2	12	4

Figure 2
MC-ARP Example

4 Experiments

In this section, it is verified whether the proposed MC-ARP technique achieves better fault detection results than the original ARP. The second goal of the experiments is to investigate the effect of the size of the test cases that fail on fault detection performance of the proposed technique. The experimental evaluation on a broader sample of test suites is important for the assessment of the newly proposed technique, because TCP techniques may have an outstanding performance for one test suite and weak results for another.

The section is divided into two main parts. The first part describes the experimental setup and other necessary aspects for the evaluation of the experiments. In the second part, the performance comparison of MC-ARP with the original ARP technique and investigation of sensitivity to the size of the test cases that fail, is presented.

4.1 Dataset

Dataset used in this section for evaluation of the proposed technique is obtained from [25]. The dataset includes 17 test suites and information about faults from six industrial systems (for overall characteristics see Table 1). Each system is covered by two to four test suites, and each test suite has 4 to 24 test cases (for more information see [8]). The projects included in the dataset are from different areas, e.g., a cashdesk system that interacts with payment terminals, or a system to manage lending of equipment/software and maintenance logs. The tested systems were modeled in a high abstraction level as control-flow Labelled Transition System models. The models represent use scenarios, which can include multiple types of control flows (standard scenario, alternative user's behavior, and exception flow that covers systems errors). Test cases were generated as paths through these models by traverses of a Depth-First Search algorithm.

Table 1

Systems overall characteristics [8]

System	Language	Size (LOC)
S1	Java	3000
S2	C	3055
S3	Java	13001
S4	Groovy grails	3693
S5	Groovy java	20713
S6	Groovy JavaScript grails	13244

Sensitivity to the size of the test case that has found a fault is defined by a relation between the sizes of test cases that fail and the rest of the test suite. The relation divides test cases into two groups. The first group contains short test cases that execute fewer steps than the average number (test cases that commonly do not traverse loops). In contrast, the second group consists of long test cases, which perform more system steps than the average. In order to evaluate this sensitivity, it is necessary to know the failed test cases in advance. Then the test suites can be divided into the following groups:

- ShortTC – test suites where every test case that fails is shorter than the average size of test case in the test suite.
- LongTC – test suites where every test case that fails is longer than the average size of test case in the test suite.
- ConstantSizeTC – test suites where all test cases have the same size.
- MixedTC – test suites which do not fit to above mentioned groups.

The terminology is taken from the original study [8], and the distribution of these groups in the dataset is shown in Figure 3.

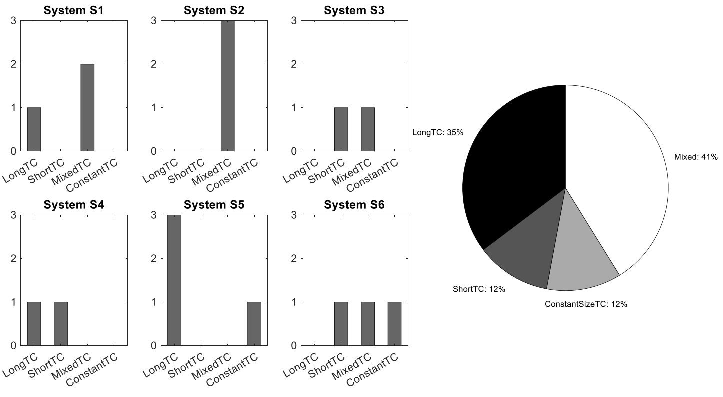


Figure 3
Test Suites Distribution

Random nature of ARP technique can affect TCP results. Therefore, the examined algorithms were executed 1000 times for each test suite (according to the suggestion in [26]); the experimental setup is shown in Figure 4.

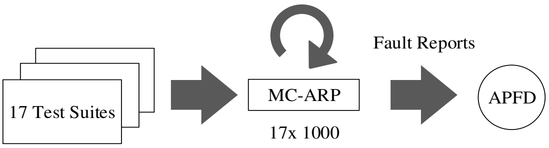


Figure 4
Setup overview

4.2 Effectiveness Metrics

To evaluate the performance of the new enhanced ARP technique and to compare results with the original methods, the following metrics are used.

Average Percentage of Faults Detection (APFD) developed by Elbaum et al. in [27] is used for evaluation of fault detection performance. The metric measures the rate of fault detection per percentage of test suite execution and is calculated as follows:

$$APFD = 1 - \frac{TF_1 + TF_2 + ... + TF_m}{mn} + \frac{1}{2n} \tag{5}$$

where n is the number of test cases, and m is the number of faults which the test suite can reveal. The TF_i is the position of the first test case that reveals the i–th fault. The APFD is a percentage (values 0 – 100), and higher values indicate better (faster) fault detection.

Non-parametric statistical test, Kruskal-Wallis test [28], is applied to compare two distributions of the APFD results. The test determines whether the difference between the results is statistically significant using a 95% confidence level (i.e. p-value < 0.05). The test resolves whether the differences are not random, but for performance comparison Vargha and Delaney's $\hat{A}_{12}$ statistic [29] is used.

Vargha and Delaney's $\hat{A}_{12}$ statistic performs a pairwise comparison of results expressed by the APFD metric from two techniques A and B. It is a nonparametric effect size measure, which is popular in the software engineering area, where randomized algorithms are involved [30]. The $\hat{A}_{12}$ metric measures the probability that running *A* produces a higher APFD than running *B*. The $\hat{A}_{12}$ can be calculated:

$$\hat{A}_{12} = \frac{\frac{R_1}{m} - \frac{m+1}{2}}{n} \tag{6}$$

Where R_1 is the rank sum of the APDF results from the first compared technique (the ranking is done through the results of both techniques). The m is the number of results from the first technique, and n is the number of results from the second technique. If the two techniques are equivalent, then $\hat{A}_{12} = 0.5$. In another case, one technique produces better results.

4.3 Results

The performance evaluation of the presented technique was done by comparison of fault detection capabilities of the original and enhanced technique. For evaluation, several variants of ARP technique with the following *Test case* *distance* functions were chosen: Jaccard (**ARP$_{Jac}$**), Manhattan (**ARP$_{Man}$**), and Similarity function (**ARP$_{Sim}$**). The original technique with Jaccard and Manhattan uses MaxMin *Test set distance* function, and Similarity function uses MaxMin and MaxMax *Test set distance* functions (marked **ARP$_{Sim1}$** and **ARP$_{Sim2}$**). The proposed Multi-criteria ARP technique is marked **MC-ARP$_{XXX}$**, where XXX distinguishes a specific variant with appropriate *Test case distance* function. The criteria weights for MC-ARP were experimentally set to: $w_{DC} = 0.5$, $w_{PC} = 0.2$ and $w_{AC} = 0.3$.

The overall fault detection results of MC-ARP and original variants are presented in Figure 5, and results for individual systems are shown in Figure 6. The pairwise Â12 results comparison can be found in Table 2.

Table 2
Effect sizes of pairwise comparisons of MC-ARP and the original technique

	ARP$_{Jac}$	ARP$_{Man}$	ARP$_{Sim1}$	ARP$_{Sim2}$
System S1	0.69	0.68	0.58	0.55
System S2	0.70	0.64	0.68	0.68
System S3	0.51	0.5	0.51	0.58
System S4	0.59	0.58	0.70	0.71
System S5	0.58	0.56	0.62	0.62
System S6	0.62	0.64	0.53	0.58
Overall	**0.59**	**0.58**	**0.59**	**0.61**

The overall comparison in Figure 5 shows that, in all cases, MC-ARP variants have higher median value (better results), and boxplots are also more compact (i.e., they have shorter interquartile ranges and more consistent results). Presented results from Table 2 show that MC-ARP improves the overall performance of ARP technique, and the improved fault detection performance is similar for each distance function variant.

Results for individual systems indicate that the more significant performance improvement is evident in systems S1, S2, and S4. These systems mostly contain MixedTC and LongTC test suites (see Figure 3). This improvement is due to Path Complexity and Additional Coverage criteria, which prefer more complex test cases with a higher amount of non-covered statements. In other cases, MC-ARP technique still has the same or better results than the original ARP variants i.e., boxplots are more compact or median values are similar or higher.

For overall results, Kruskal-Wallis test produces the highest p-value < 0.001. Therefore, the techniques present different performances than their original versions. For individual systems, the results are statistically similar (p-value > 0.05) only in the case of the system S3 for techniques **ARP$_{Jac}$**, **ARP$_{Man}$**, and **ARP$_{Sim1}$**.

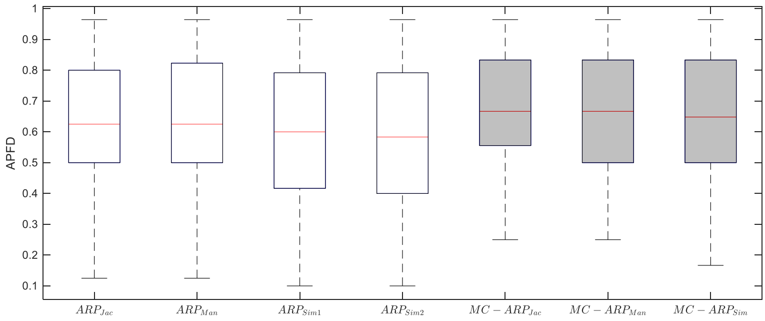


Figure 5
Comparison of overall results

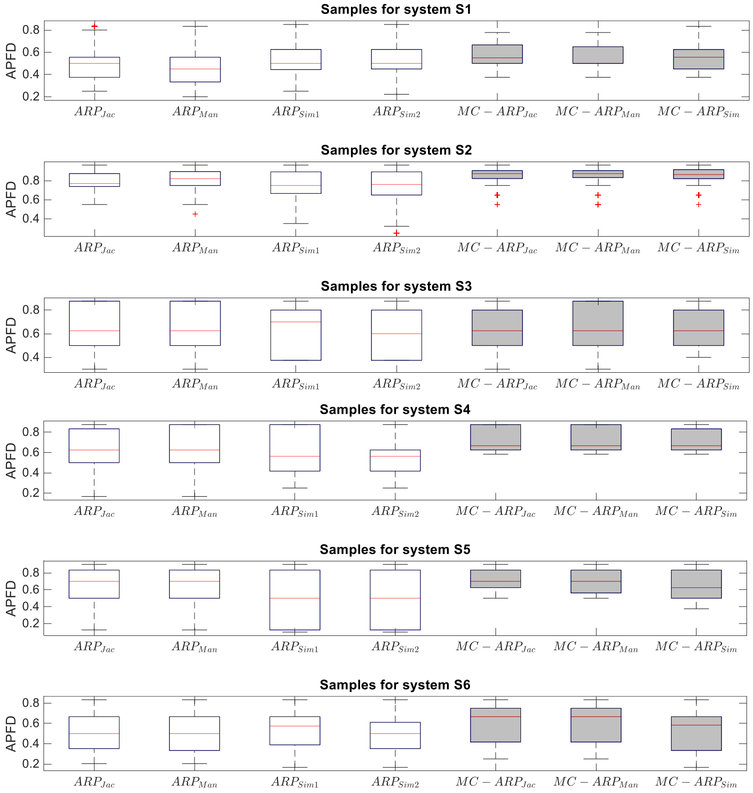


Figure 6
Performance of techniques for individual SUT

The effect of the size of the test cases that fail is investigated using ShortTC and LongTC suite samples. These samples contain test suites where failed test cases are shorter or longer than the average test cases size in that test suite.

Figure 7 presents the results for evaluated techniques. Detailed pairwise comparison is in Table 3 for ShortTC, respectively LongTC. The results show that the MC-ARP has better fault detection performance than the original ARP technique (mainly due to path complexity guidance) for LongTC samples. The MC-ARP variants have the same or higher median values and more compact boxplots. On the other hand, for ShortTC samples, Table 3 presents that MC-ARP has only slightly better performance than the original technique for Jaccard and Manhattan distance functions. The MC-ARP variant with Jaccard function has a higher median, but the boxplot is more spread out. The Manhattan variants have similar boxplots; however MC-ARP achieve higher median value. In case of Similarity function, the results are worse than the original technique, which is mainly noticeable on the boxplot for **$\mathbf{ARP_{Sim1}}$**.

In the comparison of MC-ARP and the original variants, Kruskal-Wallis test results reach maximal p-value < 0.001. Hence, the techniques also produce different performance results.

Table 3

Effect sizes of pairwise comparisons of MC-ARP and the original ARP for ShortTC and LongTC

ShortTC	ARP_{Jac}	ARP_{Man}	ARP_{Sim1}	ARP_{Sim2}
$MC\text{-}ARP_{XXX}$	0.54	0.54	0.33	0.42
LongTC	ARP_{Jac}	ARP_{Man}	ARP_{Sim1}	ARP_{Sim2}
$MC\text{-}ARP_{XXX}$	0.60	0.57	0.60	0.60

Pairwise Â12 between ShortTC and LongTC for a technique represents sensitivity to the size of the test cases that fail. The technique insensitive to the size of the test cases that fail should reach a value 0.5 when performance for both groups is the same. The results for ARP and MC-ARP techniques are presented in Table 4. At this point, the proposed technique achieved a minor decrease compared to the original ARP values. This increase of sensitivity is caused by an unequal improvement of results between ShortTC and LongTC samples. However, the sensitivity of MC-ARP still achieves decent values in comparison to other TCP techniques (for more information about these techniques see [8]).

Records for ARP technique variants are duplicated in Table 4. Values based on data from our ARP implementation have the blue italic font. The results of the original study are listed in black font. These values have been added, because they are slightly dissimilar to the original, probably due to a different implementation of the ARP algorithm.

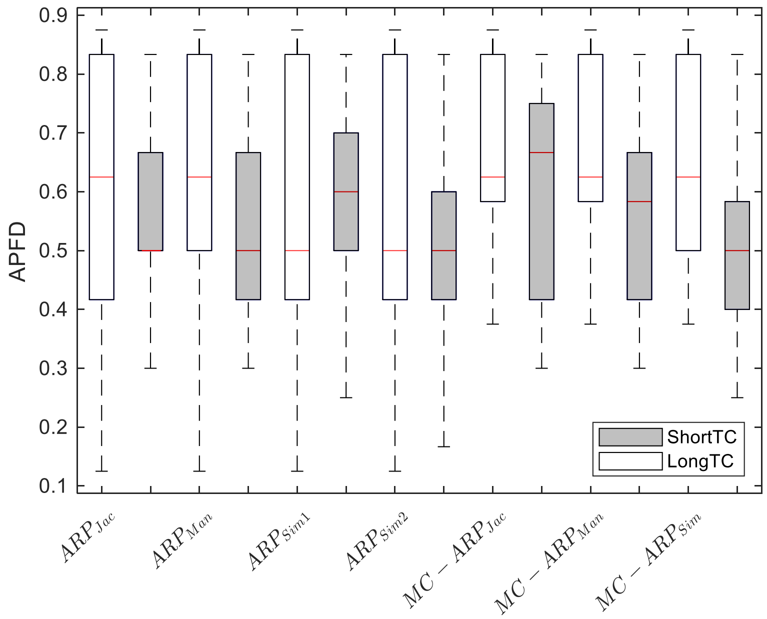


Figure 7

Boxplots of ShortTC and LongTC samples

Table 4

Effect sizes of the comparisons between ShortTC and LongTC [8]

Technique	Â12	Technique	Â12
Ran	0.4443	PC	0
ARP$_{Jac}$	0.3945	Stoop	0.5833
ARP$_{Man}$	0.374	SD$_{h}$	0.0833
ARP$_{Sim1}$	0.4725	SD$_{e}$	0.1666
ARP$_{Sim2}$	0.3892	SD$_{m}$	0
ARP$_{Jac}$	*0.44*	ST	1
ARP$_{Man}$	*0.4*	SA	0.1609
ARP$_{Sim1}$	*0.51*	**MC-ARP$_{Jac}$**	**0.35**
ARP$_{Sim2}$	*0.4*	**MC-ARP$_{Man}$**	**0.34**
FW	1	**MC-ARP$_{Sim}$**	**0.22**

Conclusions

This work presents a new Adaptive Random Prioritization technique (referred to as MC-ARP); the technique replaces the original *Test set distance* function, with a Multi-Criteria Decision-Making method. This enhancement incorporates additional criteria (other than the test case distances) into a decision which test case should be selected from the candidate set. The key idea of ARP technique performance improvement is to add guidance based on the path complexity and additional coverage techniques that have better overall fault detection performance. The PC technique calculates a score based on a path through a control flow model for each test case and the test cases with a higher score, are preferred over the others. An additional cover technique calculates how many yet, uncovered statements, will be covered when a test case will be added to the set of already the prioritized ones. The test case with the highest number of newly covered statements is preferred.

The mentioned features and distances between a candidate and already prioritized test cases are used as decision criteria in Weighted Product Model method. The method performs a pairwise comparison of all candidates. The proposed change partially limits the random nature of ARP and prefers to select more complex test cases or test cases with more uncovered statements over others.

The novel *Test set distance* function helps to improve the fault detection performance of ARP technique. Improvement of fault detection performance across all tested systems has been noticed. Moreover, experiment results show that MC-ARP has the same or better performance across all systems than the original ARP.

For test suites where every test case that fails is shorter than the average size (i.e., the test cases with less complex paths), MC-ARP with Jaccard, or Manhattan distance functions achieve slightly better results. Thanks to the guidance, MC-ARP outperforms the original technique when every test case that fails is longer than the average size of test cases in the test suite. The resulting sensitivity to the size of the test cases that fail is moderately higher than the original technique, due to uneven performance improvements in the scenarios mentioned above. However, it still achieves good results compared to other techniques.

Regarding threats to validity, the evaluation of the proposed MC-ARP technique was performed on the dataset that contains data from six industrial projects. The systems were modeled as Labelled Transition, and the test suites were generated by Depth-First Search algorithm, which traverse the models and saves paths as test cases. Therefore, the results cannot be generalized for other kinds of systems or other test case generation algorithms. However, it can be assumed, that for similar systems and test case generation approaches, the technique will perform at similar performance level. The evaluation was done only on fault detection and the effect of the size of test cases that fail, however, other aspects as the number of test cases in the dataset or the proportion of test cases that fail may affect the fault detection performance. Moreover, some test suites in the dataset are relatively small, and they may not be entirely suitable for prioritization.

In future work, the technique herein will be applied in the area of automotive Hardware-in-the-Loop (HIL) Integration Testing [31]. In this domain, MBT approach is used for Integration Testing of Electronic Control Units (ECUs). The goal of this testing phase is to estimate if a cluster of ECUs operates in synergy and functions distributed among multiple ECUs work as expected. There is a need for optimization of automatically generated test suites. Those test suites are generated from Timed Automata models using our testing tool called Taster [32, 33]. The size of a test suite depends on a Timed Automata model complexity and used a state-space traversal algorithm. In general, it can be enormous. The MC-ARP will be implemented as part of this software. The expectation is to attain test suite optimization, towards shorter test times, while maintaining a reasonable test coverage. Further experiments will be performed on a HIL testing platform, in cooperation with our Industrial Partner. This HIL testbed is dedicated to the testing of comfort and radar sensor-based driver-assistance systems.

Acknowledgement

The work was supported from ERDF/ESF project "Advanced Testing of Automotive Radars" (No. CZ.02.1.01/0.0/0.0/16_025/0007318). This support is gratefully acknowledged.

References

[1] Ammann, P., J. Offut: Introduction to software testing, 1 ed., Cambridge University Press, New York, NY, USA, 2008

[2] Utting, M., B. Legeard: Practical model-based testing: a tools approach. Elsevier, 2010

[3] Zander, J., I. Schieferdecker, and P. J. Mosterman: A taxonomy of model-based testing for embedded systems from multiple industry domains, Model-based testing for embedded systems, 2011, pp. 3-22

[4] Elbaum, S., A. G. Malishevsky, and G. Rothermel: Test case prioritization: A family of empirical studies. IEEE transactions on software engineering, 2002, 28(2): pp. 159-182

[5] Catal, C. and D. Mishra: Test case prioritization: a systematic mapping study. Software Quality Journal, 2013, 21(3): pp. 445-478

[6] Khatibsyarbini, M., et al.: Test case prioritization approaches in regression testing: A systematic literature review. 2018, 93: p. 74-93

[7] Stallbaum, H., A. Metzger, and K. Pohl: An automated technique for risk-based test case generation and prioritization. in Proceedings of the 3rd international workshop on Automation of software test. 2008

[8] Ouriques, J. F. S., E. G. Cartaxo, and P. D. Machado: Test case prioritization techniques for model-based testing: a replicated study. Software Quality Journal, 2017, pp. 1-32

[9] Triantaphyllou, E.: Multi-criteria decision making methods, in Multi-criteria decision making methods: A comparative study. 2000, Springer, pp. 5-21

[10] Jiang, B., et al.: Adaptive random test case prioritization. in Proceedings of the 2009 IEEE/ACM International Conference on Automated Software Engineering. 2009

[11] Chen, T. Y., H. Leung, and I. Mak: Adaptive random testing. in Annual Asian Computing Science Conference. 2004, Springer

[12] Jaccard, P.: Comparative study of the floral distribution in a portion of the Alps and the Jura Mountains (Étude comparative de la distribution florale dans une portion des Alpes et des Jura) 1901, 37: pp. 547-579

[13] Zhou, Z. Q.: Using coverage information to guide test case selection in adaptive random testing. in 2010 34th Annual IEEE Computer Software and Applications Conference Workshops. 2010

[14] Zhou, Z. Q., A. Sinaga, and W. Susilo: On the fault-detection capabilities of adaptive random test case prioritization: Case studies with large test suites. in System Science (HICSS), 2012 45th Hawaii International Conference on. 2012

[15] Coutinho, A. E. V. B., E. G. Cartaxo, and P. D. de Lima Machado: Analysis of distance functions for similarity-based test suite reduction in the context of model-based testing. Software Quality Journal, 2016, 24(2): pp. 407-445

[16] Kaur, P., P. Bansal, and R. Sibal: Prioritization of test scenarios derived from UML activity diagram using path complexity. in Proceedings of the CUBE International Information Technology Conference. 2012

[17] Sharma, C., S. Sabharwal, and R. Sibal: Applying genetic algorithm for prioritization of test case scenarios derived from UML diagrams. arXiv:1410.4838, 2014

[18] Kumar, A. and K. J. C. Singh: A Literature Survey on test case prioritization. 2014, 3(5): p. 793

[19] Korel, B., L. H. Tahat, and M. Harman: Test prioritization using system models. in Software Maintenance, 2005. ICSM'05. Proceedings of the 21st IEEE International Conference on. 2005

[20] Mahali, P., D. P. J. I. J. o. S. A. E. Mohapatra, and Management: Model based test case prioritization using UML behavioural diagrams and association rule mining. 2018, 9(5): pp. 1063-1079

[21] Kundu, D., et al.: System testing for object-oriented systems with test case prioritization. Software Testing, Verification Reliability, 2009, 19(4): p. 297-333

[22] Sapna, P. and H. Mohanty: Prioritization of scenarios based on uml activity diagrams. in Computational Intelligence, Communication Systems and Networks, 2009, CICSYN'09, First International Conference on. 2009

[23] Hemmati, H., Z. Fang, and M. V. Mantyla: Prioritizing manual test cases in traditional and rapid release environments. in Software Testing, Verification and Validation (ICST), 2015 IEEE 8th International Conference on. 2015

[24] Ouriques, J. F. S., et al.: Revealing influence of model structure and test case profile on the prioritization of test cases in the context of model-based testing. 2015. 3(1): p. 1

[25] Ouriques, J. F. S.: Replication of Failure Characteristics Experiment; Available from:https://sites.google.com/site/joaofso/research/experiments/replication-of-failure-characteristics-experiment

[26] Arcuri, A. and L. Briand: A practical guide for using statistical tests to assess randomized algorithms in software engineering. in Software Engineering (ICSE), 2011 33rd International Conference on. 2011

[27] Elbaum, S., A. G. Malishevsky, and G. Rothermel: Prioritizing test cases for regression testing. Vol. 25, 2000

[28] Sheskin, D. J.: Handbook of parametric and nonparametric statistical procedures. crc Press. 2003

[29] Vargha, A. and H. D. Delaney: A critique and improvement of the CL common language effect size statistics of McGraw and Wong. Journal of Educational Behavioral Statistics

[30] Poulding, S. and J. A. Clark: Efficient software verification: Statistical testing using automated search. IEEE Transactions on Software Engineering, 2010. 36(6): pp. 763-777

[31] Sobotka, J. and J. Novak: Automation of automotive integration testing process. in 2013 IEEE 7th International Conference on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS) 2013

[32] Krejci, L. and J. Novak: Model-based testing of automotive distributed systems with automated prioritization. in Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), 2017 9th IEEE International Conference on. 2017

[33] Sobotka, J. and L. Krejci: Testing of Automotive Systems-Complex vs. Simple Environment Models. in 2018 16th Biennial Baltic Electronics Conference (BEC). 2018

Selected Problems of Family Business: A Case Study from Slovakia

Tomáš Peráček[1,*], Lucia Vilčeková[2] and Ľubomíra Strážovská[2]

[1]Department of Informations Systems, Faculty of Management, Comenius University in Bratislava, Odbojárov 10, 820 05 Bratislava, Slovak Republic; tomas.peracek@fm.uniba.sk

[2]Department of Marketing, Faculty of Management, Comenius University in Bratislava, Odbojárov 10, 820 05 Bratislava, Slovak Republic; lucia.vilcekova@fm.uniba.sk; lubomira.strazovska@fm.uniba.sk

[*]Corresponding author

Abstract: The scientific paper deals with a part of the business sector, which is made up of family businesses. The paper presents the current status of such businesses from the perspective of positive and negative factors which will be linked to the problem areas. In this article, we have focused our attention on some aspects of family business, especially the managerial aspects, because the management of a family business has various differences and specifics compared to other types of businesses. In the theoretical part, we present the current state of the issue, while the empirical part of the article is based on a survey conducted among family businesses using a questionnaire. This article does not aim to highlight the contentious areas of family business. However, it brings valuable findings of business practice. More than 400 enterprises were approached, the resulting sample consisted of 185 family enterprises. Therefore, we understand the results as a case study from Slovakia. Our findings were subject to statistical analysis using several quantitative methods (t-test, regression models) and we present them in the empirical part. Based on our results, we bring the most valuable findings and ideas for further research.

Keywords: family business; business problems; t-test; regression model

1 Introduction

The importance of doing business in every market economy lies mainly in the development of the economy and job creation. The term business environment is a term known and often used, yet its definition is neither simple nor unambiguous, especially given the large number of entities involved in its design. The business environment reflects the quality of economic conditions and the basis for the economic activity of entrepreneurs [6]. The business environment of the Slovak

Republic has been examined by researchers since the political change in 1989 when the economy underwent a change from centrally planned to a market economy. For decades, the declared effort of the Government of the Slovak Republic has been to create a favorable business environment. This is essential not only for large investors but also for small and medium-sized enterprises, which are often operated as family businesses in Slovak conditions. This particular category of businesses represents the most powerful engine of the economy in the developed countries, especially in Western Europe, as they represent a number of advantages.

In essence, family and family businesses are two intertwining and interacting worlds. But there are values and principles that make up the family and shape the business. The entrepreneur's family life, personality, and interests are often the main driving force in doing business. However, an entrepreneur has to deal with a situation where his or her business threatens the family, but also with a situation where his/her own family threatens the business and thus the economic existence.

Small and middle-sized enterprises are most vulnerable to changes in the business environment, they have limited access to trades where large investments are required, they may be threatened by multinational or large enterprises, they cannot afford to employ top scientists, managers, professionals. Last but not least, they are unable to monitor available resources.

As Mura states, in recent years, several economic and political instruments have been adopted to support small and middle-sized business [24]. However, we see the problem is the fact that none of them focused exclusively on family businesses. Unlike the Slovak Republic, the European Union has adopted a number of measures aimed at increasing the competitiveness of small and medium-sized enterprises, but also of family businesses. The above-mentioned legislative measures are also adopted in connection with the administration of the family business. According to these, EU legislation measures, up to 85% of all businesses can be considered as family-owned, and these businesses employ up to 60% of all employees.

The business process, as well as family business, are much more developed in western countries, which has been caused in particular by the favorable political situation, but also by centuries of evolving the business processes. Therefore, western countries are an inspiration for our young entrepreneurs.

The past three decades in Slovakia were shaped by the emergence and subsequent development of small and medium-sized enterprises [17, 18]. The merits of the representation of these business units in the country's economy were confirmed. In the transition to a market economy, it was once again small and middle-sized enterprises that helped significantly transform the economy and laid the foundation for the functioning of smaller businesses. As part of the transformation of the economy, families have returned to their previously confiscated property and have taken ownership of various businesses. Therefore, in Slovakia the family business had to develop again from the beginning.

Family businesses represent the oldest form of business not only in Europe but also in our country. The reason that they are the object of our research is that they are long-term established companies that create jobs even in less developed regions without the interest of foreign investors and have a great perspective especially for small and medium-sized enterprises [30]. Their main advantage is usually less risk and greater stability due to conservative management and long-term sustainability. Family businesses in Slovakia have been facing problems for a long time; most often it is the changing legislation, a number of constraints, lack of skilled labor, low law enforcement, lack of advice, and assistance from public authorities [22].

During the survey, we found out what problems family businesses in Slovakia are facing at present. The main aim of the paper is to investigate the problems of Slovak family businesses. On the basis of the data obtained, we will propose options to eliminate the most serious problems that family businesses have been facing in our conditions for a long time.

2 Theoretical Background

Entrepreneurship itself can be characterized as a source of dynamic movement in the economy, and at the same time it forms the basis of a market economy, which is a living active organism. Without entrepreneurship, a market-oriented economy could not exist because it constitutes its absolute basis on which other economic operators and the national economy sector are linked. That is why business support should be at the center of interest not only for economists but, above all, for politicians, to create an appropriate and favorable environment for the development of entrepreneurial activities.

The priority role of each country's economy is to support the development of small and medium-sized enterprises, of which family businesses are also a part. Creating a suitable business environment, which means simplifying and clarifying the relevant legislation, reducing administrative, levy, and tax burdens, strengthening support infrastructure, and improving access to capital are the most important factors for its development. In the European Union, the share of SMEs is 99.8%, and approximately 2/3 of all employees are employed in this sector. Small and medium-sized enterprises are generally considered to carry a substantial part of the innovation impulses, and are also very flexible in the changing market environment and constitute an important segment for regional economic growth, writes Kachanek, Stalk, Bloch [19].

In particular, national governments should support, through appropriate instruments, the creation of new and retaining existing business units and support their continuous development conditional on further education, a favorable business environment, and available expert advice and tutoring. Support for local SMEs has undeniable advantages for all interested groups and the effort will undoubtedly pay off.

The drivers of a market economy in the 21st Century are businesses, mostly private-owned SMEs and family businesses. They are an indispensable pillar of the economy of each prosperous state and play an important role in the process of developing national economies. Their importance is also evident in the creation of the gross domestic product, on the one hand, and on the other, they have an important position in employment issues not only within the European Union but also on a global scale.

In Slovakia, the sector of small, medium-sized, and family businesses make up more than half of the business units, based on the findings of Kvašňák and Makarovičová [21]. This was the primary reason for which the government decided to start to support family businesses. These types of businesses help to reduce the unemployment rate in many regions suffering from disparities and the disinterest of larger business units.

In this context, there is a fundamental problem in Slovakia of how to approach the definition of a family business. It is not easy to define a family business because the Slovak legislature does not know this term, but in general, it is possible to speak of a family business if the business owner is a family member, the business is managed by a family member or the next generation of the founder Practice shows that in many cases family businesses are created as small and medium-sized enterprises.

There is no legal definition of the term 'family business' or 'family enterprise' in the Slovak legal order. Therefore, as stated by Mucha we can only rely on general legal definitions of the terms business, enterprise, and family [26]. The term 'company' does not have a uniform definition. The Commercial Code describes a company as a set of tangible as well as personal and intangible components of doing business.

Experts in the field such as, e.g. (Dudic et al.) say the definition of family businesses is based on different criteria than legal science [12]. According to them, first of all, small and family-run businesses have a relatively small market share and cannot influence the market significantly. According to Belas et al. the second feature is that they are managed by the owners (entrepreneurs), the families of the owners, not mediated through the formal management structure [8, 9]. The third feature is their independence as they do not form part of a larger enterprise.

To compare the differences in the definition, Zufan et al. lists several features of a family business that are applied in the US [39]. Primarily, the family has a major influence on the management and development of the business and its strategic goal is to pass the business on to the next generation. It is important that the family business is owned by the founder and his descendants. Anyakoha adds that the family business is a business of several generations [4]. This means in American practice that the family manages the business directly, the assets of the business are owned by the family and more than one member of the family is in business management.

Serina, however, points to the definition of the Massachusetts Mutual Life Company, which defines a family business by alternatively fulfilling at least one of three characters [31]. The owner considers his business to be a family business or intends to hand it over to a close relative or, in addition to it, another member of the family who is part of the day-to-day management process of the business works as a full-time employee.

However, in this context, we consider it necessary to properly include the concept of a family business in the legal order. We appreciate the legislator's intention to deal with such sensitive issues as the share of profits from the family business as well as other property rights from the operation of the family business, which creates additional property rights of the participating family members. However, Šétafy negatively assesses that these are disposable provisions of the Act which also allow for a different agreement between the participants [34].

2.1 The Nature of Family Business and Its Problems

The specificity of family business lies in the interaction of the family and the business environment. The founder of a family business is usually also a statutory body. We consider this to be the main reason why there is often an undesirable interaction between the harsh business climate and sensitive family ties. It is difficult to distinguish and separate relationships that are professional and emotional. The result of this clash is the need for an entrepreneur to address the issue of priorities between business and personal life. According to Zygmunt, however, every family has a natural endeavor to survive, especially if it is existentially dependent on the results of the business activities [38]. This, in comparison with “foreign employees”, motivates the family members to make generally greater efforts to achieve a successful outcome. Ślusarczyk and Haque point out that there should also be greater mutual trust among family members [33]. This assumption leads him to conclude that, for this reason, family businesses have longer-term stability and will survive several generations. The focus of family business is often the so-called family interest, which is created by involving multiple family members in the business. Such a strong family unit is, after all, a better prerequisite for the transfer of experience from generation to generation. The personal factor is not only a source of stability but also a tradition, which creates a strong basis for business continuity and its future prosperity [5].

One of the specifics of family businesses is, in particular, the flexibility to make decisions, giving priority to family members in managerial positions regardless of meeting the required criteria. Differences from other types of businesses, according to De Alvis can be also seen in aspects such as motivation, solidarity, and coherence [2]. These are related to the dependence of the family on the success of the business. According to Ključnikov et al. and Švec et al. is the key of success of every business, not just family business the exchange of information because we live in the information era [20, 35]. This information exchange is sincere and direct among family members, and moreover, without any speculation. Another important advantage of a family business is, according to Rahman et al. cost optimization [28]. It should be understood that family members are easier to agree on setting up substantial cost items, e.g. own wages.

In view of the fact that family businesses in the Slovak Republic are part of the category of small and middle-sized enterprises, their advantages also lie in the advantages that are characteristic of small and middle-sized enterprises. Among the most important, Nagy et al. incorporates a simple management structure, the possibility of increasing employment, innovation [27]. Meszároš and Divékyová add benefits like the ability to create self-employment, better knowledge of customer needs and the ability to address specific problems, but especially the use of regional labor, regional resources, contributing to reducing regional disparities [23].

Many family businesses use their own abilities and skills, making them more flexible to make decisions and responding more quickly to changes. From a broader perspective, they are, therefore, more stable. Their key advantage over large enterprises is their stability, as it is not their practice to move to other countries when state incentives are exhausted [36].

The change of the political system and transition to a market economy posed a serious problem for the older generation. Most employees, when they reach the age of 50, cannot find employment in the labor market if they lose their jobs because of employers' prejudices. From the point of view of several of the above-mentioned authors, it is possible to see the advantage of a family business in favor of a social care and security because the family business is not age-limited.

Family business and the related overlapping of family and making business can result in uncontrollable problems. As reported by Dudic et al. and Bure and Tengeh the problems mainly occur where there is no sincerity and mutual trust among family members [13, 10]. This results in the transfer of business conflicts to family life, and vice versa. These disputes may not only effect the business but may also result in its termination. That is why the literature recommends that the basic formal rules between individual family members should be defined before starting a business [32]. It is essential that the family is always able to name problems and solve them before they endanger the business. One possible solution, following the example of the Czech Civil Code, is to create a family council that would meet on a regular basis and solves problems before they cross the permissible limits.

Other problems in family businesses, according to Duľová et al. occur in the field of labor relations [14]. It is harder, sometimes impossible, to deduce the consequences for a breach of work discipline, failure to perform duties, or insufficient work performance towards family members compared to other employees. Our research has shown that only 15% of family businesses do not favor family members over other employees.

We also found out that up to 86% of family businesses do not address the issue of succession. Zajkovski and Domanska state that this is the most critical period of the existence of a family business [37]. Mura and Kaisar agree and state that there is also a breakthrough moment [25]. The founders usually try to delay this moment as much as possible in order to keep the business under control. The inability to leave the business and leave it to the successor is an insoluble problem for them. In the opinion of psychologists, there is a particularly inwardly suppressed fear of concern. International statistics show that only one-third of family businesses can handle generational exchange, with only a fraction of them remaining in the family's property for more than 50 years [29].

Haviernikova et al. are of the opinion that, in order to maintain business continuity, it is necessary to prepare for the transfer of succession well in advance and with the help of a psychologist [15]. In the absence of a suitable candidate, there are several possibilities. Selling a business is the easiest option. However, we consider ensuring a continuity of business management through a professional manager as a more suitable solution.

According to Ahmad et al. and Androniceanu et al. family businesses have several advantages, but also many problems [1, 3]. Difficult access to foreign sources of financing, especially credit, is increasingly plaguing family businesses. Business education is also problematic. Both the professional community and entrepreneurs have been calling for more time and space to be devoted to this issue in the educational process. There is also a need to increase spending on research and development of innovative technologies, including control of spending efficiency. Research results also show that the administrative burden is the biggest problem in the family businesses. In addition, Belas et al. see the problems in a lack of access to information, as well as a constantly changing business environment and poor availability of relevant information [7].

Before starting a business in the form of a family business, we recommend considering a number of facts, in particular the issue of coping with crisis situations. The disadvantages of family businesses can sometimes not be eliminated, but by using prevention these can be eliminated [16]. However, the benefits of family business can include a better form of cooperation, greater credibility in the workplace as well as customer awareness, easier acquisition or expansion of capital, flexible working time, family cohesion especially in difficult times, as well as informal relationships in the workplace [11].

3 Methodology

The theoretical definition of the issue was defined in the first part of this article. We are currently dealing with methodological issues. The aim of the article is to point out the problem areas of family businesses and the problems that Slovak family businesses encounter. At the same time, as a partial goal, we sought to find out, with the help of statistical analyzes, deeper links to specific issues of family businesses. In addition to the theoretical research of literary sources, we conducted a survey among Slovak family businesses to process this article. More than 400 enterprises were approached, the resulting sample consisted of 185 family enterprises willing to participate in the survey and providing the necessary information.

The methodology is an important part of the research work. Depending on the topic, we carefully selected possible methodological options. We used one of the most common methods of quantitative research - a survey with questionnaire.

The survey is a method often used in qualitative research because it provides a source of primary data and feedback, which is one of the most important components of the system. The advantage of this method is the possibility of analysis using various statistical software. On the contrary, the disadvantage is often insufficient sample size, failure to meet some prerequisites, respectively, poor sampling, sometimes the problem is the credibility of the respondent's answers.

If we want to describe the correlation between two variables, it is appropriate to use the so-called correlation coefficient having the following form:

$$r_{xy} = \frac{\overline{xy} - \bar{x}.\bar{y}}{\sqrt{\left(\overline{x^2} - \bar{x}^2\right)\left(\overline{y^2} - \bar{y}^2\right)}} \qquad (1)$$

“The correlation coefficient shall take the interval value ⟨-1,1⟩. If the value is positive, $r_{xy} > 0$, we indicate direct dependence, as the values of variable x increase, the values of variable y tend to grow. If the values of the correlation coefficient are negative, $r_{xy} < 0$, we indicate indirect dependence that is, when the values of the variable x increase, the values of the variable y tend to decrease. If $r_{xy} = 0$, both variables are linearly independent. These variables are said to be uncorrelated.”

In our case, we will consider significant values of at least a slight correlation, i.e. greater than 0.3 and less than -0.3 in the case of indirect correlation. A strong correlation can be considered one when the coefficient values are higher than 0.5 or below -0.5.

If we want to test statistical hypotheses, we need to use appropriate statistical methods, in this case we will perform nonparametric goodness o fit tests. Based on the nature of our research, we will use goodness of fit tests for variance and average.

Example of a statistical hypothesis:

H_O (zero hypothesis): the averages of the two statistical sets are the same,

H_1 (alternative hypothesis): the mean of the two statistical sets are not equal.

If the so-called p-value (determined by a statistical software) is less than the significance level α, which in our case will be 0.05, then the null hypothesis is rejected and the alternative hypothesis H_1 cannot be rejected. If the p-value is greater than the significance level, the null hypothesis cannot be rejected. The strength of statistical methods combined with reliable sampling is that the results are valid for the entire population.

In addition to the survey, hypothesis testing, elementary scientific methods (induction, deduction, analysis, synthesis), we also used t-test and regression models. These advanced methods have helped us identify a number of important facts.

4 Results and Discussion

The sample consisted of 185 family businesses willing to participate in the research. However, more than 400 family businesses were addressed. Indirect export (46 enterprises) was identified as the most used method of internationalization of entrepreneurial activity by family enterprises that participated in the research. The second most frequent way to enter international environment is a subsidiary abroad (33 companies). The research included family businesses, according to a well-defined definition of family business. These were the enterprises that the owner considered a family business or employed members of his family. The questionnaires were sent to companies and the questionnaire completion was dependent on the voluntary participation of employees in the research.

In descriptive statistics, we would like to give a description of the age structure of the sample. Based on the results of the variable age analysis presented below, it can be seen that the sample examined is skewed to the left, which means the most of the values are centered to the right of the average. Thus, on the basis of these results, we conclude that the number of older respondents prevails in the sample.

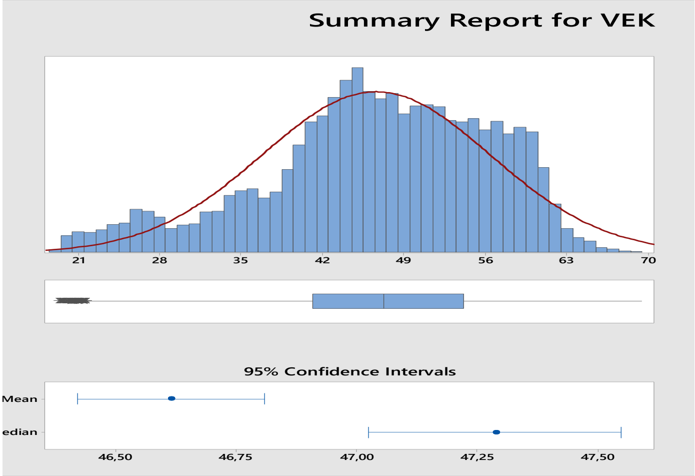


Figure 1

Characteristics of the sample structure

Concerning the distribution of the sample of respondents by age and gender, the average age of women is slightly higher than that of men. The difference between the first quartile is smaller in Figure 2.

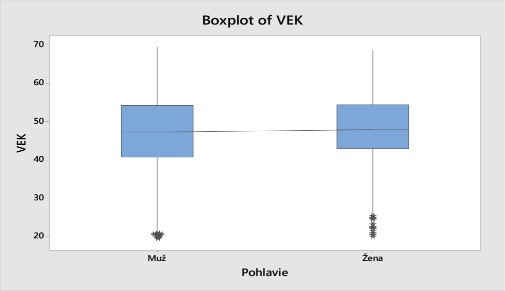


Figure 2

Age structure by gender

Within the analysis, we worked with two samples: family and non-family employees, women, and men. The variables we followed were age, education, and the involvement of entrepreneurship within a family business. By correlating age, the line of engagement, and education, we came to the conclusion that the younger generation of family employees had a higher incentive to do business in international markets, and this sample was also milder in identifying barriers to doing business. In the case of non-family employees in the companies we surveyed, the high dependency was found - the higher the level of education the respondents had, the lower were the barriers to international business. The most important motives of an entrepreneur to do business abroad was the lack of room for development in the home country followed by a wish to have higher sales.

In our research, we set out a few hypotheses that we verified. The student's t-test was used to verify them. The results were determined at the significance level of α = 0.05. The T-test was chosen based on the Fisher F-test result where the p-value was P> 0.05. By calculating the Student's t and comparing with the critical value, we came to the conclusion.

Hypothesis H1: "There is a statistically significant difference in the experience of working abroad between family and non-family employees."

There is no significant difference in experience from abroad. We were unable to confirm this hypothesis.

Table 1

Results of the T-test

T-test		
T-test values	t Stat	t crit (2)
	0.558463	2.014103

On this partial issue, we assumed that the work experience from abroad could help the Slovak family businesses. Experience from abroad, from a developed market economy, could contribute more effectively to the development of Slovak family businesses. This fact was not confirmed within our sample.

Hypothesis H2: "There is a statistically significant difference between family and non-family employees in that the family members are in the management of the enterprise."

There is a significant difference between the monitored groups and family members indeed showed significantly more managerial activities within the company. The hypothesis H2 was confirmed.

Table 2

Results of the T-test

T-test		
T-test values	t Stat	t crit (2)
	2.2907	**2.0141**

Hypothesis H3: “There is a statistically significant difference between family and non-family employees in the areas of management that led to the biggest business development.”

There is no significant difference between the variables in the identification of management areas that led to business development. Hypothesis H3 was not confirmed. On this sub-issue, we assumed that family members would approach managerial methods in different ways and practices. However, no significant difference was found within the areas of management that led to business development. We see the reason for this in several possible causes, such as the lack of gradual training, the lack of managerial experience of establishing the practice under specific conditions, the inability to exercise freely its managerial decision, etc. We have identified these reasons in the feedback of respondents.

Table 4

Results of the T-test

T-test		
T-test values	**t Stat**	**t crit (2)**
	1.5483	2.0141

The barriers to entry international markets were identified as financial demands, absence of a foreign partner, ignorance of the market, fear of an unfamiliar environment, and legislation. Further in our research, we focused on the innovation factors within the family business.

Among the innovation factors, we have included - product quality and services, customer relationship management, customer requirements, proactive approaches in marketing and employee qualification. We performed two regression models of multiple variables using R. Model 1 is shown in the following table.

Table 5

Model 1 Results

	Coefficient	*Std. Error*	*t-ratio*	*p-value*	
const	2,06621	0,0294836	70,08	<0,0001	***
Quality of products and services	0,0527984	0,0153012	3,451	0,0006	***
CRM	−0,131434	0,0221885	−5,924	<0,0001	***
Customer requirements	0,0154481	0,000556426	27,76	<0,0001	***
Proactive marketing approaches	−0,0360616	0,00124336	−29,00	<0,0001	***
Qualification of employees	−0,00194741	0,000973602	−2,000	0,0455	**

Mean dependent var	3,334289	S.D. dependent var	0,627007
Sum squared resid	2850,956	S.E. of regression	0,537040

The above-mentioned linear regression model constructed by the OLS method explains 26% of the variability of the dependent variable. The problem, however, is that the p-values for residual normality, heteroscedasticity and multicolinearity were less than 0.001, implying the rejection of the null hypotheses and thus the model assumptions were not met. For this reason, we made a correction of the heteroscedasticity model, which provided a more credible model without the presence of heteroscedasticity, multicolinearity and residual abnormality. The results are shown in the following table.

Table 6

Model 2 results

	Coefficient	*Std. Error*	*t-ratio*	*p-value*	
const	2,72620	0,0332163	82,07	<0,0001	***
Quality of products and services	0,0452121	0,0135721	3,331	0,0009	***
CRM	−0,179499	0,0190910	−9,402	<0,0001	***
Customer requirements	0,00731404	0,000569556	12,84	<0,0001	***
Qualification of employees	−0,0411669	0,00102319	−40,23	<0,0001	***
Proactive marketing approaches	−0,00012038	0,000799010	−0,1507	0,8802	

Statistics based on the weighted data:

Sum squared resid	33935,73		S.E. of regression	1,852850
R-squared	0,214326		Adjusted R-squared	0,213691

Statistics based on the original data:

Mean dependent var	3,334289		S.D. dependent var	0,627007
Sum squared resid	3020,611		S.E. of regression	0,552789

Since this model meets the prerequisites for OLS models and its overall p-value is less than the significance level, we reject the hypothesis that the model could be zero. Thus, it is statistically significant. The only statistically insignificant variable was proactive marketing approaches. For completeness, below is the picture for the Q-Q fence.

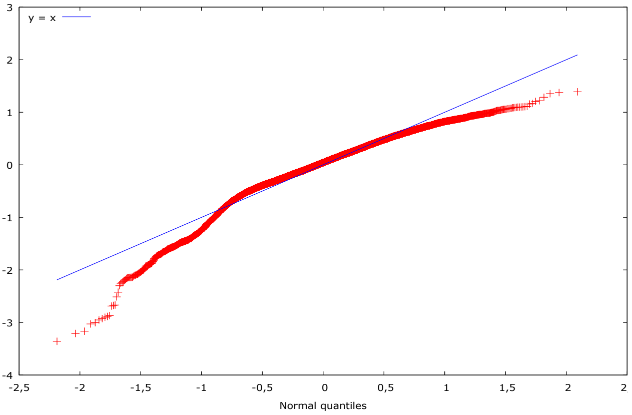


Figure 3
Q-Q plot for model 2

These results show that it is very important to apply strategic management decisions to a family business at the international level and that the elements of internationalization need to be maintained chronologically in order to be successful in foreign markets. The entrepreneur's priority is to place emphasis on the quality of the products produced, the quality of the services provided, the cultivation of fair customer relations, and the fulfillment of customer requirements.

Conclusions

Entrepreneurship is a priority for the market economy in terms of its functioning. The core of entrepreneurship in the business sector, which consists of business units. The vast majority of businesses are small and middle-sized enterprises, and in this group, a special form of business – a family business can be found. Family businesses are an important part of the economy not only of the European Union but also of Slovakia. They show a greater ability to create new jobs, which are related to their high sense of responsibility, tend to introduce new solutions for practice, bring innovative products and offer innovative services, provide personal input into mutual business relationships. These non-gas businesses are not typical. On the other hand, they also struggle with many, often specific problems.

This article discusses the important characteristics of family businesses, pointing out the burning issues that family business management has to address. In a sample of 185 family businesses, we analyzed several factors that we defined as significant. In particular, developmental and developmental factors hampering factors. We evaluated information from business units using quantitative methods. The results were presented numerically and graphically. In conclusion, we would like to point out important facts resulting from a well-known survey: the experience gained abroad does not contribute significantly to the development of

Slovak family businesses. Further in our research, we focused on the innovation factor within the family business, these factors were identified as quality of products and services, customer relationship management, customer requirements, proactive approaches in marketing and employee qualification. We constructed two regression models for multiple variables using the R program. Considering the outcome of the first model, we had to correct it for heteroscedasticity, which provided a more credible model without the presence of heteroscedasticity, multicollinearity, and residue abnormality. Since this model meets the prerequisites for OLS models and its overall p-value is less than the significance level, we reject the hypothesis that the model could be zero. Thus, it is statistically significant. The only statistically insignificant variable was only proactive marketing approaches.

Acknowledgement

This paper is part of the research carried out within the project VEGA 1/0813/19 Management of the development of innovative and start-up forms of business in the international business environment and verification of the INMARK concept.

References

[1] I. Ahmad, J. Olah, J. Popp, and D. Mate: Does Business Group Affiliation Matter for Superior Performance? Evidence from Pakistan, Sustainability, Vol. 10, issue 9, Article Nr. 3060, 2018, doi: 10.3390/su1009306

[2] C. De Alwis: Owner family and business succession in family owned companies, Acta Oeconomica Universitatis Selye, Vol. 5, issue 1, 2016, pp. 40-54

[3] A. Androniceanu, M. Comănescu and D. Jiroveanu: Factors with major influence on disparities across regions and their impact on economic development in Romania, Proceedings of the 29th International Business Information Management Association Conference - Education Excellence and Innovation Management through ision 2020: From Regional Development Sustainability to Global Economic Growth, Vienna, 2017, pp. 1743-1753

[4] C. Anyakoha: Job analysis as a tool for improved organizational performance of SMEs in Lagos, Nigeria, Central European Journal of Labour Law and Personnel Management, Vol. 2, issue 1, 2019, pp. 7-16, doi: 10.33382/cejllpm.2019.02.01

[5] C. Anyakoha: Strategic management practise and micro-small enterprises financial performance in Imo, South Eat Nigeria, Acta Oeconomica Universitatis Selye, Vol. 8, issue 1, 2019, pp. 41-52

[6] Z. Bayar, R. Remeikiene, A. Androniceanu, L. Gaspareniene and R. Jucevicius: The Shadow Economy, Human Development and Foreign Direct Investment Inflows, Journal of Competitiveness, Vol. 12, issue 1, 2020, pp. 5-21, doi:10.7441/joc.2020.01.01

[7] J. Belas, B. Gavurova and P. Toth: Impact of selected charakteristics of SMEs on the capital structure, Journal of Business Economics and Management, Vol. 19, issue 4, 2018, pp. 592-608

[8] J. Belas, E. Ivanova, Z. Rozsa and J. Schonfeld: Innovations in SME segment: importat factors and differences in the approach by size and age of the company, Transformations in Business and & Economics, Vol. 17, issue 3, 2018, pp. 55-71

[9] J. Belas, I. Kmecová and M. Cepel: Availability of human capital and the development of the public infrastructure in the context of business activities of SMEs, Administratie si Management Public, Vol. 2020, issue 34, 2020, pp. 27-44, doi: 10.24818/amp/2020.34-02

[10] M. Bure and R. K. Tengeh: Implementation of internal controls and the sustainability of SMEs in Harare in Zimbabwe, Entrepreneurship and Sustainability Issues, Vol. 7 issue 1, 2019, pp. 201-218

[11] G. Çera, M. Meço, E. Çera and S. Maloku: The effect of institutional constraints and business network on trust in government: an institutional perspective, Administratie si Management Public, Vol. 2019, issue 33, 2019, pp. 6-19, doi: 10.24818/amp/2019.33-01

[12] Z. Dudic, B. Dudic, M. Gregus, D. Novackova and I. Djakovic: The Innovativeness and Usage of the Balanced Scorecard Model in SMEs, Sustainability, Vol. 12, issue 8, Article Number 3221, 2020, doi: 10.3390/su12083221

[13] B. Dudic, Z. Dudic, J. Smolen and V. Mirkovic: Support for foreign direct investment inflows in Serbia, Economic Annals-XXI, Vol. 169, issue 1-2, 2018, pp. 4-11, doi:10.21003/ea.V169-01

[14] E. Dulová Spisakova, L. Mura, B. Gontkovicova and Z. Hajduova: R&D in the context of Europe 2020 in selected countries. Economic Computation and Economic Cybernetics Studies and Research, Vol. 51, issue 4, 2017 pp. 243-261

[15] K. Haviernikova, V. Snieska, V. Navickas and D. Burksaitiene: The attitudes of small and medium entrepreneurs toward cluster cooperation: the expectations and reality, Transformations in Business & Economics, Vol. 18, issue 3, 2019, pp. 191-205

[16] D. Halasi, P. Schwarcz, L. Mura and O. Rohacikova: The impact of EU support resources on business success of family-owned businesses, Potravinarstvo Slovak Journal of Food Sciences, Vol. 13, issue 1, 2019, pp. 191-205, doi:10.5219/1167

[17] J. Horecký and M. Blažek: Dependent work and internship, Central European Journal of Labour Law and Personnel Management, Vol. 2, issue 2, 2019, pp. 7-20, doi:10.33382/cejllpm.2019.03.01

[18] E. Ivanova: Barriers to the development of SMEs in the Slovak Republic, Oeconomia Copernicana, Vol. 8, issue 2, 2018, pp. 255-272

[19] N. Kachanek, G. Stalk and A. Bloch: What you can learn from family business, Harward Budiness Review, Boston, 2012

[20] A. Kljucnikov, L. Mura and D. Sklenar: Information security management in SMEs: factors of success, Entrepreneurship and Sustainability, Vol. 6, issue 4, 2019, pp. 2081-2094

[21] L. Kvašňák and X. Makarovičová: How to save a family business, Weekly TREND, online, cit. Oct. 2019, Analaible at: https://www.etrend.sk/trend-archiv/rok-2018/cislo-35/ako-zachranit-rodinne-firmy.html

[22] J. Lazikova, A. Bandlerova, O. Rohacikova and P. Schwarcz: Regional Disparities of Small and Medium Enterprises in Slovakia, Acta Polytechnica Hungarica, Vol. 15, issue 8, 2018, pp. 227-246

[23] M. Mészáros and K. Divékyová: Immediate termination of employment relationship by the employer, Central European Journal of Labour Law and Personnel Management, Vol. 2, issue 2, 2019, pp. 33-43, doi: 10.33382/cejllpm.2019.03.03

[24] L. Mura: Entrepreneurship internationalization – Case of Slovak family businesses, AD ALTA-Journal of Interdisciplinary Research, Vol. 9, issue 1, 2019, pp. 222-226

[25] L. Mura, and P. Kajzar: Small Businesses in Cultural Tourism in a Central European Country, Journal of Tourism and Services, Vol. 10, issue 19, 2019, pp. 40-54, doi:10.29036/jots.v10i19.110

[26] B. Mucha: Tools to increase the effectiveness of comprehensive management of emergencies affected by climate change in the Slovak republic, Proceedings of International Multidisciplinary Scientific GeoConference Surveying Geology and Mining Ecology Management, SGEM, Vol. 19, issue 5.4, 2019, pp. 573-580, doi: 10.5593/sgem2019/5.4/S23.075

[27] J. Nagy, J. Olah, E. Erdei, D. Mate and J. Popp: The Role and Impact of Industry 4.0 and the Internet of Things on the Business Strategy of the Value Chain—The Case of Hungary, Sustainability, Vol. 10, issue 10, Article Nr. 3491, 2018, doi:10.3390/su10103491

[28] A. Rahman, MT. Rahman and J. Belas: Determinants of SME Finance: Evidence from Three Central European Countries, Review of Economic Perspectives, Vol. 17, issue 3, 2017, pp. 263-285

[29] E. Rogalska: Multiple-criteria analysis of regional entrepreneurship conditions in Poland, Equilibrium - Quarterly Journal of Economics and Economic Policy, Vol. 13, issue 4, pp. 707-723, 2018, doi: 10.24136/eq.2018.034

[30] N. Selivanova-Fyodorova, V. Komarova, J. Lonska and I. Mietule: Differentiation of internal regions in the EU countries, Insights into Regional Development, Vol. 1, issue 4, 2019, pp. 370-384

[31] P. Serina: Family business in Slovakia, online, cit. 08. 10. 2019, Bratislava March 2011, avalaible at: http://www.sbagency.sk/sites/default/files/file/studia_rodinne_podnikanie_na_slovensku.pdf

[32] R. Suliková and N. Meyer: Motivating by flexibility: which role plays the company´s culture, Proceedings of the Managing global diversities Conference, 2018, pp. 41-50

[33] B. Ślusarczyk and B. Ul Haque: Public services for business environment: challenges for implementing Industry 4.0 in Polish and Canadian logistic enterprises, Administratie si Management Public, Vol. 2019, issue 33, 2019, pp. 57-76, doi:10.24818/amp/2019.33-04

[34] J. Šétafy: Family business needs a new modern definition also in Slovakia, online, cited 15. 11. 2019, Bratislava May 2014, avalaible at: https://www.sfa.sk/sk/novinky/detail/rodinne-podnikanie-potrebuje-aj-na-slovensku-novu-modernu-definiciu

[35] M. Svec, A. Madlenak and J. Horecky: GDPR and its impact on the direct marketing management, Proceedings of 15th Annual International Scientific Conference on Marketing Identity - Digital Mirrors Location, Book Series: Marketing Identity, 2018, pp. 344-353

[36] M. Švec and A. Madleňák: Legal frameworks for the phygital concept, European Journal of Science and Theology, Vol. 13, issue 6, 2017, pp. 209-217

[37] R. Zajkowski and A. Domańska: Differences in perception of regional pro-entrepreneurial policy: does obtaining support change a prospect?, Oeconomia Copernicana, Vol. 10, issue 2, 2019, pp. 359-384

[38] J. Zygmunt: Entrepreneurial activity drivers in the transition economies. Evidence from the Visegrad countries, Equilibrium - Quarterly Journal of Economics and Economic Policy, Vol. 13, issue 1, 2018, pp. 89-103

[39] J. Žufan, M. Civelek, I. Hamarneh and Ľ. Kmeco: The Impacts of Firm Characteristics on Social Media Usage Of SMEs: Evidence from the Czech Republic, International Journal of Entrepreneurial Knowledge, Vol. 8, issue 1, 2020, pp. 102-113, doi:10.37335/ijek.v8i1.111

Model Predictive Control for Automated Vehicle Steering

Ahmad Reda, Ahmed Bouzid, József Vásárhelyi

University of Miskolc, Institute of Automation and Info-communication,
Egyetemváros, 3515 Miskolc, Hungary
{autareda, qgebouzid, vajo}@uni-miskolc.hu

Abstract: The autonomous vehicle steering system, a multi-input multi-output (MIMO) system, is challenging to design using traditional controllers due to the interaction between inputs and outputs. If PID controllers are used the control loops are executed independently of each other as there is no interaction between the loops. Designing a larger system increases the controller parameters requiring tuning. Model Predictive Control (MPC) overcomes this problem, as it is a multi-variable control method taking into account the interactions of the variables in the target system. Achieving a high safety level is also critical for autonomous vehicle systems. This can be provided by an MPC controller, which can handle constraints such as maintaining a safe distance from other cars. Wider applicability of the Model Predictive Controller calls for more efficient hardware architectures for implementation. The aim of this paper is to achieve optimal implementation of the MPC controller by increasing the computational speed in order to reduce execution time for optimization. An MPC controller is used to control the steering system of an autonomous vehicle to keep it on the desired path. A traditional MPC controller is used to control the system where the plant dynamics do not change, whereas an Adaptive MPC controller is used when the system is nonlinear or its characteristics vary with time (the longitudinal velocity changes as the vehicle moves). Results are discussed in terms of performance, resource utilization, cost, and energy-effective implementations taking into consideration a reasonable size number of constraints handled by the controller.

Keywords: Autonomous Vehicle; Steering System; Model Predictive Control (MPC); Field Programmable Gate Array (FPGA); System on Chip (SOC)

1 Introduction

In recent years, research in the automotive industry has been growing in order to address the challenges of this application domain. Automotive control applications require high performance and cost reduction at the same time [5]. The control system requirements are becoming higher, and to achieve the improvement in control performance, the optimization process is incorporated into the control system design. The optimization process is subject to an increased number of factors, such as physical, safety, and economic constraints (power consumption, actuator saturation, etc.). In this context, Model Predictive Control (MPC) is a powerful optimization strategy for feedback control based on the model of the system. Basically an MPC controller runs a set of forecasts forward in time on the system model for different actuation strategies. MPC determines the immediate next control action based on the optimization. Next, it reinitializes the optimization in order to define the next control input [7]. The current and future control inputs are determined based on minimizing the difference between the target setpoint and the predicted output [13]. MPC features and capabilities are very effective in terms of meeting the requirements and achieving the optimization tasks. A basic MPC controller solves Linear Programming (LP) problems, which can be formulated as quadratic programming (QP) problem [12]. Also, the MPC controller has a natural capability to handle soft and hard constraints. That means, the requirements that are imposed by the operating conditions can be managed and formulated using the constraints. However, MPC controller implementation has several challenges such as high computational load and high power consumption, whereas the embedded system applications have limitations in their hardware resources.

One of the most effective solutions in order to achieve MPC implementations for embedded system applications which have constraints related to the computational time, is the use of the hardware acceleration. In this context, the deployments of an embedded MPC controller can achieve using reconfigurable hardware such as Field Programmable Gate Array (FPGA) or System on Chip (SoC), which is popular due to its high computational capabilities, parallel processing and development framework [11]. In this context, the main contributions of this paper are the study and the analysis of the efficiency of implementing control methods, in addition to the use of rapid prototyping methods (here hardware/software co-design using Embedded Coder and HDL Coder) for the implementation of embedded systems dedicated for digital signal processing considering performance, execution time and resources consumption. The research applied functional on-target rapid prototyping using Embedded Coder and HDL coder. The suggested implementation method is based on taking the optimization problem of the control method through MATLAB Simulink, Fixed-Point Designer, Embedded Coder and HDL coder. The suggested method allows the authors to focus on the verification, the validation and the test of the embedded system rather than programming, which in turn gives the ability to refine the design, tune the MPC controller parameters and see the results in the real-time. Finally, different optimization strategies were implemented and the obtained results were compared in terms of reducing the execution time and hardware resources consumption.

FPGA based systems have been applied for a variety of applications, such as image and signal processing, aerospace, energy, autonomous vehicles, telecommunications (5G) and medical field. In paper [1] an analytical study for Adaptive MPC controller under external disturbances signals was provided, the Lipschitz-based approach was used and provides satisfactory stability and robustness. Saragih et al. used the MPC controller for visual-based control system application (face tracking system) to control the motion of a robot, where the MPC controller was implemented to control the camera movements in order to keep the tracked face at the center of the camera – see [21]. Paper [4] provides an overview of a real-time optimization problem for automotive and aerospace applications with a focus on MPC controller. The optimal control problem was formulated based on the cost function and the system constraints, in addition, numerical algorithms and their implementations on an embedded computing platform were discussed. The improvement of fuel economy for power-split hybrid electric vehicles (HEV) was discussed in [2]. The energy management system was formulated as a nonlinear and constrained system. The MPC controller was used to split the power between the combustion engine and electrical machines at the different system operating conditions. The proposed approach provided an improvement compared to the controllers in commercial Powertrain System Toolkit (PSAT) software. The research reported in [8], proposed a control approach based on combining steering and braking MPC controllers. The authors in the paper introduced two model predictive controllers. The first one was implemented on a four-wheel vehicle model which determines the steering angle and braking torques to track the desired trajectory. The second MPC controller was implemented on a simplified bicycle model with a smaller number of inputs. The obtained results showed that the first controller provides good performance in terms of tracking the reference trajectory at low and high-speed, but the computation was time-consuming. On the other hand, the second controller showed unsatisfactory performance at high speed due to the simplicity of the vehicle model [8].

Paper [24] presented research of edge cloud on the Internet of Things (IoT) where the Model Predictive Controller evaluates the system properties. The paper presented the potential of merging the IoT, 5G, and cloud computing with the efficiency of deploying the automatic control system for time-sensitive and mission-critical processes. Haidegger et al. in [20] stated that the predictive and model-based control gives satisfactory performances only in the case of providing the accurate system's behavior and cascaded control approach. An empirical design with the use of Smith predictor for a telesurgical robot system was suggested in order to deal with the large latencies. In the same context of paper [20], the article [10] suggested a cascaded control structure to deal with the time delay in a teleoperation robot system. The suggested method used the extended Kessler's method sported by a predictive control method. Fuzzy–PID controller was also suggested to improve the performance. Using the extended Kessler's method with Smith predictor provides good control. MPC controller deals with linear-time-invariant (LTI) plant model, which allows predicting the future behavior of the system [22]. Nevertheless, paper [17] suggested a strategy to control heterogeneous traffic flow. Linear Parameter-Varying (LPV) model was suggested where the model deals with a non-linear traffic flow system which contains autonomous and human-driven vehicles with different operating conditions. LPV provides the ability to control the nonlinear system which uses different linear controllers for different operating points. LPV model uses a scheduling variable to enable the controller based on the current operating point of the system [6]. This paper discusses the use of an MPC controller for an autonomous vehicle steering system and its implementation using MATLAB Simulink and an FPGA board. The implementation on FPGA is conducted using HDL coder.

This paper is organized as follows: in this first section, a review of the MPC formulation, previous work, and literature are presented. The second section describes the plant (the vehicle) for which the controller was implemented. Section three describes the simulation and implementations. Section four presents the obtained results and analyzes the implementation. Finally, the conclusions are provided and directions for future work are suggested.

2 MPC and Adaptive MPC Working Principles

In a control problem, basically, the goal of the controller is to calculate the input variables to the plant so the plant responds in a way that makes its output track the reference output. Figure 1 shows the standard control loop diagram.

2.1 Model Predictive Controller (MPC)

Model Predictive Control (MPC) uses a future prediction strategy in order to calculate the input. To ensure that the output of the plant follows the target reference output, the MPC controller uses what is called an optimizer. The prediction strategy is based on the use of a plant model (car model) by the MPC controller to simulate the car's path in the next P time steps, where P is the prediction horizon which represents the time, the MPC controller looks forward in the future to make the prediction. The Model Predictive Controller simulated different future scenarios in a systematic way, and here the optimizer comes to the picture by determining the best scenario which achieves the minimum error between the reference and the predicted trajectory. The minimum error corresponds to the minimum cost function, which means the scenario of the predicted trajectory with the minimum cost function provides the optimal solution. Figure 2 shows the traditional MPC controller, and Figure 3 shows a future prediction strategy, where each scenario represents a series of steering wheel movements in order to follow the reference trajectory, and as mentioned above the optimal scenario is the one which achieves the minimum cost function.

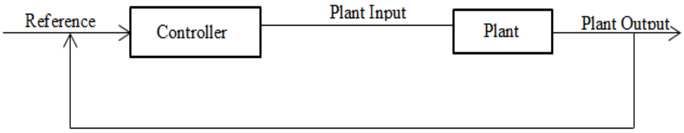


Figure 1
Standard Control Loop

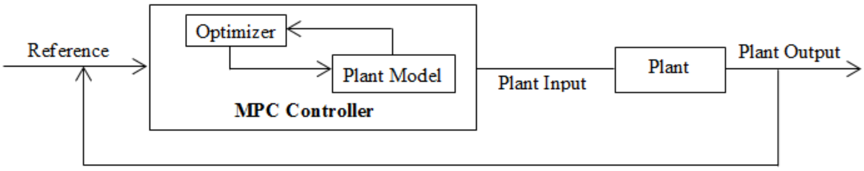


Figure 2
Traditional MPC control diagram

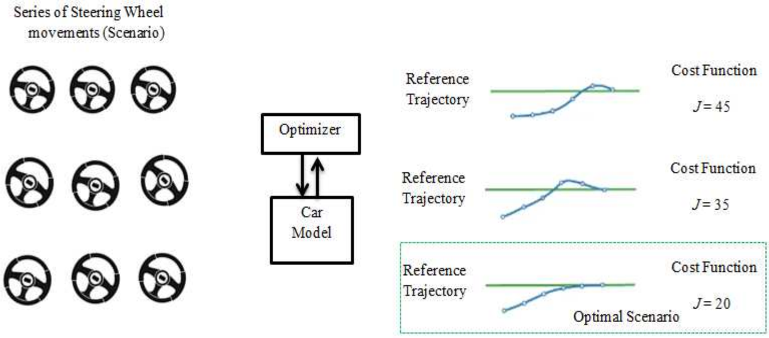


Figure 3
Future Prediction Strategy for Optimization Problem

The scenario with the minimum cost function J = 20 is the optimal solution, which achieves the optimal reference trajectory tracking.

The design presented in this article proposes that the new state of the car model can be measured, while in the case of the state model cannot be measured. The MPC controller uses the so-called “state estimator” to estimate the state of the system and feed it back to the controller. The MPC controller uses static Kalman Filter (KF) in order to update the controller states (plant model states, measurements noise model state and disturbance model state).

2.2 Adaptive Model Predictive Controller

The traditional MPC controller is unable to deal with the changing dynamics systems effectively since it uses a constant internal plant. When the system is nonlinear or its conditions vary with time, the accuracy will be negatively affected and the performance becomes unacceptable. To deal with these systems, an Adaptive MPC (AMPC) controller is used. AMPC controller handles the changes in operating conditions by providing a new linear model at each time step to achieve accurate prediction for the new conditions, as shown in Figure 4.

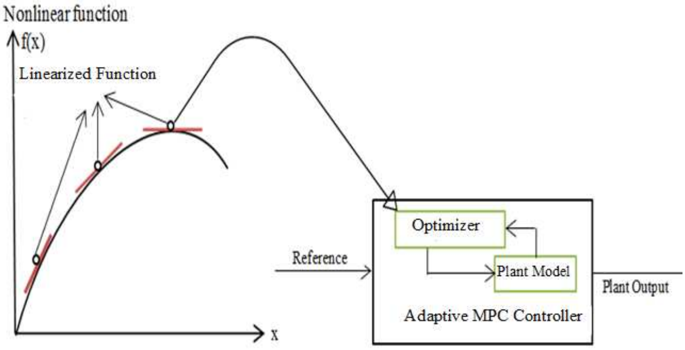


Figure 4
Adaptive MPC Controller [14]

The optimization problem in the Adaptive MPC controller remains the same, which means the same number of states and constraints for the varied operating conditions. The Adaptive MPC controller requires a discrete plant model, which means, the continuous-time state space needs to be converted to discrete-time (zero-order hold method). The Adaptive MPC Controller receives the updated discrete-time state space containing the following:

- A: n_x by n_x matrix signal, where n_x the number of plant model states.
- B: n_x by n_u matrix signal, where n_u the total number of plant inputs.
- DX: Vector signal of length n_x

$$DX = Ax_k + Bx_k - x_k \tag{1}$$

where DX is computed by equation (1), which provides the updated discrete-time state where u_k and x_k are respectively the inputs and the state values for the current time step k.

2.3 The optimization Problem

The MPC controller solves an online optimization problem, which is a Quadratic Problem (QP) for specific at each control interval. The optimization problem includes the followings:

Cost Function: also called objective function, it measures the controller performance, and the goal is to be minimized.

Constraints: It represents the soft and hard constraints which must satisfy the system conditions such as the physical bound.

To achieve the optimization, the MPC controller needs to calculate the control inputs driving the output of the plant that are very close to the desired reference. This process is performed in a systematic way by applying different scenarios and minimizing the cost function of the optimization problem. The cost function J of the autonomous vehicle's steering system can be formulated as:

$$\sum_{i=1}^{P} w_e e_{k+i}^2 \ \sum_{i=0}^{P-1} w_{\Delta u}\, \Delta u_{k+i}^2 \qquad (2)$$

where w_e is the weight of the predicted error e_{k+1} and $w_{\Delta u}$ is the weight of the steering angle increments Δu_{k+1}. Cost function goals are to minimize both, the error between the predicted trajectory and the reference and the change in the steering angle between the consecutive time steps. The optimal solution corresponds to the smallest value of the cost function.

Decision: Modify the manipulated variables in order to achieve the minimization of the cost function and to satisfy the constraints.

The MPC controller computes the manipulated variable by solving the quadratic problem using a custom QP solver which in turn converts the linear optimization problem to the general form of the QP problem. Figure 5 shows the control algorithm of the Model Predictive Controller.

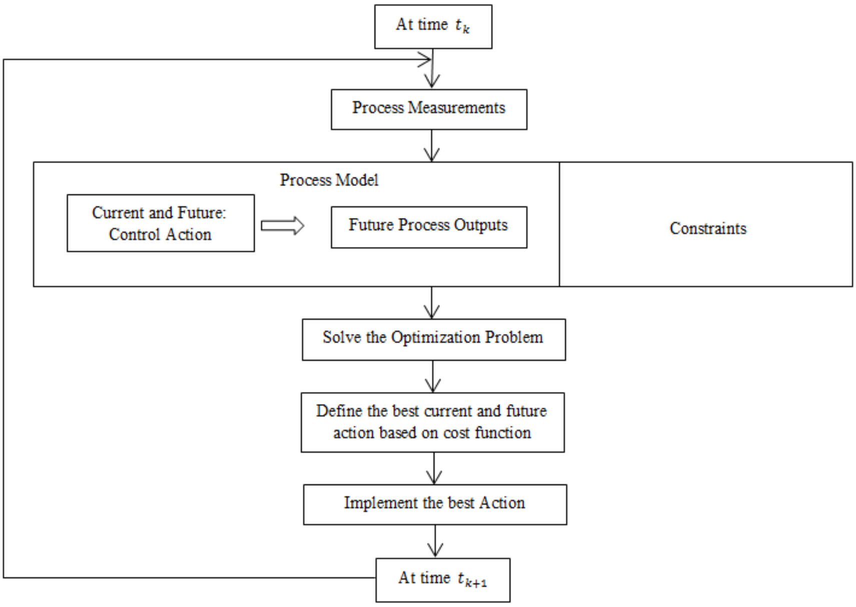

At time t_k
Process Measurements
Process Model
Current and Future: Control Action
Future Process Outputs
Constraints
Solve the Optimization Problem
Define the best current and future action based on cost function
Implement the best Action
At time t_{k+1}

Figure 5

MPC Control Algorithm

2.4 Model Predictive Controller Design Parameters

Designing the MPC controller takes into consideration the required constraints such as the steering angle limits. Figure 6 presents the main parameters and terms of the MPC controller, where the following nomenclature applies: k is the current sampling step and T_s the Control Time Step. Prediction horizon (P): number of time steps (the time on which the MPC controller looks forward to the future to make the prediction). Control Horizon (M): number of the possible control moves to time step $k+P$. The design parameters of the MPC controller are very important as this affects the performance and the computational complexity of solving the optimization problem. The choice of the design parameters should achieve the balance between the computational load and the performance. There are general recommendations, which can be taken into consideration for the parameters.

Sample time (T_S): determines the rate that the controller executes the control algorithm. In the case of Control Time Step T_s interval is too long, the controller will not be able to respond in time to the disturbance, which means that the performance will be negatively affected. On the other hand, if T_s is too short, the controller's response will be faster, but this causes a significant increase in computational load. The recommendation, in this case, is to choose T_s between 10 to 20 samples of the Rise Time T_r in an open-loop system, where T_r is the required time that the response takes to rise from 10 % to 90% of the steady-state as Figure 7 shows [15].

Prediction horizon (P): should be chosen in a way that covers the dynamic changes of the system and the recommendation are to choose P to have 20 to 30 of samples covering the open-loop transit system response [15], [18], [26] and [29].

Control Horizon (M): Only the two control moves have a significant impact on the response behavior, choosing a large control horizon will only increase the computation complexity, based on that, the recommendation is to choose M to be 10 to 20 of the prediction horizon. A small value of M provides stability while in contrast, large values reduce the robustness. It is recommended to choose M to be between 3-5 – as presented in [9], [15], [18], and [25].

For the model in this paper, the following strategy was used in order to choose the parameters which achieve satisfactory control performance: First, we initialized the parameters based on the recommendations above regarding the Sample Time, Prediction Horizon, and Control Horizon. Next step, is about tuning the parameters and then evaluating the MPC controller performance using the MPC Designer MATLAB toolbox until the optimal values provided the best control performance were determined. The weights of the inputs and outputs were determined using the MPC Designer by setting nonzero values to the inputs and outputs which need to track a reference value. Based on that, the weight equal is set to zero for the steering angle as it does not track a target. The weight of the

Lateral Position and Yaw angle were determined with nonzero values as the main objective is position tracing.

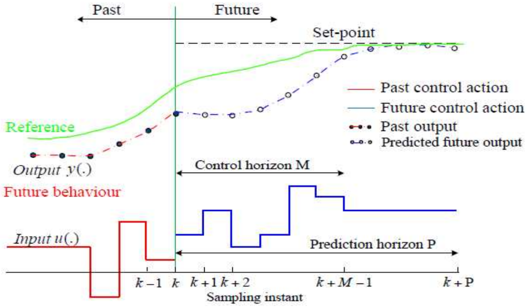


Figure 6

MPC schema for the main terms [27]

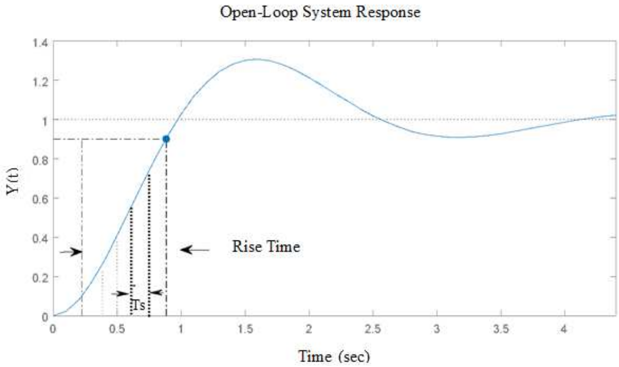


Figure 7

Control Time Step T_s and Rise Time T_r

3 The vehicle Model

MATLAB MPC designer application was used to design the controller that steers the vehicle autonomously. Figure 8 shows the global position of the vehicle in X and Y axes where (X, Y) are the vehicle's global position, v_y is the lateral velocity and v_x is the lateral longitudinal velocity. The parameters that need to be controlled are: Yaw angle Ψ and the front steering angle δ. The state-space of the model is given by the following equations:

$$\frac{d}{dt}\begin{pmatrix}\dot{y}\\ \psi\\ \dot{\psi}\end{pmatrix}=\begin{pmatrix}\frac{-2C_{af}+2C_{ar}}{mV_x} & 0 & -V_x-\frac{2C_{af}l_f-2C_{ar}l_r}{mV_x}\\ 0 & 0 & 1\\ \frac{2l_fC_{af}-2l_rC_{ar}}{I_zV_x} & 0 & \frac{-2{l_f}^2C_{af}+2{l_r}^2C_{ar}}{I_zV_x}\end{pmatrix}\begin{pmatrix}\dot{y}\\ \psi\\ \dot{\psi}\end{pmatrix}+\begin{pmatrix}\frac{2C_{af}}{m}\\ 0\\ \frac{2l_fC_{af}}{I_z}\end{pmatrix}\delta(2) \qquad (1)$$

$$\dot{y} = v_x\psi + \ v_y \qquad (2)$$

where v_x is longitudinal velocity at the center of gravity of the vehicle, m is the total mass of the vehicle, l_z is yaw moment of inertia of the vehicle, l_f and l_r are the longitudinal distance from the center of gravity to the front tires, C_{af} is cornering stiffness of tires and y is the lateral position.

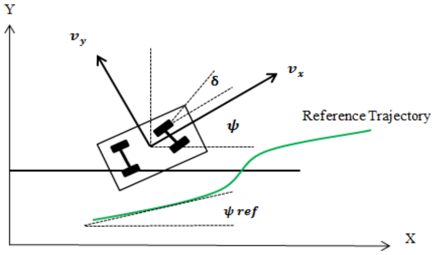


Figure 8
The global position of the vehicle

The MPC controller performs all the calculations using discrete-time state space. When a plant model is specified for the MPC controller, the following process needs to be performed [16]:

Conversion to state space: the model is converted to linear time invariant (LTI) state space model.

Discretization or resampling: in the case of difference sample time between the model and the MPC controller the following occurs:

- In the case of a continuous model, it must be converted to a discrete–time dynamic system model.
- In the case of the discrete model, the discrete-time dynamic system model is resampled in order to generate equivalent discrete–time model with a new sample Time T_S.

There are different ways to discretize a continuous model, in the proposed one, the continuous-time dynamic system model was discretized using zero–order hold on the inputs and sample time of T_S. This can be used also for resampling the discrete-time dynamic system model with new sample time T_S.

4 Design of the MPC Controller and HDL Code Generation

Based on the MPC control diagram the Simulink model was built. First, the required blocks (Plant model and Reference) were added to the workspace and linked to the MPC controller. The first input of the controller is the measured output and the second one is the reference trajectory, which was created using the Driving Scenario Designer Toolbox in MATLAB. As mentioned before, the MPC controller was designed using MPC Designer, where the internal plant model and the scenario are defined and the designing parameters such as sample time and control horizon were set using the strategy defined in section (2.4). In addition, the hard and soft constraints and their weights for the inputs and outputs such as the steering angle and the rate of change were set. In the case of an unchanging dynamics system, the input of the vehicle model is the output of the Model Predictive Controller (the steering angle) and the outputs are the lateral position and Yaw angle. Figure 9 presents the MPC controller model for linear systems (unchanging dynamics system). On the other hand, in the case of changing dynamics system, the longitudinal velocity is a second input for the vehicle model and the Adaptive MPC controller will use the plant mode output (State) to perform the new prediction for the updated model state. Figure 10 presents the Adaptive MPC controller model for nonlinear systems (changing dynamics system) with the Update Plant Model block.

Manual coding is time-consuming compared to the automatic code generation, which in turn lets the designers to focus on verification, validation and testing rather than programming. The model-based design generally provides an effective improvement in terms of system reliability and reduces the total project time up to 33% and the cost by 20% compared to the traditional methods (hand–written code) [23].

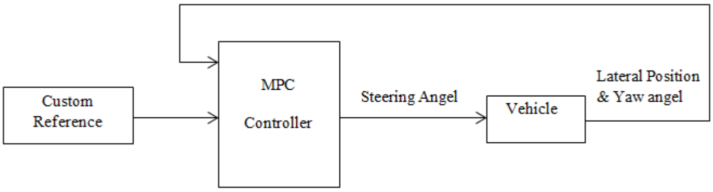


Figure 9

MPC controller model for linear system (Constant longitudinal velocity)

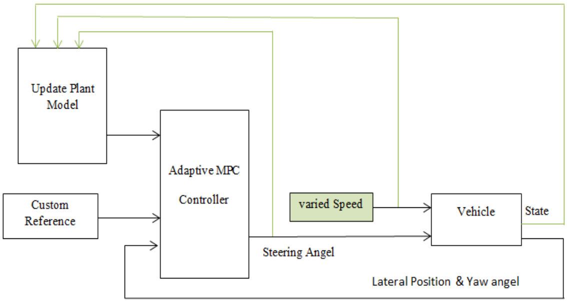


Figure 10
Adaptive MPC controller model for nonlinear system (varied longitudinal velocity)

The floating-point model needs to be converted to fixed point in order to reduce the hardware resources [19]. The steering system was designed and simulated using MATLAB Simulink and implemented on SoC (System on Chip) target using embedded coder and HDL coder. The working methodology is presented in Figure 11. First, the MPC controller model was created and the parameters were determined in MATLAB (see Table 1), followed by the HDL coder model and functional verification. Intellectual Property (IP) was created by Vivado. The MPC controller project was created and the MPC IP was connected to the Processing System (PS) through AXI interface. Figure 12 shows the block design of the MPC system.

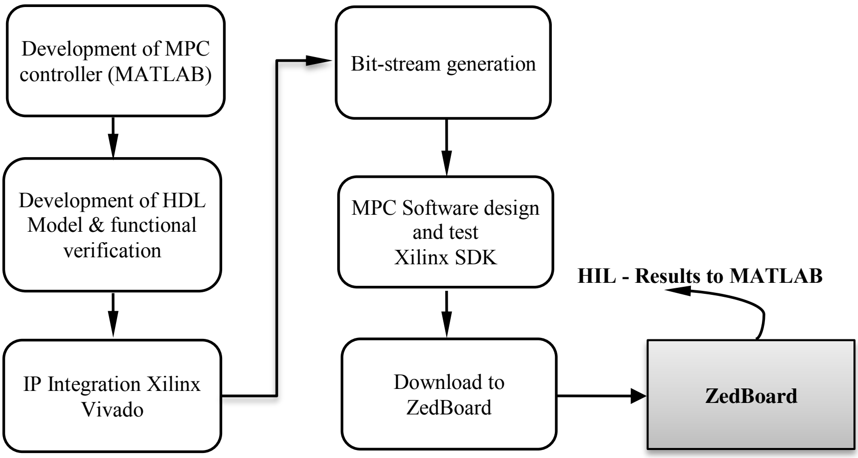

Development of MPC controller (MATLAB)
Development of HDL Model & functional verification
IP Integration Xilinx Vivado
Bit-stream generation
MPC Software design and test Xilinx SDK
Download to ZedBoard
HIL - Results to MATLAB
ZedBoard

Figure 11
The design workflow of the proposed solution [2]

Table 1

Values of the main MPC controller parameters and constraints

MPC Parameters	
Parameter	**Value**
Sample Time Ts	0.1 seconds
Prediction Horizon (P)	10 seconds
Control Horizon (M)	3 seconds
Constraints	
Steering Angle	[-0.5 - 0.5] rad
Steering Angle (changing rate)	[-0.26 - 0.26] rad

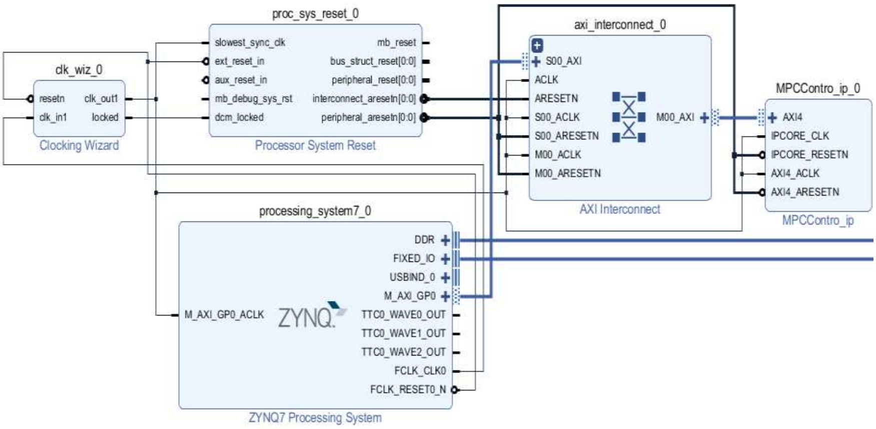


Figure 12

Vivado Block Design

The next step of the development (see Figure 11) was the bit-stream generation and export to the software development system (Xilinx SDK). The last step of the development was the software design and test. The generated project in Xilinx SDK together with the bit-stream downloaded and the target FPGA was programmed. In MATLAB Simulink the MPC model and MPC hardware system were tested and checked with Hardware In the Loop (HIL) simulation. The results are presented in the next section.

5 Simulation and Implementations Results

5.1 MATLAB Simulink Implementations

The steering system model was tested using MATLAB Simulink for both MPC and Adaptive MPC. Figure 13 shows the performance of MPC controller at a constant longitudinal velocity, and Figure 14 shows its performance at varied longitudinal velocity. The obtained results in Figure 13 and Figure 14 show that the MPC controller achieved satisfactory performance for the constant operating conditions, while it failed to handle the system with changing longitudinal velocity. Figure 15 shows the performance of the Adaptive MPC controller for the changing dynamic system (varied longitudinal velocity). Results demonstrate that using the Adaptive MPC controller for the changing dynamics system yields good performance in terms of tracking the reference (lateral position and yaw angle).

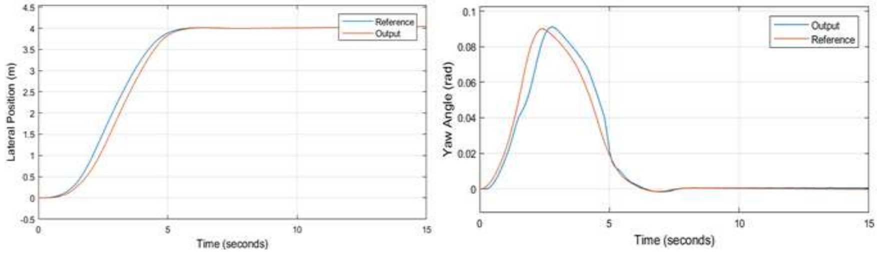


Figure 13

MPC controller performance at constant velocity

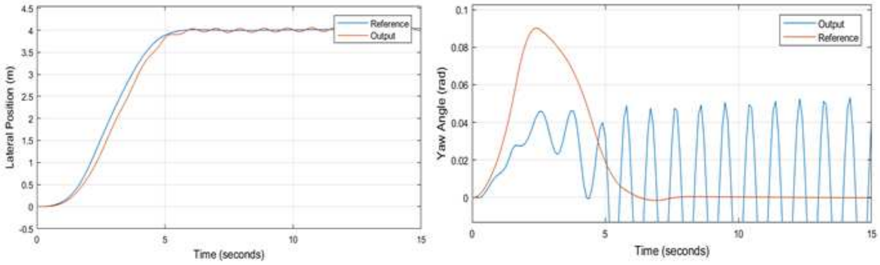


Figure 14

MPC controller performance at varied velocity

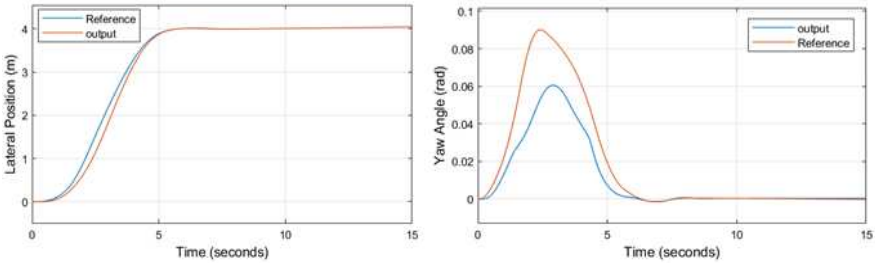


Figure 15

Adaptive MPC performance at varied velocity

5.2 FPGA Implementations

Both models (MPC and Adaptive MPC controller) were implemented on FPGA and the results were compared with the results obtained using MATLAB Simulink. The experiments showed slight differences in terms of performance between the implementations (Simulink and FPGA). Figure 16 and Figure 17 show the performance of the MPC controller at constant longitudinal velocity, and the performance of the Adaptive MPC controller at varied longitudinal velocity, respectively. Figure 18 and Figure 19 clearly show the difference in performance between the two controllers' implementation.

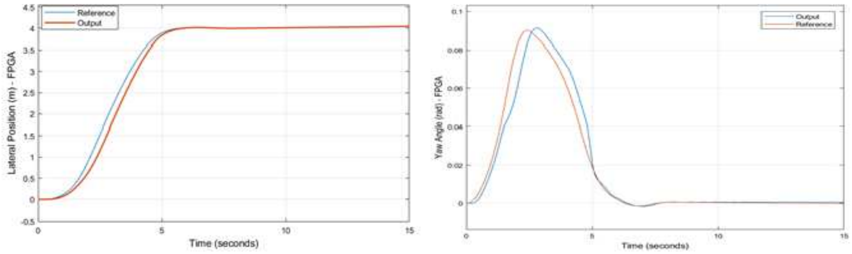


Figure 16

MPC controller performance at constant longitudinal velocity (FPGA)

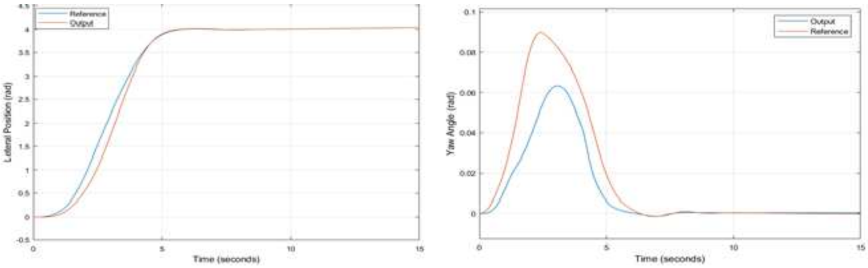


Figure 17

Adaptive MPC controller performance at varied longitudinal velocity (FPGA)

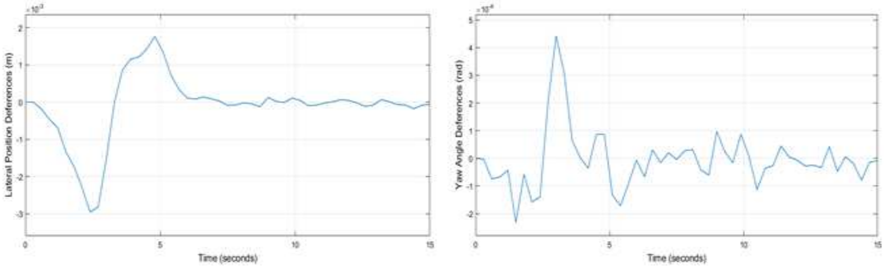


Figure 18

MPC implementation using Simulink and FPGA: Performance compression

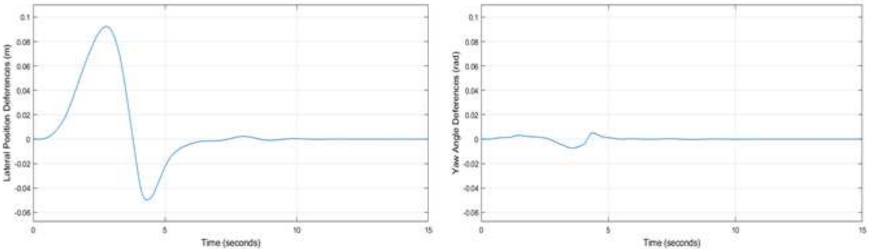


Figure 19

Adaptive MPC implementation using Simulink and FPGA: Performance compression

The implementations of MPC and Adaptive MPC controllers on FPGA were analyzed also in terms of resource utilization and power consumption using three different strategies for implementation to achieve the optimization as Table 2 and Table 3 show. In general, the implementations involve Logical optimization, placement of logic cells, and routing the connections between cells [28]. Implementation "Defaults strategy" balances runtime with trying to achieve timing closure. "Performance_ExplorePostRoutePhysOpt" strategy uses multiple algorithms for optimization, placement, and routing in order to get potentially better results. In "Flow_RuntimeOptimized" strategy, each implementation step trades design performance for a better run time [28].

Table 2

Resource utilization using different strategies

Defaults strategy					
Resource	Utilization		Available	Utilization %	
	MPC	Adaptive MPC	Adaptive MPC - MPC	MPC	Adaptive MPC
LUT	204	208	53200	0.38	0.39
FF	361	361	106400	0.34	0.34
BUFG	3	3	32	9.38	9.38

Performance_ExplorePostRoutePhysOpt strategy					
LUT	181	184	53200	0.34	0.35
FF	329	329	106400	0.31	0.31
BUFG	3	3	32	9.38	9.38
Flow_RuntimeOptimized strategy					
LUT	177	231	53200	0.33	0.43
FF	329	361	106400	0.31	0.34
BUFG	3	3	32	9.38	9.38

Table 3

Power consumption – different implementation strategies

Name	**Strategy**	**Total Power (W)**	
		MPC	Adaptive MPC
Impl_1	Implementation Defaults	1.791	1.791
Impl_2	Performance_ExplorePostRoutePhysOpt	1.792	1.792
Impl_3	Flow_RuntimeOptimized	1.793	1.791

Table 4

Power Consumption on chip - Summary

	Power Consumption	**Power on Chip**	
Dynamic	91%	Clocks	Less than 1%
		Signals	Less than 1%
		Logic	Less than 1%
		MMCM	6%
		PS7	91%
Static	9%	PL Static	100%

The results in Table 2 show that the implementation of the MPC controller on FPGA using the “defaults” strategy has the highest resource utilization, whereas the “Flow_RuntimeOptimized” strategy achieved the lowest resource utilization, where the utilization of LUTs (Lookup Tables) and FF (Flip-Flop) were reduced by 13.2% and 8.86% respectively. For BUFG (Global Buffer) there is no change. On the other hand. the implementation of MPC controller using “Performance_ExplorePostRoutePhysOpt” strategy achieved the lowest resource utilization. Table 3 shows that the power consumption for all applied strategies is almost the same.

Table 4 shows that 91% of the total power was used by the Processing System (PS), whereas only 9% was used by Programmable logic (PL) and only 6% of MMCM (Mixed-Mode Clock Manager) were used for both MPC and AMPC implementations.

Conclusions

This paper discussed the implementations of MPC and adaptive MPC controllers to control an autonomous vehicle steering system. The implementations were performed for both constant and changing dynamics systems. The models were implemented on FPGA using MATLAB HDL coder and different strategies were adopted to optimize resource utilization. The results showed that the MPC controller provides a satisfactory control for a constant dynamics system, but it couldn't handle operating conditions that are changing, while adaptive MPC provides good control for changing dynamics systems. In addition to analyzing the performance of the controllers, the implementations were discussed in terms of resource utilization and power consumption using different strategies.

The results showed a very slight improvement regarding the total power consumption. Based on the findings of this study, in future work, the implementations of MPC and adaptive MPC controller will be performed using System Generator in order to improve the power consumption and results will be compared with the results obtained in this paper.

Acknowledgement

This research was supported by the European Union and the Hungarian State, co-financed by the European Regional Development Fund in the framework of GINOP-2.3.4-15-2016-00004 project, aimed to promote the cooperation between the higher education and the industry.

References

[1] V. Adetola, M. Guay, Robust adaptive MPC for constrained uncertain nonlinear systems. International Journal of Adaptive Control and Signal Processing, 2011, pp. 155-167

[2] H. Borhan, A. Vahidi, A. Phillips, M. Kuang, I. Kolmanovsky, S. Di Cairano, MPC-Based Energy Management of a Power-Split Hybrid Electric Vehicle. IEEE Transactions on Control Systems Technology 20, 2012, pp. 593-603

[3] L. Crockett, D. Northcote, C. Ramsay, F. Robinson, R. Stewart, Exploring Zynq MPSoC: With PYNQ and Machine Learning Applications, https://www.zynq-mpsoc-book.com, UK, 2019, pp.

[4] S. Di Cairano, I. Kolmanovsky, Real-time optimization and model predictive control for aerospace and automotive applications. In: 2018 Annual American Control Conference (ACC), USA, 2018, pp. 2392-2409

[5] S. Di Cairano, I. Kolmanovsky, Automotive applications of model predictive control, Handbook of Model Predictive Control. Control Engineering, 2019, pp. 493-527

[6] Gy. Eigner, Control of physiological systems through linear parameter varying framework. Acta Polytechnica Hungarica, Vol. 14, No. 6, 2017, pp. 185-212

[7] P. Falcone, F. Borrelli, H. Tseng, J .Asgari, D. Hrovat, A Hierarchical Model Predictive Control Framework for Autonomous Ground Vehicles. American Control Conference, 2008, pp. 3719-3724

[8] P. Falcone, H. Tseng, F. Borrelli, J .Asgari, D. Hrovat, MPC-based yaw and lateral stabilization via active front steering and braking. Vehicle System Dynamics: International Journal of Vehicle Mechanics and Mobility, 2008, pp. 611-628

[9] J. Garriga and M. Soroush, Model predictive control tuning methods: A review. *Ind. Eng. Chem. Res.*, Vol. 49, No. 8, 2010, pp. 3505-3515

[10] T. Haidegger, L. Kovacs, S. Preitl, R. E. Precup, B. Benyo, Z. Benyo, Controller Design Solutions for Long Distance Telesurgical Applications. International Journal of Artificial Intelligence, Vol. 6, No. S11, 2011, pp. 48-71

[11] M. Lau, S. Yue, K. Ling, J. Maciejowski, A Comparison of Interior Point and Active Set Methods for FPGA Implementation of Model Predictive Control. Proceedings of European Control Conference, 2009. pp. 156-161

[12] K. Ling, B. Wu, J. Maciejowski, Embedded Model Predictive Control (MPC) using a FPGA. The International Federation of Automatic Control, 2008, pp. 1930-1935

[13] K. Ling, S. Yue, J. Maciejowski, A FPGA Implementation of Model Predictive Control. American Control Conference, 2006, pp. 15250-15255

[14] MathWorks ***, I. 2018. Linearize Nonlinear Models, URL: https://www.mathworks.com/help/slcontrol/ug/linearizing-nonlinear-models.html#responsive_offcanvas, Last accessed 16 March 2020

[15] MathWorks ***, I. 2018. Choose Sample Time and Horizons, URL: https://www.mathworks.com/help/releases/R2018a/mpc/ug/choosing-sample-time-and-horizons.html?s_eid=PSM_15028, Last accessed 16 March 2020

[16] MathWorks ***, I. 2018. MPC Modelling, URL: https://www.mathworks.com/help/mpc/gs/mpc-modeling.html, Last accessed 25 March 2020

[17] B. Németh, G. Péter, LPV design for the control of heterogeneous traffic flow with autonomous vehicles. Acta Polytechnica Hungarica, Vol. 16, No. 7, 2019, pp. 233-246

[18] Q. T. Nguyen, V. Veselý, D. Rosinová, Design of robust model predictive controller with input constraints. International Journal of Systems Science, Vol. 44, No. 5, 2013, pp. 896-907

[19] N. Othman, F. Mahmud, A. K. Mahamad, M. H. Jabbar, N. A Adon, Cardiac Excitation Modeling: HDL Coder Optimization towards FPGA stand-alone Implementation, In: 2014 IEEE International Conference on Control System, Computing and Engineering, 28-30 November, 2014, pp. 507-511

[20] T. Haidegger, L. Kovács, R. E. Precup, S. Preitl, B. Benyó, Z. Benyó, Cascade Control for Telerobotic Systems Serving Space Medicine. IFAC World Congress, Vol. 44, No. 1, 2011, pp. 3759-3764

[21] C. F. D. Saragih, F. M. T. R. Kinasih, C.Machhbub, P. H. Rusmin, A. S. Rohman, Visual Servo Application Using Model Predictive Control (MPC) Method on Pan-tilt Camera Platform. 6th International Conference on Instrumentation, Control, and Automation (ICA), August 2019, pp. 1-7

[22] M. Schetzen, Linear Time-Invariant Systems. John Wiley & Sons, New York, 2003

[23] Y. Siwakoti, G. Town, Design of FPGA-Controlled Power Electronics and Drives Using MATLAB Simulink. IEEE ECCE Asia Down under conference, 2013, pp. 571-577

[24] P. Skarin, W. Tärneberg, K. E. Årzen, M. Kihl, Towards Mission-Critical at the Edge and Over 5G. in 2018 IEEE International Conference on Edge Computing (EDGE), 2018, pp. 50-57

[25] S. E. Tuna, M. J. Messina and A. R. Teel, Shorter horizons for model predictive control. Proceeding of the 2006 American Control Conference, 2006, pp. 863-868

[26] K. Worthmann, Estimates of the prediction horizon length in MPC: A numerical case study, Proc. IFAC Conf. Nonlinear Model Predictive Control, 2012, pp. 232-237

[27] Y. Xiaoliang, L. Guorong, L. Anping, L. Van Dai, A Predictive Power Control Strategy for DFIGs Based on a Wind Energy Converter System, Energies, Vol. 10, No. 8, 1098, 2017, pp. 2-24

[28] Xilinx ***, Vivado design suite user guide: Implementation. UG904, v2016.2, 2016

[29] A. S. Yamashita, A. C. Zanin, D. Odloak, Tuning of model predictive control with multi-objective optimization. *Brazilian J. Chem. Eng.*, Vol. 33, No. 2, 2016, pp. 333-346

Analyzing the Relationship between Leadership Style and Corporate Social Responsibility in Hungarian Small and Medium-sized Enterprises

Peter Karácsony

J. Selye University, Department of Economics, Bratislavská cesta 3322, SK 94501 Komarno, Slovakia, e-mail address: karacsonyp@ujs.sk

Abstract: In recent decades, Corporate Social Responsibility (CSR) has become an important issue in the global business world. The topic has been studied from many viewpoints, but due to its continuous development and constant change, it remains a special and interesting area of scientific life to this day. The essence of CSR is that in addition to economic aspects, companies take into account the interests of society in their business and economic behavior. These behaviors can have many segments, so they can take into account their business partners, suppliers, employees, and the surrounding environment. Despite the fact that CSR literature has grown significantly since the turn of the millennium, researchers are mostly focused on the CSR activities of multinational companies, and there is limited research on SMEs. The relevance of the research topic is undoubtedly proven by the fact that many domestic and international researchers refer to the role of the leadership in CSR. Practical implementation basically depends on the behavior of the organization, and ultimately the attitude of the leader to CSR. In developing this study, I aimed to get to know the motivations, views, and attitudes of examined Hungarian small and medium-sized enterprise leaders regarding CSR. In my opinion, the companies that want to be successful, now and in the future, need to integrate CSR into their business strategy.

Keywords: leadership style; Corporate Social Responsibility; Hungary, employee-oriented leadership style

1 Introduction

More and more businesses are recognizing the need for CSR practices, and yet alongside this, they are trying to re-establish the irresponsible and short-term beneficial decisions that have led to global problems [10].

According to McWilliams and Siegel [32] CSR is defined as actions that acquiesce in the promotion of some social good, beyond the interests of the organizations and its shareholders and beyond what is required by law.

In my opinion, mostly the profit-oriented companies (Exxon Shipping Company, Union Carbide Corporation, Chisso Corporation, etc.) are causes of global social problems, for which the leaders have prime responsibility. The individual values that appear in the leaders' personality can play a decisive role in social responsibility. Thus, the examination of the impact of leaders' values on CSR was well founded. In my research, I would like to focus on the relationship between leadership and the practice of corporate social responsibility. My goal was to get to know the factors that influence leaders toward corporate social responsibilities.

The main benefit of my study is a new approach to analyzing the relationship between the style of leaders in small and medium-sized enterprises and CSR. According to literature, CSR activities have been approached mainly by ethical leadership ([3], [26], [44]) and transformational leadership ([17], [23], [47]), while, I analyze the impact of selected Hungarian small and medium-sized enterprise leaders on their CSR activities by the Michigan State University leadership model.

2 Literature Review

CSR has no precise definition to this day, probably because of the largeness of the topic. CSR is known under a number of names such as corporate responsibility, corporate accountability, corporate ethics, or corporate citizenship [15].

The CSR concept is often very positively labeled as ethical, honest and responsible behavior on the part of companies, and sometimes it is considered a "small green lie, a marketing trick, or a managerial boast" [38], which improves the reputation of companies.

CSR means "companies voluntarily incorporate social and environmental considerations into their business and their relationships with their partners". Its objectives include employment and social policy, environmental protection, consumer protection, governance and the sustainable development of these dimensions [4].

The first period of CSR's emergence can date to the 1950s. The rapid development of the economy and the growing demands of the consumer society brought the CSR concept. In 1953, Bowen interpreted terminology as a social obligation with issues of business responsibility [8].

The definition of CSR began in the 1950s and '60s, but initially, the focus was not on the responsibility of the company but on the responsibility of "businessmen".

Friedman [20] argues that companies are responsible only for increasing profits while spending money on social responsibility or environmental protection is just a money-wasting act. Another argument is the invisible hand theory. According to this, there is no need to intervene in the economy because the regulation of the 'invisible hand' creates the socio-economic optimum. Accordingly, in some Member States of the United States, until the early 1990s, companies were prohibited by law to give charitable donations to social organizations [37].

Today, Friedman's words have been disproved by the fact that CSR in the long term is now essential to ensuring organizational profitability because industrial accidents and corporate scandals have disrupted many companies' financial performance over the past decades. The former skepticism that companies cannot afford “charity„ has become obsolete for several reasons. The first reason is the behavior of a responsible corporate, which may show to be costlier in the short term but can be productive in the long run with a good company image. Secondly, the role of companies has unintentionally become increasingly important in society. They are responsible for the environment and everyday life, and because of this, they have to show themselves as trustworthy actors in every case [40].

Loew et al. [31] identify four CSR theoretical models. According to CSR1, companies, as a part of society, have responsibilities, and have certain obligations towards society when they use the resources of society. CSR2 or Corporate Social Responsiveness has been spreading since the 1970s and examines how a company can respond to societal challenges [1]. CSR3 is Corporate Social Rectitude, which means fair corporate conduct – and has been around since the 1980s. In essence, ethical considerations must be embedded in relevant corporate decision making processes [19]. Finally, CSR4 “Cosmos, Science, Religion” from the late 1990s highlights the role of individual companies, emphasizing the importance of science and its role in building social institutions.

From the turn of the millennium, CSR has become part of the corporate strategy while contributing to both business and social activities [25]. European companies are convinced that CSR-oriented organizations are gaining a competitive edge and will be at the forefront of the global economy in the future while reducing both their costs and environmental pressures [14]. CSR is not seen as a separate task but as a coordinated part of the corporate unit.

Many have studied the relationship between corporate performance and corporate social responsibility. Orlitzky et al. [35] noted that CSR not only reduces business risk, but it also increases the efficiency of internal resources and attracts a higher quality workforce.

According to Paine [36], the motivation behind the ethical behavior of companies can be negative or positive. It is negative if the company wants to avoid, prevent and avoid scandals, fines, costs, risks with CSR, and positive if the goal is to improve trust and reputation.

According to Carroll [12] although companies are responsible for the social problems, the damage and problems they cause are the responsibility of corporate leaders, and thus "social responsibility is not the commitment of a company, but of a person that takes into account the impact of their decisions and actions on the entire social system.

Carroll [11] describes the levels of responsibility of economic operators and companies in his famous pyramid model: economic responsibility means that economic operators must make a profit because they can meet other expectations of society only on a strong financial and economic basis. Compliance with the rules is based on, legal responsibility, but there is a higher level of ethical responsibility to consider and respect the interests of those concerned. Finally, Carroll identifies an exceptionally high-level, a human-friendly approach: philanthropic responsibility, which recognizes that the economic operator's ultimate goal is to allocate resources to increase social well-being and quality of life too. This responsibility is voluntarily assumed by the company and no one can be forced to do so.

Another popular interpretation is based on the stakeholder theory. According to this theory, the essence of CSR is that companies take into account the interests of stakeholder groups in addition to their corporate shareholders. In other words, companies have to find an ethical balance between the interests of the various groups [2]. The basic idea is that an organization needs to manage its relationship with the many stakeholder groups that are affected by its business decisions [13]. Social responsibility includes all the economic, legal and ethical actions undertaken by organizations to meet the expectations of all stakeholders [41].

Companies that do nothing to assume their social responsibility and claim that they gain no significant benefit from CSR because stakeholders do not value this type of activity are losers in today's fierce market competition [27].

Sood and Arora [42] argue that the motivation of companies for social responsibility activities depends on the organizations' leadership.

Nowadays, the importance of leadership has increased due to the continuously increasing competitive business environment [9]. According to Varney [45] leadership has been defined as an interpersonal process for influencing individuals and groups to achieve organizational goals. Leadership is a kind of power where one person has the ability to influence or change the behavior of another person [21].

Blake, Shephard, and Mouton [7] developed a two-factor model of leadership behavior, calling the factors 'concern for people' and 'concern for output'. From the 1960s, research into leadership began to analyze the effects of certain leadership behaviors, predominantly task-oriented and relationship-oriented leadership [5]. In the task-oriented leadership style, the leader focuses extensively on goal achievement and well-defined communication. In relationship-oriented leadership the leader gives strong emphasis to respect for their followers, looking out for their welfare, expressing appreciation, and providing emotional support.

Drucker [16] states briefly: “the only definition of a leader is someone who has followers.” It is about influencing a group of individuals. These group leaders direct their energies toward common goals [34].

Lewin [30] identified two main leadership styles, which are autocratic and democratic. In an autocratic leadership style, the leading power and decision making are centralized, the leader does not consult employees, and the motivation is a structured set of rewards and punishments. The democratic leadership style is also called the participative style, in which the leader encourages employees to take part in the decision making.

Despite it being a key topic there is a lack of theoretical and empirical research on the leadership aspect of CSR. In today's society, there is a growing interest in corporate leaders who are committed to CSR through actions that promote some social good, as required by company regulations and law [47].

The leaders’ motives for CSR can be divided into extrinsic and intrinsic motives. Extrinsic motives are aligned to business strategy and external pressures and are financially motivated, while intrinsic motives are morally, ethically, and philanthropically driven [22].

According to Waldman [48], ignoring the nature of the relationship between CSR and management is unfortunate, as a company's decision to take social responsibility must be considered at the strategic level.

According to Visser [46] the type of leadership required by CSR has elements in common with the contingency/interactionist leadership style, which is about the interaction between the individual leader and his framing context [18].

A CSR leader is someone who inspires followers and supports action towards a better world. According to Du et al. [17] transformational leaders are more effective in socially responsible organizations than transactional leaders. Transformational leadership increases follower motivation and performance more than transactional leadership [5]. Lazányi [28] argues that transformational leaders use emotions to motivate employees, communicate a vision, and stimulate followers to work towards long-term ideals and strategic objectives.

In summary, I can state that common in the CSR definitions is that all professionals pay particular attention to voluntary expression, as this activity is not prescribed by law, is not mandatory, but is a manifestation of ethical corporate behavior. Furthermore, based on the literature, I can say that it is worth exploring the topic in more detail and doing further research on the relationship between leadership and CSR. In the next part of my study, I want to introduce research methodology and the results of empirical research.

3 Methodology

The purpose of the present study is to identify if there is any correlation between leadership style and CSR practice in Hungarian small and medium-sized enterprises. The research is based on a survey conducted at 277 enterprises and uses data I collected through a self-completion questionnaire between 2018 and 2019. The sample includes respondents who manage SMEs (employing fewer than 250 people, with a turnover of less than EUR 50 million). All of the respondents hold significant leadership roles in the enterprise. Simple random sampling was used among the database of Hungarian small and medium-sized enterprises. Before my own research, pilot testing was conducted to help identify and change confusing, awkward, or offensive questions and techniques, thereby enhancing the validity and reliability of the research instruments. Feedback from the pilot test was generally agreed by the respondents that the questionnaire had been constructed in a clear way.

The questionnaire is divided into 2 sections. In the first one, the background information part is designed in order to collect demographic characteristics, such as gender, age, and education of respondents. The other part includes questions from the scales that were chosen to measure variables analyzed in the current study, namely leadership and CSR.

A total of 400 questionnaires was distributed, and 277 of those were returned completed, resulting in a response rate of 69.2%.

Participants were advised that the completion of the questionnaire should take no longer than 30 minutes. This was confirmed during the pilot study conducted prior to the main research.

All participants were advised that their participation was voluntary. Respondents were also assured that their own identity, together with the name of the organizations they work for, will remain confidential. It was explained to participants that the questionnaire is completely anonymous.

The analysis was performed by using single and multivariate statistical methods, in applying the SPSS program.

4 Results of Empirical Research

The research is based on three main pillars. In the first part, I examined the demographic characteristics of leaders and the main characteristics of evaluated enterprises. These results give an overview of the type of enterprises of the examined segment, its main activity, the market of its products/services, and the scope of its employees. Research on the demographics of leaders highlights the gender, education, and work experience proportion of the examined leaders. In the second part, the respondent evaluates the CSR activity of their enterprises using the 5-point Likert scales. This provides an opportunity to prioritize leaders in terms of CSR. In the third part, I was curious about the leadership style of examined leaders. Here, I examine the style of leaders and the impact of their attitudes toward corporate social responsibility.

In my primary research, I targeted Hungarian small and medium-sized enterprises because we can find in the scientific literature much less study than relating to multinational organizations. At the same time, small and medium-sized enterprises play an important role in economic life, in terms of employment and their role in the local economy. The small-sized organizations represent the majority (60.28%) of my sample, followed by medium-sized (24.18%) and micro-enterprises (15.52%).

Table 1 shows the main demographic characteristics of the respondents, 68.23% of respondents were male and 31.77 % were female. In terms of age, 32.85% of respondents declared themselves to be between 35 and 44 years, while 29.24% were over 45 years. The age group with the smallest representation in the sample (26 persons, 9.75%) was the group between 18 and 24 years of age. If we take a look at the educational background of the respondents, most of them, 51.99% have finished their secondary school education. The following group featured the respondents with a university qualification (38.63%). Those who finished primary school make up below 9.39% of the respondents and are mainly from the older age groups. It was found that 27.55% of the employees had 2 to 5 years of experience, while 11.91% had more than 10 years of work experience.

Table 1
Demographic characteristics of respondents

Age	**Frequency**	**Percent**
18-24	27	9,75
25-34	78	28,16
35-44	91	32,85
45+	81	29,24
Gender	**Frequency**	**Percent**
Male	189	68,23
Female	88	31,77
Education	**Frequency**	**Percent**
Primary	26	9,39
Secondary	144	51,99
University degree	107	38,63
Work experience	**Frequency**	**Percent**
0-2 year	66	23,83
2-5 year	104	37,55
6-10 year	74	26,71
more then 10 year	33	11,91

The sectoral proportion of the examined enterprises is shown in the following Table 2. The table above shows that services (20.58%), manufacturing (14.08%), and clothing (13.36%) sectors had the highest proportion among respondents.

Table 2

Sectoral proportion of examined SMEs, percentage

	Number of responses	**Percent**
Agriculture	11	3,97
Electric material	7	2,53
Trade	15	5,42
Wholesale	20	7,22
Construction	22	7,94
Restaurant	17	6,14
Transports and distribution	24	8,66
Clothing	37	13,36
Telecommunications	28	10,11
Services	57	20,58
Manufacturing	39	14,08

The CSR activity of the examined enterprises can be divided into 2 main sections, the inward and the outward CSR activity. The following Table 3 shows the main elements of the CSR activities of the evaluated small and medium-sized enterprises.

Table 3

Elements of CSR in examined SMEs

Inward CSR	**Percent**	**Outward CSR**	**Percent**
Protecting the health of employees	26.71	Sport sponsorship	14.80
Flexible, family-friendly workplace	22.38	Cultural sponsorship	22.38
Employee volunteer programs	13.36	Community Support	25.63
Employment of disadvantaged people	5.05	Support for environmental protection	24.19
Education, learning support	24.91	Support for health promotion	6.86
None of them	7.58	None of them	6.14

Most of the companies in the sample indicated in the first place protecting the health of their employees (26.71%) in terms of internal CSR activity, followed by the importance of education, learning support (24.91%). In outward CSR activity, the most performed by examined SMEs was community support (25.63%) followed by support for environmental protection (24.19%) and cultural sponsorship (22.38).

From the literature, I selected the task (work) oriented and employee (relationship) oriented leadership model for further analysis. Based on the literature ([29], [33], [39], [43]) these models are one of the frequently utilized leadership styles by small and medium-sized enterprise leaders. The main characteristics of this leadership model are: 1) Task-oriented behavior – stresses getting work done, followers are like tools that can be used to complete work and achieve goals. 2) Employee-oriented behavior – focus on the personal aspect of work where the leader looks at worker individuality and attends to each subordinate's personal needs [24].

In my research model, I indicated 8 leadership traits (Table 4) to determine which direction the examined leader represents.

Table 4

Main personality traits of task/employee-oriented leaders

Task-oriented leader	**Employee-oriented leader**
accuracy	communicative
objective driven	cooperative
proactive	inspiring
straight forward	sympathetic

The results of the personality characteristics of examined leaders are shown in Figure 1. Leaders marked on the Likert-scale from 1 to 4 their own personality traits (1-not very characteristic, 4-very characteristic). Based on the obtained results it can be stated that the leaders of the examined Hungarian small and medium-sized enterprises are mainly characterized by the communicative (3.77) personality trait, followed by the cooperative (3.49) and sympathetic (3.42) personality traits.

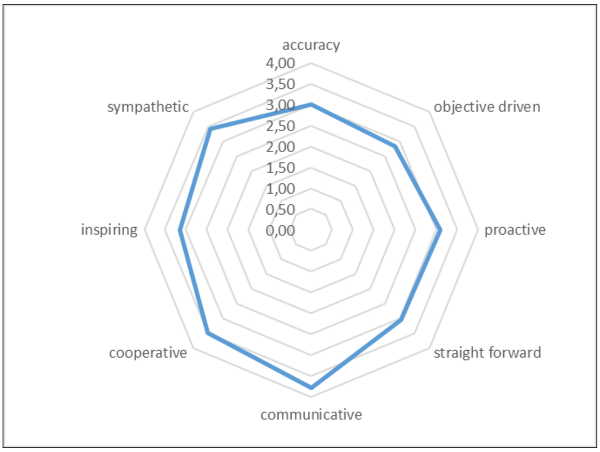


Figure 1

The main characteristics of the examined leaders, Likert-scale

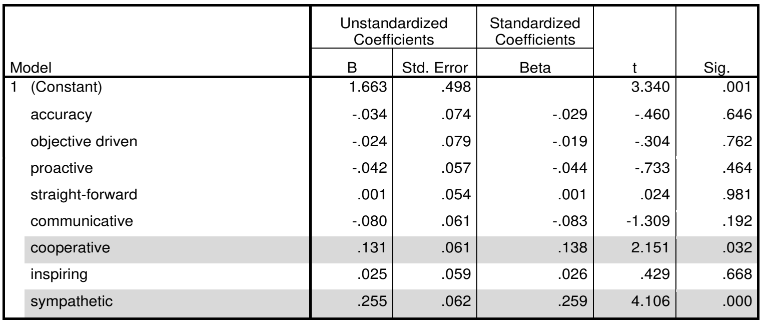
Based on the above results, the leaders of the examined Hungarian small and medium-sized enterprises mostly belonged to the employee-oriented leadership style. This is confirmed by Bass and Avolio [6] who described the concept of employee-oriented leadership behavior, which is measured by indicators like idealized attribution and behavior, individualized concern, intellectual stimulation, and inspirational motivation.

In my opinion, the personality of leaders has a significant impact on CSR activity in small and medium-sized enterprises. To prove this statement, I analyzed the personality traits of examined leaders relating to CSR activity.

Table 5 shows that two of the personal characteristics of the evaluated leaders show significant effects on their CSR activity. In the case of these 2 factors (cooperative and sympathetic), the significant value of less than 0.05. This statement also confirms the fact that leaders who are socially sensitive to people are more committed to CSR.

Table 5

Impact of a leader's personal characteristics on CSR activity



Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	1.663	.498		3.340	.001
	accuracy	-.034	.074	-.029	-.460	.646
	objective driven	-.024	.079	-.019	-.304	.762
	proactive	-.042	.057	-.044	-.733	.464
	straight-forward	.001	.054	.001	.024	.981
	communicative	-.080	.061	-.083	-1.309	.192
	cooperative	.131	.061	.138	2.151	.032
	inspiring	.025	.059	.026	.429	.668
	sympathetic	.255	.062	.259	4.106	.000

The results of the correlation analysis between the factors of task-oriented/employee-oriented leadership and CSR are shown in Table 6. This result also confirms that the employee-oriented leadership style has a significant relationship with CSR.

Table 6

Correlation between leadership styles and CSR

	Task-oriented leadership style	Employee-oriented leadership style
CSR	0.167	0.529

Correlation is significant at the 0.01 level (2-tailed)

The factors of employee-oriented leadership (0.529) have a strong positive correlation with CSR at a significance level of 0.000. This signifies that leaders who possess these attributes of employee-oriented leadership are more likely to stimulate their enterprises for CSR activity.

Regression analyses have been conducted to validate the hypothesis of my study: *The leadership style of leaders has a significant influence on CSR in small and medium-sized enterprises.*

Because the employee-oriented leadership style had a stronger correlation with CSR as a task-oriented leadership style, after this part I will evaluate deeply the employee-oriented leadership style and its impact on CSR to prove my hypothesis.

In the model the employee-oriented leadership was the independent variable and CSR the dependent variable. The regression analysis (Table 7) indicates that employee-oriented leadership has a considerable impact on the CSR activities of small and medium-sized enterprises.

Table 7

The result of regression analysis

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.529[a]	.280	.278	1.780	2.039

a. Predictors: (Constant), employee-oriented leadership

b. Dependent Variable: CSR

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	339.011	1	339.011	107.021	.000[b]
	Residual	871.119	275	3.168		
	Total	1210.130	276			

a. Dependent Variable: CSR

b. Predictors: (Constant), employee-oriented leadership

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2.042	.278		7.334	.000
	employee-oriented leadership	.553	.053	.529	10.345	.000

a. Dependent Variable: CSR

As the Model Summary section of Table 7 shows, the R value is 0.529. The R value represents the correlation value between employee-oriented leadership and CSR. The R Square value is 0.280 and the Adjusted R Square value is 0.278. Since the Adjusted R Square value is 0.278, it can be concluded that the independent variable accounts for nearly 28% of the variation in the dependent variable. The ANOVA section of Table 7 illustrates that the F value is 107.021. Since the F statistic is significant at 0.000 it can be said that leadership plays an important role in determining the CSR orientation of small and medium-sized enterprises. The Beta value is 0.529 at a significance level of 0.000. It indicates that leadership contributes significantly to CSR. The t value is 10.345 and the associated p value is 0.000. As the p value is smaller than 0.05 it can be concluded that the independent variable (leadership) reliability predicts the variation in the dependent variable (CSR) and the relationship between them is significant. This clearly indicates that leadership has a positive impact on the CSR initiatives undertaken by small and medium-sized enterprises and thereby supports my hypothesis.

One of the novel results of my study was that I approached the CSR activities of small and medium-sized enterprise leaders with the help of employee-oriented leadership style, which could provide new research opportunities for those interested in the topic. As there are few studies on the relationship between leadership and CSR, and even less dealing with small and medium-sized enterprises, my study can contribute to the literature of selected topic.

Conclusions

The results of my research confirm that CSR also plays an important role in the business life of Hungarian small and medium-sized enterprises. Based on the results obtained, it can be concluded that more than 90% of the surveyed companies had both outward and inward CSR activities. The corporate social responsibility of SMEs is expressed to stakeholders via environmental protection and community supports, while inward CSR primarily manifests itself in protecting the health of workers and in training and development of subordinates.

The activities of companies committed to social responsibility appear as a competitive advantage both in the short and long-term. I suggest to the leaders of the examined small and medium-sized enterprises that they should place more emphasis on the promotion of CSR activities in their corporate communication.

The regression analysis conducted on leadership and CSR indicates that leadership does have a positive impact on CSR. Of the examined SME leadership styles, the employee-orientation leadership style plays an important role in CSR activity. The results of my own research also highlighted the importance of leadership values for CSR. Most important are the leadership values that are manifested in behavior with subordinates. The responsible manager feels a sense of responsibility towards his subordinates and the "outside world" too. This is supported by Pless [37] who argues that the leaders' personal values are reflected in the practices that companies adopt.

An important factor in the dissemination of CSR is the personal commitment of the leaders, which is well illustrated by the example of the enterprises surveyed. Therefore, it is very important for leaders to prioritize a company's CSR activities, as opposed to their own individual interests, so that not only companies but also communities can benefit in the long run.

Even though small and medium-sized enterprises do not have that resources background that is available for leaders of large companies, SME leaders strive to conform to certain CSR values. Among companies, there are many small and medium-sized enterprises that try to run their business in a socially and environmentally responsible way. Even today, there are many factors that hinder the effective CSR activities of small and medium-sized enterprises. In order to have a successful and efficient corporate governance, it would be important that leaders explore the factors that hinder the implementation of CSR within their enterprises.

Competitive advantage as an opportunity stimulates active CSR of these small and medium-sized enterprises, which improves the reputation of the enterprise, as well as the acquisition and retention of employees.

In 2020, the COVID-19 crisis shocked the world. The crisis will certainly have an impact on the CSR activities of companies, too, especially of small and medium-sized enterprises, as they will have to deal with the retention of their employees in addition to the financial problems caused by the crisis. Thus, due to the COVID-19 crisis, companies have to re-evaluate the issues of both external (environmental subsidies, community programme supports, etc.) and internal (occupational health expenditures, employees' programmes, etc.) CSR activities. In my opinion, leaders of truly responsible companies should not change their basic business and moral attitude even during the crisis, because CSR can help enterprises to recover more quickly. As a result of responsible activities and appropriate communication, they can retain their business partners as well as their talented employees.

References

[1] Ackerman, R. W. and Bauer, R. A.: *Corporate Social Responsiveness*. Virginia, Reston Publishing, 1976

[2] Adda, G., Azigwe, J. B. and Awuni, R. A. (2016): Business ethics and corporate social responsibility for business success and growth. *European Journal of Business and Innovation Research,* Vol. 4, No. 6, 2016, pp. 26-42

[3] Adkins, S.: *Cause related marketing: Who cares wins*. Oxford. Butterworth Heinemann, 1999

[4] Andriof, J. and Waddock, S.: *Unfolding Stakeholder Thinking. Theory, Responsibility and Engagement.* Sheffield: Greenleaf Publishing, 2002, pp. 19-42

[5] Bass, B. M.: From Transactional to Transformational Leadership: Learning to share the Vision. *Organisational Dynamics*, Vol .18, No. 3, 1990, pp. 19-31

[6] Bass, B. M. and Avolio, B. J.: The Implication of Transactional and Transformational Leadership for Individual, Teams, and Organizational Development. *Research in Organizational Behavior*, Vol. 4, 1990, pp. 231-272

[7] Blake, R. R., Shepard, H. A. and Mouton, J. S.: *Managing intergroup conflict in industry*. Houston, Texas: Gulf Publishing Co., 1964

[8] Bowen, H. R.: *Social Responsibilities of the Businessman*. NewYork City, NY: Harper & Brothers, 1953

[9] Baranyai, Zs., Gyuricza, Cs. and Vasa, L.: Moral hazard problem and cooperation willingness: some experiences from Hungary. *Actual Problems of Economics*, Vol. 138, No. 12, 2012, pp. 301-310

[10] Buldybayeva, G.: Both Sides of CSR Practice: A Case from Oil and Gas Industry in Kazakhstan. *Acta Polytechnica Hungarica*. Vol. 11, No. 2, 2014, pp. 229-248

[11] Carroll, A. B.: The Pyramid of Corporate Social Responsibility: Toward the Moral Management of Organizational Stakeholders. *Business Horizons,* 1991, pp. 39-48

[12] Carroll, A. B.: Corporate social responsibility: Evolution of a definitional construct. Business and Society, Vol. 38, No. 3, 1999, pp. 268-95

[13] Clarkson, M. B. E.: A stakeholder framework for analyzing and evaluating corporate social performance. *Academy of Management Review,* Vol. 20, No. l, 1995, pp. 92-117

[14] Csehné Papp, I., Bilan, S. and Dajnoki, K.: Globalization of the labour market – Circular migration in Hungary. *Journal Of International Studies,* Vol. 12, No. 2, 2019, pp. 182-200

[15] Dahlsrud, A.: How Corporate Social Responsibility is Defined: an Analysis of 37 Definitions. *Corporate Social Responsibility and Environmental Management,* 15, 2008, pp. 1-13

[16] Drucker, P. F: *Managing the nonprofit organization.* HarperCollins Publishers, New York, 1990

[17] Du, S., Swaen, V., Lindgreen, A. and Sen, S.: The roles of leadership styles in corporate social responsibility. *Journal of Business Ethics*, Vol. 114, No. 1, 2012, pp. 155-169

[18] Fiedler, F. E.: *Leadership*. General Learning Press, 1971

[19] Frederick, W. C.: Moving to CSR4. *Business & Society,* Vol. 37, No. 1, 1998, pp. 40-60

[20] Friedman, M:. *Capitalism and Freedom.* University of Chicago Press, Chicago, 1962

[21] Ganta, V. C. and Manukonda, J. K.: Leadership During Change And Uncertainty In Organizations. *International Journal of Organizational Behaviour & Management Perspectives*, Vol. 3, No. 3., 2014, pp. 1183

[22] Graafland, J. J. and Ven van de, B.: *Strategic and moral motivation for corporate social responsibility.* Munich Personal RePEc Archive, Working paper no 20278. Tilburg University, Netherlands, 2006

[23] Groves, K. S. and LaRocca, M.: An empirical study of leader ethical values, transformational and transactional leadership, and follower attitudes toward corporate social responsibility. Journal of Business Ethics, Vol. 103, No. 4., 2011, pp. 511-528

[24] Johns H. E. and Moser, H. R.: From trait to transformation: The evolution of leadership theories. *Education*, Vol. 110, No. 1, 2001, pp. 115-122

[25] Karácsony, P.: The Role of Corporate Social Responsibility in Environmental Sustainability. In: Behnassi, M.; Gupta, H.; Pollmann, O.: *Human and Environmental Security in the Era of Global Risks*. Springer International Publishing, 2019, pp. 377-386

[26] Kim, M-S. and Thapa, B.: Relationship of Ethical Leadership, Corporate Social Responsibility and Organizational Performance. *Sustainability.* Vol. 10, 2018, pp. 1-16

[27] Kotler, P. and Lee, N.: *Corporate social responsibility. To do good for a cause and for the company*. HVG Kiadó, Budapest, 2007

[28] Lazányi, K.: The role of leaders' emotions. *Applied Studies in Agribusiness and Commerce. APSTRACT.* 2009, pp. 103-108

[29] Leitch, M. C., McMullan, C. and Harrison, T. R.: Leadership development in SMEs: an action learning approach. *Action Learning: Research and Practice*, Vol. 1, No. 3, 2009, pp. 243-263

[30] Lewin, K.: *The consequences of an authoritarian and democratic leadership.* In A. W. Gouldner, Studies in leadership: Leadership and democratic action. New York: Russell & Russell, 1965

[31] Loew, T. Ankele, K., Braun, S. and Clausen, J.: *Significance of the CSR debate for sustainability and the requirements for companies.* Münster/Berlin, 2004

[32] McWilliams, A. and Siegel, D.: Corporate social responsibility and financial performance: correlation or misspecification? *Strategic Management Journal*, Vol. 21, 2000, pp. 603-609

[33] Morrison, A.: SME management and leadership development: Market re-orientation. *Journal of Management Development,* Vol. 22, No. 9, 2003, pp. 796-808

[34] Northouse, P.: *Leadership. Theory and practice.* (5th ed.) London: SAGE Publications, 2010

[35] Orlitzky, M., Schmidt F. L. and Rynes, S. L.: Corporate Social and Financial Performance: A Meta analysis. *Organization Studies,* Vol. 24, No. 3, 2003, pp. 403-441

[36] Paine, L. S.: Does Ethics Pay? *Business Ethics Quarterly*, Vol. 10, No. 1, 2000, pp. 319-330

[37] Peattie, K.: *Green marketing*. London: Pitman, 1992

[38] Piercy, N. F. and Lane, N.: Corporate Social Responsibility: Impacts on Strategic Marketing and Customer Value. *The Marketing Review*, 9, 2009, pp. 335-360

[39] Pless, N. M.: Understanding Responsible Leadership: Role Identity and Motivational Drivers. *Journal of Business Ethics*, Vol. 74, No. 4, 2007, pp. 437-456

[40] Porter, M. E. and Kramer, M. R.: *The Competitive Advantage of Corporate Philantropy*. Harvard Business Review, December, 2002

[41] Schwartz, M. S. and Carroll, A. B.: Corporate Social Responsibility: A three domain approach. *Business Ethics Quarterly*, Vol. 13, No. 4, 2003, pp. 503-530

[42] Sood, A. and Arora, B.: *The political economy of corporate responsibility in India.* United Nations Research Institute for Social Development UNRISD., 2006

[43] Soriano, R. D. and Martinez, M. C. J.: Transmitting the entrepreneurial spirit to the work team in SMEs: the importance of leadership. *Management Decision*. Vol. 45, No. 7, 2007, pp. 1102-1122

[44] Tourigny, L., Han, J., Baba, V. V. and Pan, Y.: Ethical Leadership and Corporate Social Responsibility in China: A Multilevel Study of Their Effects on Trust and Organizational Citizenship Behavior. *Journal of Business Ethics.* Vol. 158, No. 4, 2019

[45] Varney, S.: Leadership learning: key to organizational transformation. *Strategic HR Review*, Vol. 7, No. 4, 2008, pp. 5-10

[46] Visser, W.: Corporate citizenship in South Africa. *Journal of Corporate Citizenship*, Vol. 18, 2005, pp. 29-38

[47] Waldman, D.: Defining the socially responsible leader. *The Leadership Quarterly*, Vol. 19, No. 1, 2008, pp. 117-131

[48] Waldman, D. A., Siegel, D. S. snd Javidan, M.: Components of CEO transformational leadership and corporate social responsibility. *Journal of Management Studies*, Vol. 43, No. 8, 2006, pp. 1703-1725

Exploitation vs. Prevention: The Ongoing Saga of Software Vulnerabilities

László Erdődi, Audun Jøsang

University of Oslo, Gaustadalléen 23 b, 0371 Oslo, Norway
laszloe@ifi.uio.no, josang@ifi.uio.no

Abstract: Online IT systems are frequently exposed to cyber-attacks. An Exploit is an advanced attack tool that takes advantage of some software vulnerability to attack and cause harm to IT infrastructures. Developers and manufacturers of operating systems and hardware put huge effort into the prevention of vulnerability exploitation (e.g. Data Execution Prevention, Control Flow Integrity, etc.). However, the number and severity of attacks show that new exploit methods are continuously being invented despite the increasingly sophisticated protection methods. The present article summarizes the current, known and most relevant software vulnerability exploitation methods, as well as, the possible methods used to protect against these exploits. Moreover, the effectiveness of both the exploitation and prevention methods (as seen from both the attacker's and the defender's sides) is analyzed to find a possible future direction, to eliminate exploit attacks against an IT infrastructure.

Keywords: vulnerability; exploitation; protection; control-flow

1 Introduction

Software coding errors can become vulnerabilities that can allow malicious exploits to take control over computer systems. Using deliberately malformed input data attackers can cause unintended or unanticipated behaviors in a software package that contains a particular type of vulnerability. Depending on the type of vulnerability an exploit can be a sequence of commands, a chunk of data or a piece of software to cause malicious code execution for the sake of the attackers. Exploits can be categorized according to their capability (e.g. remote code execution, DOS), the platform they can be applied to (e.g. Windows, Linux, IoS, etc.) and also according to the way of execution (local, remote). Some websites allow the public to register known exploits, such as, the exploit database [1], where users can submit ready-to-use exploits. Exploitalert [2] is another website that reports exploits with detailed data found on the Internet. Another exploit collection is the Metasploit framework [3] which contains several exploits in a unified form which makes the exploitation very easy and automatic for the attackers.

Figure 1 shows the number of the available exploits over the years, according to the Exploit database [1]. Even if this figure and the available sources do not contain all the existing exploits, it is nevertheless, interesting to observe the trend. The number of new exploits was on the top in December 2009 when nearly 600 new exploits were added during one month. The Data Execution Prevention (DEP) [4] and the Address Space Layout Randomization (ASLR) [5] became basic feature of operating systems around that time, which can explain the significant decrease in the number of new exploits after 2009. Another reason for the decrease can be the appearance of the dark web.

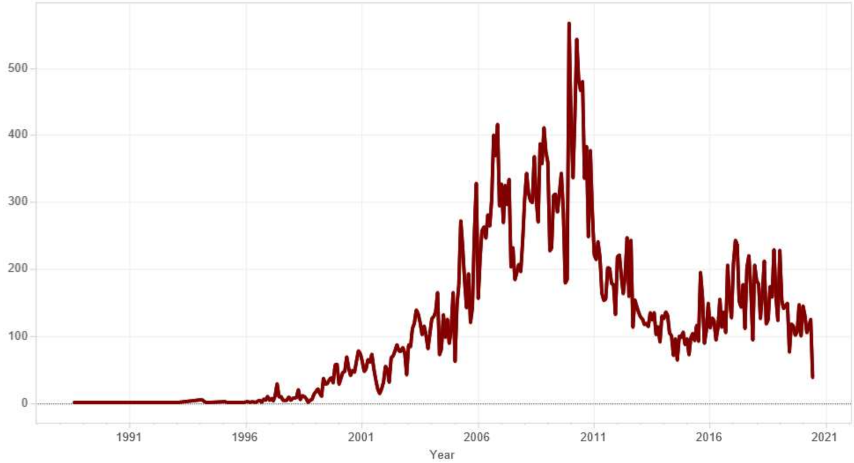


Figure 1

Number of recorded new exploits per month in the exploit database [1]

An exploit is usually able to take advantage of one particular vulnerability in a particular piece of software, but there are some exceptions. A general exploit can affect multiple platforms as it customizes itself for the actual version of the software. Some exploits use two or more different vulnerabilities at the same time to achieve their goals [6]. For a modern web browser exploitation, sometimes three different vulnerabilities are necessary: one for obtaining the ASLR randomization offset, one for exploiting the vulnerability and a third one to break out from sandboxing.

From a vulnerability point of view, two major categories can be created according to our categorization: The configuration error based and the software error-based exploits. The exploit that takes advantage of a configuration error can use e.g. default passwords, access hidden content or bypass protections by misusing the system. In all of these cases the vulnerability is connected to inappropriate configuration. In this paper we focus on the other case when the configuration is correct, but the software code contains vulnerability. Since we use different software layers that are based on each other the bug can be on different levels too. The level of the bug significantly determines the difficulties of the detection and protection possibilities. For example, a Content Management System (CMS) uses a kind of server side scripting code which is executed by the webserver software of the operating system. The web server software uses the operating system API which is based on the kernel level code of the operating system. So a bug in a php code, on the CMS level, has different effect than a bug in a kernel driver. Figure 2 shows the different layers.

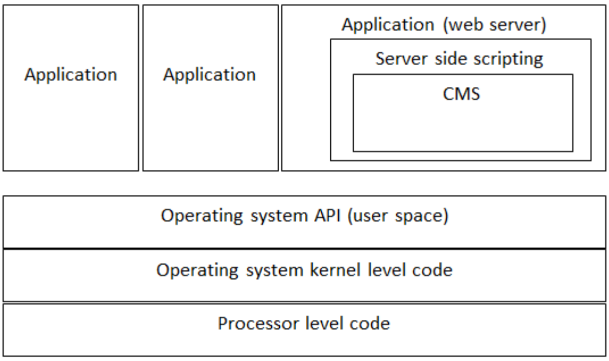


Figure 2
Software code levels

If the vulnerability is e.g. in a kernel driver, then the exploit has the system right to execute the malicious code. In the user space the exploits have the same right as the application that contains the vulnerability. In these cases, e.g. a crafted PDF file is the exploit itself that is opened by the PDF reader (application). If the application provides services, then the attack surface will be increased. In the case of a web server application the vulnerability can be inside the application code or in the high level server side code (e.g. php based SQL injection). In other cases, the Content Management System (CMS) contains the vulnerable server side code (e.g. Drupal SQL injection [7]). Exploits can be created in all of these cases, but obviously the form of the exploit is totally different for a kernel driver bug and for a Drupal SQL injection.

The CVE database [8] contains the distribution of different vulnerabilities. It contains a huge amount of webserver-side coding vulnerabilities but the number of lower level coding vulnerabilities like memory corruption is also significant. This paper focuses on the lower level type of vulnerabilities, where the exploitation is carried out directly within the virtual memory.

We can also categorize the exploits according to the vulnerability exposure date. If the vulnerability was previously unknown, then the exploit would be called a zero day (0day) exploit. In other cases, the vulnerability is known but the exploit is still actual since the vulnerability is not patched everywhere or cannot be patched. In this case it can be referred as 1st day exploit.

Protecting the system against a first day exploit is usually not a real challenge, because the manufacturer has to provide a patch to remove the security gap after the vulnerability disclosure. The main focus of the exploit prevention is to protect the system against the 0day exploits, when concrete attack signatures cannot be used. This is possible by providing a secure execution environment which prevents the exploitation of an unknown vulnerability of the software. Several exploitation and attacking techniques exist and the main focus is to stop the exploitation without significant resource usage overhead. Since hardware based techniques hardly slow down the normal execution speed they are more preferable. In Chapter 2 different exploitation and protection techniques are summarized, while Chapter 3 focuses on future potential exploitation techniques and their analyses.

2 The Evolution of Software Vulnerability Exploitation and Protection

2.1 Early Exploitations

In the early years of software vulnerability exploitation, the aim was to find some coding error types that could lead to compromises, such as, arbitrary code execution. In this context there is no specific protection against vulnerability exploitation; everything is based on code correctness. The operating system focuses on the fast and efficient code execution within the virtual memory without any protection that considers coding errors. The program code and the shared libraries are loaded into the virtual memory to a code segment of the virtual address space having the operating system API. Each thread of the application has its own stack segment that consists of the method call stack frames. The whole process has some common heaps, where the dynamically allocated objects are stored. Each object has a virtual method table that contains the actual addresses of the virtual methods during runtime. For the sake of the effective and fast memory allocation and free in runtime, every heap is organized as series of linked list chunks with different sizes. A simplified figure of the virtual address space is presented in Figure 3.

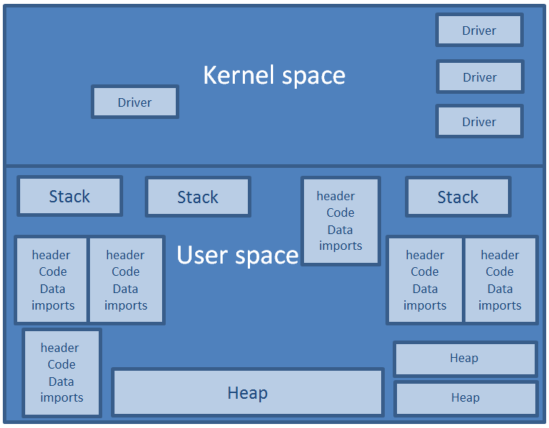


Figure 3
Virtual address space layout

In the early years of exploitation, the security of a software was only provided by the coding. If the code had no vulnerabilities, then the software would not be compromised. Unfortunately, this not the usual case, and with a single coding error, the attacker can force the software to execute malicious code. This malicious code execution is possible using several well-known techniques, such as, the stack overflow [9] the heap overflow [10], the format string vulnerability [11] or the use-after-free bug [12].

In the case of stack overflow [9] a local variable of a method (e.g. a string or an array) is overwritten inside the stack frame. Since the stack frame contains the method return pointer too, the attacker can redirect the execution to an arbitrary place by providing a new return pointer inside the local variable. By placing the attack payload in the corrupted local variable on the stack, the attacker can redirect the execution to the stack itself and the malicious payload is executed there.

In the case of heap overflow [10] the overwritten variable is in the heap. By overrunning a heap chunk the attacker will be able to modify the linked list pointers of the current heap. During the process of merging the freed heap chunks the chunk pointers are used for writing data. With an appropriate pointer modification, the attacker can write arbitrary data to an arbitrary place when the heap is freed. This is the way how the execution is redirected to the code where the malicious content is previously placed.

In the case of format string vulnerability [11] the attacker provides a series of formatting characters of which no data belong to for a *printf* type of functions. Choosing the formatting parameters appropriately, the attacker can write almost arbitrary data to an arbitrary place. By overwriting sensitive data in the virtual memory such as the stack method return pointer or a virtual address table pointer the execution is redirected to the attacker controlled place where the malicious payload is executed.

The use-after-free exploitation technique [12], is based on the modification of an object virtual method table pointer. If the vulnerability consists of an object that can be used after being freed then the attacker can try to allocate a fake object to the same place in the virtual memory where the original object was to redirect the execution. To achieve this, the attacker has to allocate multiple fake objects, with fake virtual method tables in the heap, that are pointing to the malicious code, that has already been placed in advance (heap spraying). When a virtual method of a freed vulnerable object is called, then the malicious code is executed.

It is easy to draw the conclusion from these early exploitations, that software security cannot be based only on the code correctness; additional protections are also necessary to avoid software bug exploitation.

2.2 Early Protections

The early solutions focused on the protection of the critical data in the virtual memory. For example, the stack frame return pointer overwriting, is aimed to be protected by the stack cookie [13]. As the stack cookie is placed between the method local variables and the method return pointer, any modification outside the real memory range of the local variables results in the modification of the stack cookie too. Therefore, the stack cookie modification indicates the stack frame corruption for the operating system. If stack cookie is placed in each stack frame, then this protection will be good enough to filter the stack frame corruption. However, it comes with a significant speed performance penalty.

The heap chunk header exploitation is prevented by the secure heap chunk unlink process [14] that validates the chunk header pointers before it is merged with another chunk. The secure structured exception handling [15] is another special defense that was introduced early against the exploitation of the exception handling vulnerabilities. This protection validates the exception handler pointer before it is executed.

In addition, several more robust protections appeared in the middle of the 2000s. These protections such as the Data Execution Prevention [4] and the Address Space Layout Randomization [5], aim to make software exploitation more complicated in general. Data execution prevention enforces memory page rights for the different types of segments in the virtual address space. Reading, writing or executing the page data are all different types of operations and DEP ensures that a memory page cannot be written and executed at the same time. DEP stopped several previously mentioned exploitation methods such as the stack overflow, since the payload can be written to a writable memory place but it cannot be executed due to the stack DEP protection.

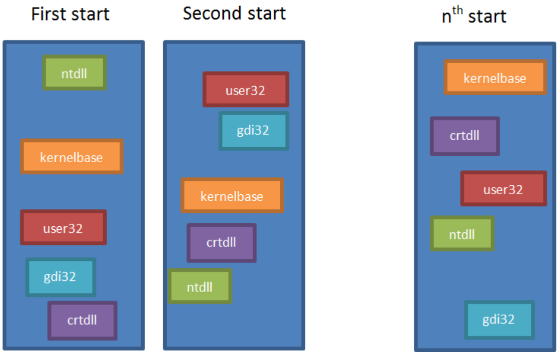


Figure 4

Address Space Layout Randomization in Windows [4]

Address Space Layout Randomization [5] is about to prevent the reuse of the already existing code parts in the virtual memory for malicious purposes. If the locations of the different segments in the virtual memory are randomized every time when the program is launched (Figure 4) then the attacker cannot rely on the known memory addresses of the shared libraries. It is also important to provide sufficient entropy for the randomization to prevent code reuse exploitations with guessing the ALSR offsets.

2.3 Advanced Exploitations

With the introduction of Data Execution Prevention [4], exploit writers could no longer place their own code to be executed. Attackers had to apply new techniques and the main idea became to execute the already existing code parts in the virtual memory that have the right to be executed, and that is the code reuse.

The first applied technique was the *return to libc* [16] type of exploitations where the corrupted method is redirected to an operating system API by placing its address as a return pointer in the corrupted stack frame. However, this technique is can execute only one operating system method but, selecting the right method, such as, the *WinExec* or *Execve,* with the right parameters can be sufficient.

A significant break-through for the code reuse was the invention of the Return Oriented Programming (ROP) [17]. This technique divides the desired payload into small code parts (gadgets) and searches for same code parts in the code libraries in the virtual address space. Since the gadgets are part of the virtual memory there is no need for own code to be placed, the payload is only a series of the gadget addresses and their parameters. Each gadget contains some assembly instructions with a *ret* type instruction at the end. When the corrupted method exits, the execution will be directed to the first gadget by its address. Because of the *ret* instruction at the end of the gadgets, the execution is directed to the next gadget every time by taking the next address on the corrupted stack frame by the *ret* instruction. ROP is proven to be Turing complete, the only limitation is the gadget catalog provided by the virtual address space. According to our current experiences there is no practical limitation, the attacker can almost always find enough gadgets in the virtual address space to turn off the DEP and continue the payload execution in the traditional way.

Jump Oriented Programming (JOP) [18] is a generalization of ROP and also capable of bypassing the DEP protection in a very sophisticated way. Similarly, to ROP, JOP executes the payload step by step by using small code parts called the functional gadgets. Each functional gadget has an indirect jump instruction at the end to redirect the instruction pointer to a special code part called the dispatcher gadget. The functional gadget addresses are stored in the dispatcher table that has to be placed in the virtual memory before the exploitation.

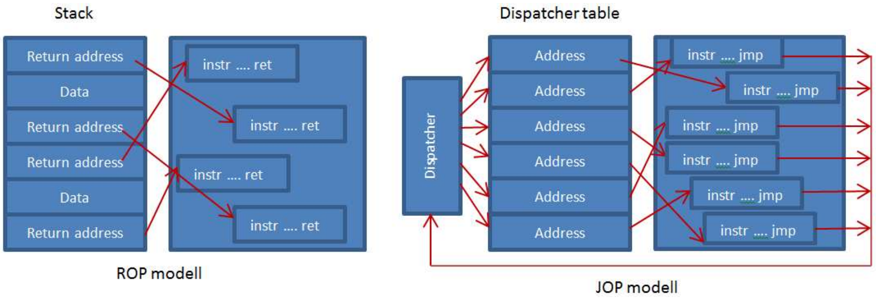


Figure 5
Return Oriented and Jump Oriented Programming [17] [18]

The dispatcher gadget maintains a register which always points to the next functional gadget in the dispatcher table to be executed. Instead of relying on the stack and the *ret* type instructions, JOP realizes its own stack like structure by the dispatcher table and the concatenation of the gadgets are ensured by the indirect jump instructions of the functional gadgets and the indirect call instruction of the dispatcher gadget.

There exist some other forms of scattered code reuse technique and these are under research such as the Sigreturn Oriented Programming (SROP) [19] or the Call Proceeded Return Oriented Programming (CPROP) [20]. SROP is based on the kernel context switching that saves the current execution context in a frame on the stack. The saved execution context contains the saved registers, as well as, the flags. In the case of stack overflow the instruction pointer is overwritten in the saved execution context. This is how the execution is redirected when the OS gets back the register values from the stack to resume the previous context. Contrary to ROP, SROP exploits are usually portable across different binaries and can also bypass ASLR in some cases.

Call Proceeded Return Oriented Programming applies whole functions as a gadget in order to bypass the control flow protections. With this approach every *ret* like instruction is legitimate during the payload execution and cannot be discovered with method return address validations.

Even if bypassing DEP is possible with the listed techniques, it is important to state that the gadget addresses should be known in order to apply these techniques. With ASLR this condition is not satisfied so attackers have to also consider ASLR bypassing, which is always a challenge. In some cases, ASLR can be bypassed by simple guessing the randomization offset [21] or by taking advantage of another vulnerability that leaks the randomization offset [6]. Special techniques to bypass ASLR and DEP together already exist: The Blind Return Oriented Programming (BROP) [22] and Just in Time Return Oriented Programming (JIT-ROP) [23]. BROP maps the virtual address gadgets by systematic guessing, while JIT-ROP does a just in time payload customization relying on an ASLR offset leak.

2.4 Current Exploitations

Secure software development is a fundamental question and several protections exist. Even compilers, operating systems and hardware manufactures try to mitigate software exploitation as much as possible, several exploits are still successful. Analyzing the exploits found in the wild, published by researchers and white hat hackers, it is clear that attackers have to consider the DEP and the ASLR together as a basic elements of the modern operating systems nowadays. Some browser exploits appeared at the end of 2016 and the most popular exploitation method was the Just in time Return Oriented Programming. A Firefox/Tor exploit (CVE-2016-9079) is revealed [24] at the end of 2016 that maps the WindowsPE structure in runtime to find appropriate ROP gadgets. The ROP code turns off the DEP with the *kernel32.VirtualAlloc* method then the rest of the payload is executed in the conventional way. Another DEP and ASLR bypassing exploit is related to the chakra JavaScript [6]. This exploit uses two different vulnerabilities. CVE 2016-7200 is used for the ASLR bypass, the *mshtml.dll* randomization offset is obtained with that bug, while CVE 2016-7201 is used to execute a short ROP code to turn off the DEP. Both cases belong to the Just in Time Return Oriented Programming category. ROP based exploits are used everywhere e.g. against network devices too. A vulnerability (CVE 2017-3881) [25] in the Cisco Cluster Management Protocol (CMP) processing code in Cisco Software could allow an unauthenticated, remote attacker to execute code with elevated privileges.

Based on the currently available software exploits, it is obvious that the main technique is still Return Oriented Programming. DEP and ASLR in combination were thought to provide very strong protection, but the current examples show that they can be bypassed routinely in several cases. The next step from the protection point of view is to disable ROP, where a possible approach is to enforce the right control flow during the code execution. In Section 3 several control flow bypassing techniques will be analyzed.

2.5 Enhanced Protections

Because the software vulnerability exploitation is still successful several advanced practical solutions are available to protect the systems, however these techniques have to keep up with the new challenges. One of the most-frequently applied ASLR bypass methods is guessing increasing the entropy of the Address Space Layout Randomization [26], it is a kind of mitigation, since it decreases the chance of a successful, brute-force, guessing attack. Forcing ASLR is another way to achieve better protection. Microsoft tried to prevent 0day exploitation with the Enhanced Mitigation Experienced Toolkit (EMET) [27] that provided some special advanced protections such as the anti-ROP technique. In 2016 Microsoft admitted that EMET is not proper for preventing 0day exploits and abandoned further development efforts. Microsoft has also introduced some new protections for the Edge browser [28] in 2016 such as the separated heap for the html objects or the delayed free to prevent the exploitation of use-after-free bugs. Other products, such as the Palo Alto exploit prevention [29], provides a wide choice of different protections, such as, detection of heap spraying and detection of ROP.

Since the main intension is stop the ROP-like exploitation several ideas are about to maintain and verify the correct control flow of a software [30]. One of the main questions of the protection over the efficiency is the performance. It is quite unfavorable if the exploit prevention comes with a performance penalty and slows down the execution speed significantly. Similarly, to DEP one good direction from performance point of view can be to provide hardware assisted anti-ROP protection. Such a solution is the Intel's Control Flow Enforcement (CFE) [31] which is a very promising technology.

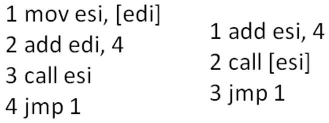
CFE provides two components for the protection: the shadow stack and the indirect jump verifier. The shadow stack is a not accessible data storage place, where the copy of the method return pointers are placed during runtime. Each time a method exists, it obtains the return pointer from the normal data stack and the shadow, then the two return addresses are compared as a control. With this technique the execution of small code gadgets, with unintended *ret* instructions is prevented. The indirect jump verifier is a procedure which controls the indirect jumps during the code execution. The idea is to mark each legitimate indirect jump instruction with a *nop-like* special instruction. Whenever an indirect jump is executed this special *nop-like* instruction must follow it. If an unintended indirect jump is executed, the operating system can observe it.

Even this protection seems to be impossible to bypass and some new designs have already arisen, that have the potential to bypass it.

3 Analysis of Control Flow Enforcement Bypassing Exploitations

Control flow integrity protections such as the Intel's Control Flow Enforcement are promising plans to stop Return Oriented Programming without any speed decrease. The main question from software vulnerability exploitation point of view is still whether the software bug exploitation will be stopped or significantly decreased by making *ROP-like* techniques totally impossible or is it just a step of the exploitation-protection fight that makes exploitation techniques even more sophisticated. There are several ongoing research projects on new software vulnerability exploitation methods, such as, the Loop Oriented Programming [32] or the Data Oriented Programming (DOP) [33] and also the Counterfeit Object-oriented Programming (COOP) [34].

The main engine of the LOP is the loop gadget. The loop gadget is a special code fragment that realizes a loop and calls a method with indirect call instruction in each step. Figure 6 illustrates some theoretical examples of possible X86 loop gadgets:



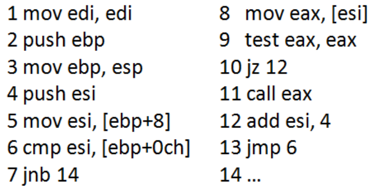
```
1 mov esi, [edi]
2 add edi, 4
3 call esi
4 jmp 1
```

```
1 add esi, 4
2 call [esi]
3 jmp 1
```

Figure 6
Minimal loop gadgets

In the two presented cases the codes contain a loop and the instructions inside are repeated infinitely. Similarly, to JOP there is a register (*edi* in the first case and *esi* in the second example) which points to a memory (dispatcher table) and the pointer is moving to the next table entry in each step of the loop by the *add* instruction. The gadgets also contain an indirect *call* and that is how the functional gadgets are executed by reading the next address from the dispatcher table in every step. A better loop gadget is presented in Figure 7. This code fragment not only executes the functions in the dispatcher table but has a condition to quit from the loop and finish the program.

Since every functional gadget is a whole legitimate function, there is no shadow stack being compromised. Since each *ret* instruction has the *call* instruction pair, thus every *ret-like* instruction will be legitimate. From the functional gadgets point of view LOP has strict limitations. To bypass CFE only whole functions can be used as functional gadgets and especially only those methods which have the indirect jump marker at the beginning. Satisfying all these conditions CFE cannot prevent LOP execution, since the stack return pointer is not compromised and all the indirect jumps are legitimate. Figure 8 shows the control flow of LOP.



```
1 mov edi, edi          8  mov eax, [esi]
2 push ebp              9  test eax, eax
3 mov ebp, esp          10 jz 12
4 push esi              11 call eax
5 mov esi, [ebp+8]      12 add esi, 4
6 cmp esi, [ebp+0ch]    13 jmp 6
7 jnb 14                14 ...
```

Figure 7
Loop gadget in msvcr.dll [32]

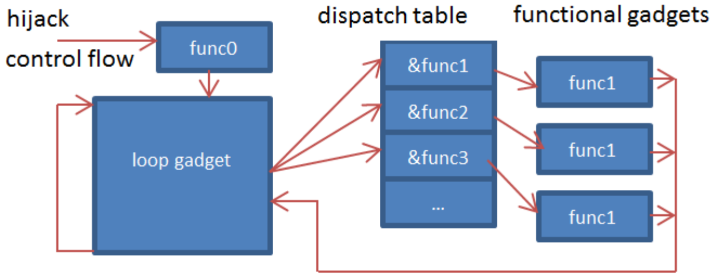


Figure 8
Loop Oriented Programming [32]

In the case of DOP [33] the main exploitation engine is the gadget dispatcher. Similarly, to the previous code-reuse techniques (JOP, LOP) a special code part controls the whole payload execution. The gadget dispatcher also has a loop but the functional gadgets are invoked in a different way than in case of LOP. DOP operates with six functional gadget types, but they implement different types of instructions: arithmetic/logical operation, assignment, load, store, jump, conditional jump. These functional gadget executions are repeated in various order during the payload execution with different parameters. The gadget dispatcher has a selector which sets which functional gadget should run in the next step and also sets the parameter of the next functional gadget. Figure 9 shows the control-flow of DOP.

Since the DOP functional gadgets implement general tasks, the gadget dispatcher of the DOP has more tasks than the LOOP gadget. It does not only invoke the next functional gadget, but sets the right parameters for the execution by customizing the input for the functional gadget. A practical example of a DOP gadget dispatcher is presented in Figure 10.

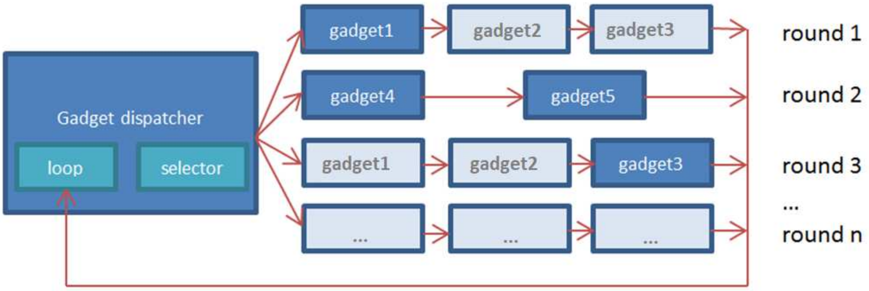
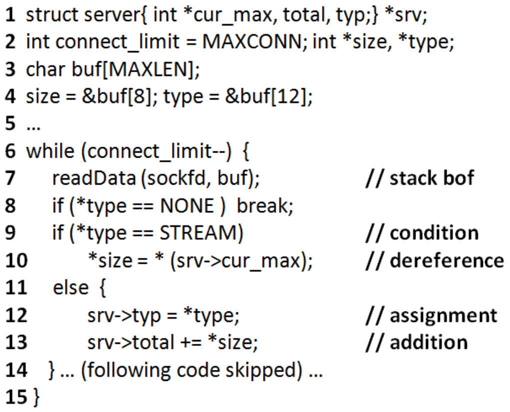


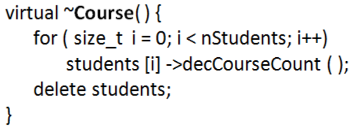
Figure 9
Data Oriented Programming [33]



```
struct server{ int *cur_max, total, typ;} *srv;
int connect_limit = MAXCONN; int *size, *type;
char buf[MAXLEN];
size = &buf[8]; type = &buf[12];
...
while (connect_limit--) {
    readData (sockfd, buf);            // stack bof
    if (*type == NONE ) break;
    if (*type == STREAM)               // condition
        *size = * (srv->cur_max);      // dereference
    else {
        srv->typ = *type;              // assignment
        srv->total += *size;           // addition
    } ... (following code skipped) ...
}
```

Figure 10
Example gadget dispatcher of Data Oriented Programming [33]

The Counterfeit Object Oriented Programming [34] is based on the virtual method calls of the Object Oriented Programming. Because of the inheritance, the object class is determined runtime in the case of virtual method call execution and the method addresses for each objects are stored in *vtable* structures. If the attacker manages to redirect the execution to a special virtual function, called the *main loop,* then they will be able to provide parameters to execute a Turing complete program without violating the Control Flow Enforcement. In the case of COOP, the dispatcher is the main loop and similarly to other loop techniques the task is to execute the functional gadgets in the right order and with the right parameters. The main loop as well as the functional gadgets are all legitimate virtual methods, so they are no longer really gadgets but long legitimate code parts. Figure 11 represents a main loop candidate, which is a destructor.



```
virtual ~Course( ) {
    for ( size_t  i = 0; i < nStudents; i++)
        students [i] ->decCourseCount ( );
    delete students;
}
```

Figure 11

A possible main loop for COOP [34]

As the attacker can set the *nStudent* parameter and the address pointing to the *student's* array, with the appropriate stack arrangement they can execute an arbitrary payload. Figure 12 shows a possible arrangement of the stack [34]. In Figure 12 the *students* array points back to the stack so the attacker can set the number of virtual methods to be executed, the address of the virtual methods to be executed, the order of the methods and also the method parameters.

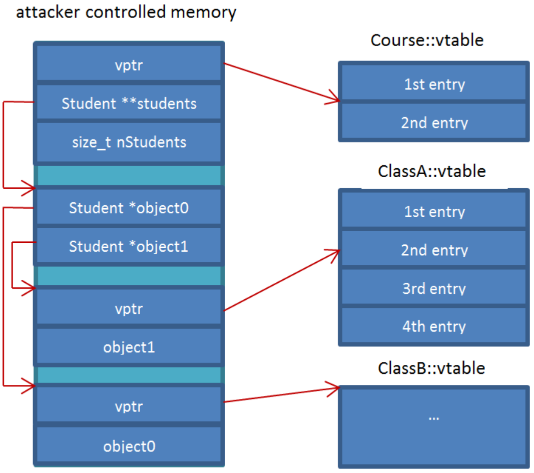


Figure 12

Stack arrangement for COOP [34]

Figure 13 shows the execution flow of COOP: taking advantage of a vulnerability the attacker redirects the code execution to the main loop and sets the stack pointer to a place where the main loop parameters are placed previously. COOP seems to be a very powerful technique against CFE as most of the programs currently use OOP.

According to our analysis it is important to distinguish between three different techniques considering the evolution of software vulnerability exploitation: In the first group we classify the techniques where the attacker can place and execute his own payload, like stack overflow, or classical use after free exploitation. Our second group contains the normal code reuse techniques, where the attacker executes the already existing code parts of the virtual memory, assembling the payload from small code parts that are not necessarily intended instructions called the gadgets. We call that group *ROP-like* techniques. Our third group contains the latest exploitation techniques where the payload is assembled from legitimate functions and the execution is controlled by a code part containing a loop. We refer this technique as *LOP-like* techniques. Table 1 contains a summary of the different techniques and their main techniques of incursion.

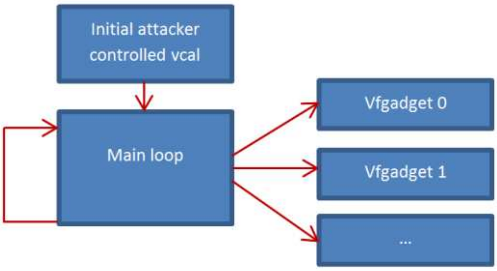


Figure 13
Counterfeit Object Oriented Programming [34]

Table 1
Software exploitation techniques

	Classical techniques	**ROP-like techniques**	**LOP-like techniques**
Method of payload execution	The payload to be executed is placed directly by the attacker	The payload is consist of small code parts (gadgets) from the virtual address space	The payload consist of legitimate methods from the virtual address space
DEP bypass	No	Yes	Yes
ASLR bypass	Not necessary	With additional vulnerability or memory leak	With additional vulnerability or memory leak
Shadow stack verification bypass	Stack overflow: No	ROP: No JOP: Yes	Yes
Indirect jump verification bypass	Use after free: No	ROP: Yes JOP: No	Yes
CFE bypass	No	No	Yes
Turing completeness	Yes	Yes, but depends on the gadget catalog	Yes, but depends on the method catalog
Example techniques	Stack overflow, use after free	ROP, JOP	LOP, DOP, COOP

The latest exploitation techniques are definitely able to bypass the Control Flow Enforcement technique [31]. So it is clear that if CFE will be used in the future then attackers will turn to *LOP-like* techniques. On the other hand, it is important to mention that there is no practical experience on the usability of these techniques.

Table 2
Control flow bypassing exploitations

	Loop Oriented Programming	**Data Oriented Programming**	**Counterfeit Object Oriented Programming**
Control gadget name	Loop gadget	Gadget dispatcher	Main loop
Control gadget functionality	Calls the methods step by step according to the dispatcher table	Selects the type of function first and call them step by step	Calls the virtual methods step by step with their parameter according to the stack arrangement
DEP bypass	Yes	Yes	Yes
ASLR bypass	With additional vulnerability or memory leak	With additional vulnerability or memory leak	With additional vulnerability or memory leak
Shadow stack verification bypass	Yes	Yes	Yes
Indirect jump verifier bypass	Yes	Yes	Yes
CFE bypass	Yes	Yes	Yes
Turing completeness	Yes, but depends on the method catalog	Yes, but depends on the method catalog	Yes, but depends on the virtual method catalog

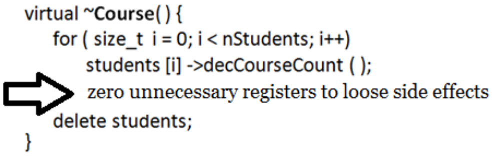
LOP-like techniques have to satisfy three conditions according to our analysis:

1) The virtual address space must have proper *loop-like* gadget
2) Possibility to redirect the code execution to the loop with the appropriate parameters
3) Appropriate method catalogs to execute the desired payload

The first and the third conditions are influenced by the content of the virtual address space. The second condition is influenced by the type of the vulnerability, as well as, the characteristics of the *loop-like* gadget dispatcher. Table 2 summarizes and compares the main behavior of the LOP-like exploitation methods. As it can be seen in Table 2 all three methods use a very similar idea: There is a loop which gets the control by a vulnerability with an initial setting. Then the loop continuously invokes legitimate methods from the virtual address space according to the previously placed method table and parameters by the attacker. However there is no hardware assisted Control Flow Enforcement yet, but the presented three exploitation techniques seem to be a real option to bypass CFE.

According to our analysis, preventing such an attack type is only currently possible during compilation time. From the point of view of the requirements the following things would be necessary to avoid such exploitations:

1) The key element of the exploitation is the loop-like gadget. The compilers should check and at least provide a warning message if a loop like gadget is available after the compilation.
2) Avoiding unwanted code redirection would be a basic prevention, but considering the current state this cannot be guaranteed. Almost all type of software vulnerabilities can achieve unwanted control flow change. Since software vulnerabilities cannot be totally excluded the prevention cannot be built on this either.
3) Preventing the creation of dispatcher table is also not realistic. With OOP different user controlled objects can be created in the heap. The only thing that is necessary from the attacker's point of view is to place the dispatcher table in a predictable place. This can be carried out together with a memory leak.



```
virtual ~Course( ) {
    for ( size_t  i = 0; i < nStudents; i++)
        students [i] ->decCourseCount ( );
        zero unnecessary registers to loose side effects
    delete students;
}
```

Figure 15
Loosing side effects of virtual methods

According to our analysis the only option to prevent such exploitations, is to prevent the loop like gadget compilation. On the other hand, in some cases, such as, in Figure 12, the loop like code block was created on purpose (iterating through the students). In such cases, our suggestion is to append the code and zero all registers, except for the return value in each step of the loop (Figure 15). With this solution the virtual methods negate the unwanted side effects that the attacker can use in these exploitations.

Conclusions

Based on previous experiences, we cannot simply let system security be based on the assumption of having perfect software, without vulnerabilities, to avoid software vulnerability exploitations. Additional advanced protections are necessary. From a performance point of view, hardware based solutions are preferred, such as DEP. However, ROP, which is the most popular technique of today's exploitations, can bypass DEP. Control Flow Integrity techniques, such as, CFE, aim to prevent *ROP-like* techniques, but new exploitation ideas, such as, LOP, DOP or COOP, have appeared recently. In this study, the main stages of software bug exploitations are analyzed, with a special focus on the behavior and capabilities of the cutting-edge techniques. We conclude, currently, it is not clear if there is any protection that is capable of stopping the exploitation of unknown software bugs; the best thing that can be done on the protection side, is to mitigate the potential for successful exploitation.

To avoid *LOP-like* exploitations, we suggested possible solutions to mitigate the risk of such attacks. According to our analysis the most feasible way of preventing loop oriented programming type attacks, can be implemented during the compilation stage. With code blocks presented in Figures 7 and 8, the compiler should try to avoid them, or at least provide a warning message if such code is created. For other loop like code blocks, such as, in COOP the compiler should try to insert extra code, that force the virtual methods to negate the side effects. With these added instructions, the attackers would not be able to create useful gadget chains.

References

[1] Offensive Security. Offensive securitys exploit database archive. https://www.exploitdb.com/

[2] Exploitalert website. http://exploitalert.com

[3] Blogger technology. Metasploit. https://blgtechn.blogspot.no/2012/08/metasploit.html

[4] Microsoft. A detailed description of the data execution prevention (dep) feature in windows xp service pack 2, windows xp tablet pc edition 2005, and windows server 2003, https://support.microsoft.com/en-us/help/875352/a-detailed-description-of-the-dataexecution-prevention-dep-feature-in-windows-xp-service-pack-2-windows-xp-tablet-pcedition-2005-and-windows-server-2003, 2006

[5] R. Seka Lixin Li, James E. Jus. Address-space randomization for windows systems. http://seclab.cs.sunysb.edu/seclab/pubs/acsac06.pdf, 2012

[6] B. Pak. Microsoft edge (Windows 10) - 'chakra.dll' info leak / type confusion remote code execution. https://www.exploit-db.com/exploits/40990/, 2017

[7] D. Dörr. Drupal 7.32 - sql injection (php), 2014, https://www.exploit-db.com/exploits/34993

[8] Cve details - the ultimate security vulnerability datasourse. http://cvedetails.com

[11] E. Levy. Smashing the stack for fun and profit. Phrack Mag, 49(14), 8 1996

[12] M. Kaempf. Smashing the heap for fun and profit. Phrack Magazine, 57(11), 8 2001

[13] Scut / team teso. Exploiting format string vulnerabilities. https://crypto.stanford.edu/cs155/papers/formatstring-1.2.pdf, 2001

[14] CWE Common Weakness Enumeration. Cwe-416: Use after free. https://cwe.mitre.org/data/definitions/416.html, 2012

[15] P. M. Wagle. Stackguard: Simple buffer overflow protection for gcc. In Proceedings of the GCC Developers Summit, pp. 243-256, 2003

[16] J. N. Ferguson. Understanding the heap by breaking it. http://www.blackhat.com/presentations/bh-usa-07/Ferguson/Whitepaper/bh-usa-07-ferguson-WP.pdf

[17] Microsoft. Preventing the exploitation of structured exception handler (seh) overwrites with sehop. https://blogs.technet.microsoft.com/srd/2009/02/02/preventing-the-exploitationof-structured-exception-handler-seh-overwrites-with-sehop/, 2009

[18] S. El Sherei. Return to libc. https://www.exploit-db.com/docs/28553.pdf

[19] H. Shacham, E. Buchanan, R. Roemer, and S. Savage. Return-oriented programming:Exploitation without code injection. https://www.blackhat.com/presentations/bh-usa-08/Shacham/BH_US_08_Shacham_Return_Oriented_Programming.pdf

[20] T. Bletsch, X. Jiang, and V. Freeh. Jump-oriented programming: A new class ofcode-reuse attack. In 17th ACM Computer and Communications Security, 2010

[21] E. Bosman and H. Bos. Framing signalsa return to portable shellcode. In SP ’14 Proceedings of the IEEE Symposium on Security and Privacy, pp. 243-258, 2014

[22] N. Carlini and D. Wagner. Rop is still dangerous: Breaking modern defenses. https://people.eecs.berkeley.edu/daw/papers/rop-usenix14.pdf, 2014

[23] H. Shacham, M. Page, B. Pfaff, Eu-Jin Goh, N. Modadugu,and D. Boneh. On the effectiveness of address-space randomization. http://benpfaff.org/papers/asrandom.pdf, 2004

[24] A. Bittau, A. Belay, A. Mashtizadeh, D. Mazieres, and D. Boneh. Hacking blind. http://www.scs.stanford.edu/sorbo/brop/bittau-brop.pdf, 2015

[25] L. Davi, C. Liebchen, K. Z. Snow, and F. Monrose. Isomeron: Code randomization resilient to (just-in-time) return-oriented programming. In NDSS Symposium 2015, 2015

[26] Ars Technica. Firefox 0-day in the wild is being used to attack tor users. https://arstechnica.com/information-technology/2016/11/firefox-0day-used-against-tor-users-almost-identical-to-one-fbi-used-in-2013/, 2016

[27] A. Kondratenko. Cve-2017-3881 cisco catalyst rce proof-of-concept. https://artkond.com/2017/04/10/cisco-catalyst-remote-code-execution/. 2017

[28] K. Johnson and M. Miller. Exploit mitigation improvements in windows 8. https://media.blackhat.com/bh-us-12/Briefings/M_Miller/ BH_US_12_ Miller_Exploit_Mitigation_Slides.pdf

[29] Microsoft. The enhanced mitigation experience toolkit. https://support.microsoft.com/en-us/help/2458544/the-enhanced-mitigation-experience-toolkit, 2012

[30] M. V. Yason. Understanding the attack surface and attack resilience of project spartans (edge) new edgehtml rendering engine. https://www.blackhat.com/docs/us-15/materials/us-15-Yason-Understanding-The-Attack-Surface-And-Attack-Resilience-Of-Project-Spartans-New-EdgeHTML-Rendering-Engine-wp.pdf, 2015

[31] Paloalto Networks. Traps administrators guide. https://www.paloaltonetworks.com/documentation/33/endpoint/endpoint-admin-guide, 2017

[32] J. Tang. Exploring control flow guard in windows 10. http://sjc1-te-ftp.trendmicro.com/assets/wp/exploring-control-flow-guard-in-windows10.pdf, 2016

[33] Intel. Control-flow enforcement technology preview. https://software.intel.com/sites/default/files/managed/4d/2a/control-flow-enforcementtechnology-preview.pdf, 2016

[34] Y. Li, B. Lan, H. Sun, C. Su, Y. Liu, and Q. Zeng. Loop-oriented programming: A new code reuse attack to bypass modern defenses. In 2015 IEEE Trustcom/BigDataSE/ISPA, pp. 91-97, IEEE Computer Society

[35] H. Hu, S. Shinde, S. Adrian, Z. Leong Chua, P. Saxena, and Z. Liang. Data-oriented programming: On the expressiveness of non-control data attacks. https://www.comp.nus.edu.sg/ ~shweta24/publications/dop_oakland16.pdf

[36] F. Schuster, T. Tendyck, C. Liebcheny, L. Davi, A. Sadeghiy, and T. Holz. Counterfeit object-oriented programming- on the difficulty of preventing code reuse attacks in c++ applications. syssec.rub.de/media/emma/veroeffentlichungen/2015/03/28/ COOP-Oakland15.pdf, 2015

Improving the Prediction Accuracy of Objective Video Quality Evaluation

Nenad Stojanović, Boban Bondžulić, Boban Pavlović, Marko Novčić, Dimitrije Bujaković

Military Academy, University of Defence in Belgrade, Generala Pavla Jurišića Šturma 33, 11000 Belgrade, Serbia
e-mails: nenad.m.stojanovic@vs.rs, boban.bondzulic@va.mod.gov.rs, boban.pavlovic@va.mod.gov.rs, marko.novcic@vs.rs, dimitrije.bujakovic@va.mod.gov.rs

Abstract: Three different approaches for improvement of objective video quality evaluation are presented in this paper. Improvement is obtained through quality guided temporal pooling, information content weighted temporal pooling, and multiscale analysis. The analysis was performed using five objective video quality assessment measures on two publicly available datasets with subjective quality scores. Only the videos with H.264, H.265, and MPEG-2 types of compression from two datasets were considered. The level of agreement between the subjective and objective quality scores are given through the Spearman rank-order correlation coefficients on complete datasets and subsets of video sequences with the same type of compression. Obtained results show that the performance of objective measures is dependent on the choice of the dataset. The greatest improvement is given by multiscale analysis.

Keywords: information pooling; objective video quality assessment; temporal pooling; video compression; video resolution

1 Introduction

In recent years there has been a rapid development of systems for digital processing, transmission and display of video content [1, 2]. This development has led to great interest in reliable, computationally efficient objective quality assessment measures. A subjective quality assessment is the most reliable way to determine the quality of video signals, but subjective tests are very expensive and time-consuming, and an alternative is sought in the form of objective quality assessment measures. There are three categories of objective quality assessment measures, No-Reference (NR), Full-Reference (FR), and Reduced-Reference (RR) [1-3]. This classification is based on the availability of the source signal on the receiving side. NR measures can be used in all applications where quality testing is required because this type of metrics do not need knowledge of the source signal. FR metrics require full information of source signal and for that reason, this category cannot be used in some real-time applications where the knowledge of the original signal on the receiving side is not possible. RR techniques are between the two previously described categories and in these techniques only the most important part of the source signal is needed for quality evaluation. Objective image/video quality assessment measures have found numerous applications. Most applications are in situations where the quality of the modified version of the image/video needs to be evaluated.

Algorithms for video quality assessment usually have two phases. In the first, quality is evaluated on local spatial/temporal level, and in the second, spatial/temporal pooling of local scores produces a final value of quality [4]. Spatial and temporal integrations are closely related to visual significance. Estimation of visual significance identifies information on motion image which notably effect on observer during forming an impression of the quality. This allows for increasing the impact of essential information on the final score of the evaluation. Generally, strong degradation in space and/or time has a great effect on the final impression of quality. Strong distortions give low values of similarity between reference and test signals, so using the scores with the lowest quality, the final quality value can be formed. Also, the resolution of video during processing and display can have significant effect on final quality assessment.

The increasing number of video services and the increase in the resolution of the video display devices have led to the requirement for higher coding efficiency compared to the H.264 compression algorithm capabilities [5]. Therefore, a novel compression algorithm, H.265, was developed [6, 7], and a new compression standard is under progress [8, 9]. The goal of introducing the H.265/HEVC standard was to maintain subjective video quality by reducing the bit rate of 50% compared to H.264 [7].

The aim of this paper is to analyze the performance of objective quality assessment measures on sequences with MPEG-2, H.264 and H.265 compressions, using three different approaches for improving the prediction accuracy of objective video quality estimation. Objective quality assessment measures performance was analyzed on two publicly available, subject rated video datasets. H.264, H.265 and MPEG-2 compression algorithms are most commonly used algorithms in video systems, and therefore they are chosen for the analysis.

The quantitative measure adopted by the ITU [10] – Spearman's Rank Order Correlation Coefficient (SROCC) between subjective and objective quality scores, was used in the performance analysis of objective quality assessment measures.

In the second part of the paper are described used FR objective quality measures and the most important information of two datasets for video quality is provided. Three possible directions to improve the video quality estimation with results are given in the third part of the paper. The conclusions and further research directions are given at the end of the paper.

2 Overview of Objective Measures and Video Quality Datasets

The five objective video quality assessment measures were used in the analysis. Peak Signal to Noise Ratio (PSNR) [11], is the first measure. PSNR is an unavoidable measure in image/video quality analysis, although is often criticized [11, 12]. The Structural Similarity Index (SSIM) is the second measure, which is present in almost all tests of image/video quality measures [13].

Table 1
Comparison of used video datasets

Video Dataset	FERIT-RTRK			CSIQ Video		
Number of original sequences	6			12		
Number of tested (distorted) sequences	H.264	30	90	H.264	36	72
	H.265	30		H.265	36	
	MPEG-2	30				
Number of degradation levels	5			3		
Degradation types	H.264, H.265, MPEG-2			H.264, H.265, MJPEG, SNOW, packet loss, AWGN		
Tested degradations	H.264, H.265, MPEG-2			H.264, H.265		
Resolution	1920x1080 pixels			832x480 pixels		
Length	5 seconds			10 seconds		
Frame rates	60 fps			24, 30, 50, 60 fps		
Number of observers	30			35		

SSIM has numerous modifications, such as GMSM (average of local quality values of the gradient magnitude information preservation) and GMSD (standard deviation of local quality gradient magnitude similarity scores) [14]. GMSM and GMSD are third and fourth used measures. The fifth objective video quality assessment measure is VQ^{AB} [15]. VQ^{AB} is based on the analysis of the spatial information preservation (through the gradient magnitude and gradient orientation information preservation), the temporal information preservation (through the preservation of information on changes between frames) and the color information preservation.

The analysis was performed on two video quality datasets: FERIT-RTRK [16] and CSIQ Video [17]. Table 1 shows data comparison between two used video datasets. FERIT-RTRK dataset has more degradation levels and tested (degraded) sequences than the CSIQ dataset, but has less referent video sequences and shorter videos length.

3 Video Quality Analysis

Subject rated video quality datasets are of great importance because they can help in developing reliable objective measures. Quality guided temporal pooling, information content weighted temporal pooling and multiscale analysis are some of the approaches for improving the level of agreement between subjective and objective quality scores and performance of objective quality measures.

3.1 Quality Guided Temporal Pooling

It has been shown that regions of poor image quality significantly affect on a human estimation of visual quality [18]. This fact is used for quality guided lowest percentile temporal pooling approach, where p% (p percent) of the frames with the lowest quality scores are used. Parameter p represents a number of used frames in percent, in the step by 2%. The temporal pooling process is carried out in the following way. After determining objective quality scores on a frame-by-frame basis, their sorting is done in rising order, after which the final quality score is determined as the mean value of p% of the lowest scores of frames quality. Values beyond this range are rejected. This approach is guided by the hypothesis that the frames with poor quality can have a dominant role in the subjective impression of quality [18]. Measure GMSD has an inverse scale, so the sorting is done in descending order, after which the final quality score is determined as the mean value of p% of the highest scores of frames quality.

Figure 1 shows the normalized values of objective quality scores of frames for two sequences with H.265 compression. Graphics show significant quality variations during the lasting of the video. Also, from Figure 1, a periodic repetition of the local maximums of quality is observed, which is the consequence of the I frames present in the degraded sequences. In addition, it can be noticed that objective quality measures in different ways respond to changes that occur in video sequences. Thus, from the Figure 1 (b), between the 100th and 300th frame, can be noted that PSNR and SSIM objective values are increasing then decreasing, the values of GMSD and VQ^{AB} objective measures decrease, while GMSM values increase.

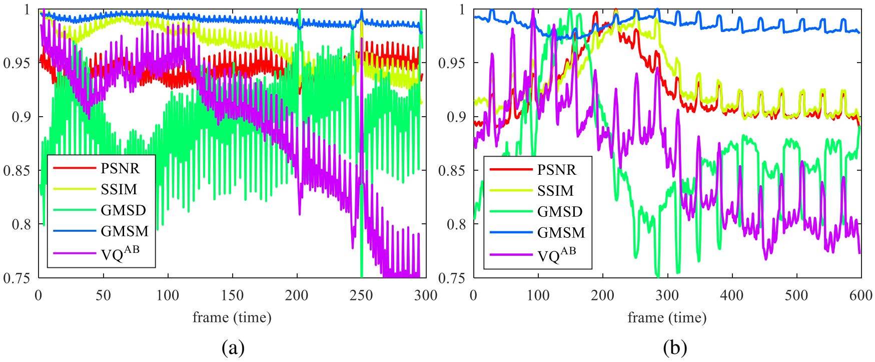


Figure 1

Objective quality scores of frames of the analyzed sequences: (a) sequence yac_H265_1 from FERIT-RTRK dataset and (b) sequence BQTerrace_832x480_dst_18 from CSIQ dataset

Figure 2 shows the mean values of p% of the lowest quality scores for two objective measures – PSNR and VQ^{AB}. All test sequences from both video datasets are analyzed. From Figure 2 it is noticed that dynamic ranges of the objective values differ on analyzed datasets. Thus, the dynamic range of PSNR measure is 4 dB narrower on the FERIT-RTRK dataset than on CSIQ dataset. Furthermore, both measures have lower objective quality values on the sequences from the FERIT-RTRK dataset. This observation is also valid for other analyzed objective measures – SSIM, GMSM and GMSD (this measure has an inverse scale, so the higher quality scores are obtained on the FERIT-RTRK dataset).

Due to the content of the test videos, it can be explained why the results of the p% of the lowest objective quality scores differ between these two datasets. Used video sequences should represent the real world images, i.e. datasets contain a wide range of content. A variety of content of a dataset can be characterized using Spatial Activity (SA), Temporal Activity (TA) and colorfulness index. In this work, the spatial complexity of video sequences, SA, was analyzed based on the mean values of the gradient magnitude of the frames. Sobel operator was used to determine the gradient magnitude. Figure 3 shows SA values per frame of all sequences in CSIQ and FERIT-RTRK datasets.

Dynamic ranges of the spatial activity of the distorted sequences on these two datasets are significantly different, which can be seen from Figure 3. The dynamic range of the spatial activity values is almost two times bigger in the CSIQ dataset than in the FERIT-RTRK dataset because video sequences from the CSIQ dataset are richer with details. This can be a consequence of the format of the delivered videos. Namely, CSIQ dataset sequences are delivered in raw format – YUV420, while all sequences (including reference) of the FERIT-RTRK dataset are delivered in the compressed format – mp4.

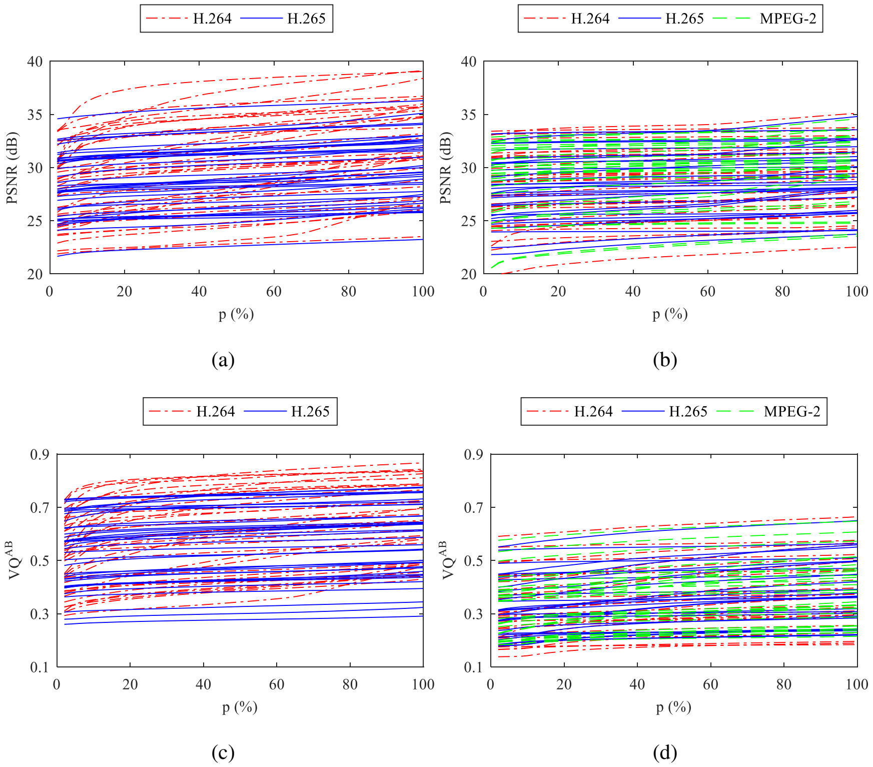
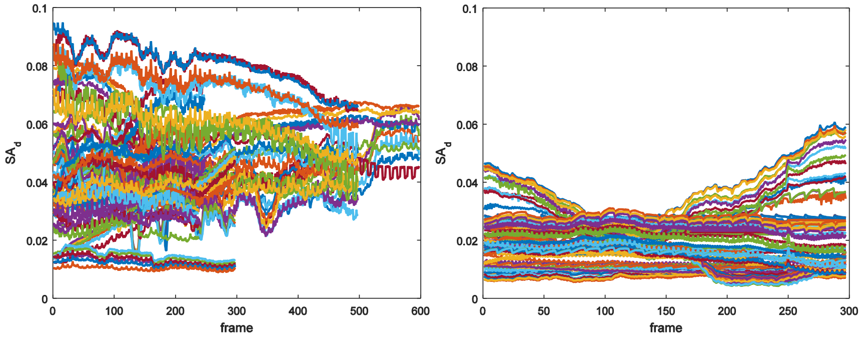


Figure 2

Mean values of the p% of the lowest objective quality scores: (a) PSNR on CSIQ dataset, (b) PSNR on FERIT-RTRK dataset, (c) VQ^{AB} on CSIQ dataset and (d) VQ^{AB} on FERIT-RTRK dataset

A similar analysis was carried out in the analysis of the TA [15] per frame of the distorted sequences, whereby the conclusion that the dynamic ranges of the TA of the sequences from these two datasets are approximately the same.

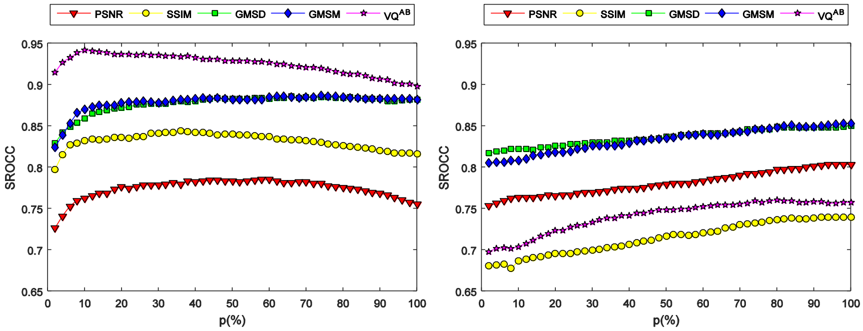
The influence of the selection of the frames with poor quality on objective assessment was analyzed through the SROCC at the level of complete datasets and at the level of subsets of sequences with the same type of degradation. Results of the correlations with subjective quality impressions on the global level are shown in Figure 4.



(a) (b)

Figure 3

Spatial activity values per frame of the test (degraded) sequences on the: (a) CSIQ dataset and (b) FERIT-RTRK dataset



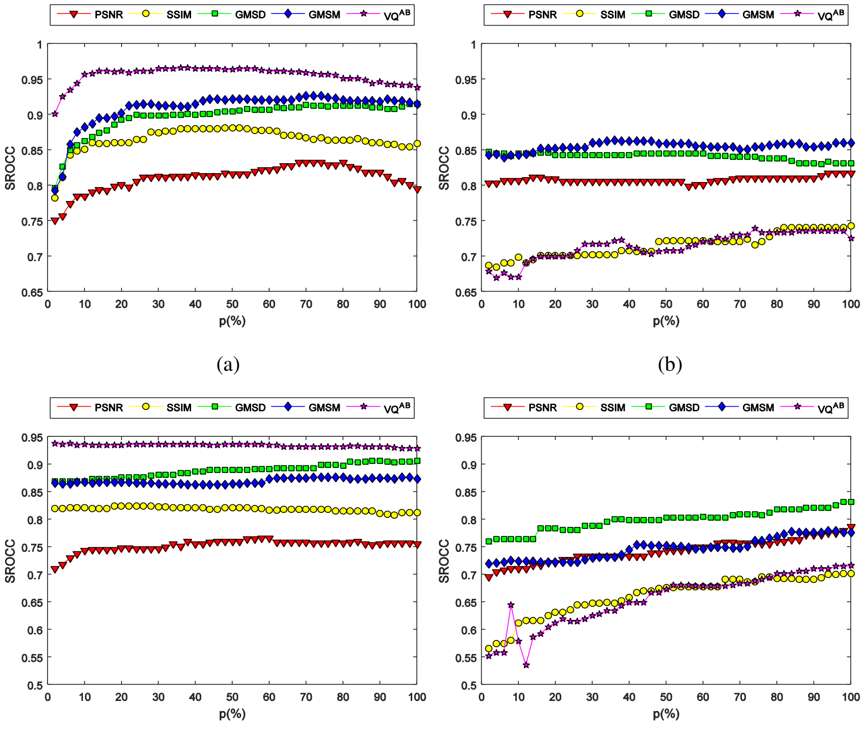
(a) (b)

Figure 4

Rank order correlation (SROCC) between subjective and p% of the lowest objective quality scores on: (a) CSIQ dataset and (b) FERIT-RTRK dataset

From Figure 4 it is noticed different trends of SROCC values on these two analyzed datasets. Unlike the FERIT-RTRK dataset, where the best agreement of subjective and objective quality scores is achieved if all frames of the test sequences (p=100%) are used, in CSIQ dataset can be noted that using the lowest scores of frames quality can improve the performance of objective measures (VQ^{AB}, SSIM, and PSNR). In this dataset, the greatest gain would be obtained using 10% of the lowest scores of the VQ^{AB} objective measure (level of agreement by using all frames is 0.9, while the level of agreement by using 10% of the lowest quality scores is 0.94). Objective quality assessment measure, VQ^{AB}, has the best performance on the CSIQ dataset (in the entire range values of parameter p). However, the performance of this measure is significantly worse on the FERIT-RTRK video dataset. The performance of all other tested objective measures is worse on the FERIT-RTRK than on the CSIQ dataset, where the biggest performance drop is noticed at VQ^{AB} and SSIM measurements.

The performance of objective measures on subsets of videos with the same type of degradation is presented in Figure 5. From this figure, it can be noticed that on H.264 compressed videos, the performance of objective measures depends on the values p. On the other side, in a subset of sequences with H.265 compression, the performance of objective measures is almost independent of the choice of values p. The performance of objective measures on corresponding subsets of the FERIT-RTRK dataset is worse than on subsets of the CSIQ dataset.



(a)

(b)

(c)

(d)

Figure 5

Rank order correlation (SROCC) between subjective and p% of the lowest objective quality scores on subsets of video sequences: (a) H.264 subset of CSIQ dataset, (b) H.264 subset of FERIT-RTRK dataset, (c) H.265 subset of CSIQ dataset and (b) H.265 subset of FERIT-RTRK dataset

3.2 Information Content Weighted Temporal Pooling

The problem with non-uniform distribution of frames quality over time can be solved by assigning them a time-varying significance (weight) [19]. The general shape of a temporal weighting approach is given by:

$$Q_f = \frac{\sum_{i=1}^{N} w_i Q_i}{\sum_{i=1}^{N} w_i} \quad (1)$$

where w_i is the weight associated with the i-th temporal location (frame), N is the number of frames in the degraded/reference sequence and Q_i is the quality value at the i-th temporal location. The weights are determined by the frame information content (using reference or distorted frames or both of them).

A list of 18 weighted functions is given in Table 2. The significance associated with the estimates of the frames quality during the time is derived from the spatial activity of the reference (SA_r) and distorted (SA_d) video sequences, and from the temporal activity of the reference (TA_r) and distorted (TA_d) video sequences. Impact of the significance of the frames during the time is given through the rank order correlation between subjective and objective quality scores on a global level – on complete datasets. Spatial and temporal activities are combined in an additive and multiplicative manner or as their maximum value. In Table 2, for objective measures, with +/- are marked situations in which weighting led to an improvement/deterioration of the performance of the objective measure, while the value presents the gain/loss relative to the SROCC of the standard method of pooling the frames quality scores (averaging). All presented results are on the relation with correlation coefficients where 100% of the frames are included from the previous subchapter.

From the Table 2, it can be noticed that all measures have some improvement on the CSIQ dataset within all weighted functions, except the GMSD measure which has an improvement for only four weighted functions. In all objective measures, the use of temporal activities (TA_r/TA_d) results in a higher gain than the use of spatial activities (SA_r/SA_d). Also, it can be concluded that weighted functions in multiplicative form lead to a greater agreement between subjective and objective quality scores than weighted functions in additive form or weighted functions with the selection of maximum. The highest improvement on this dataset was obtained by using PSNR objective measure overall analyzed weighted functions. The greatest gain in the PSNR measure was achieved by applying multiplicative weighted functions $TA_d \cdot TA_r$ (0.03), $SA_r \cdot TA_r$ (0.024) and $SA_d \cdot TA_d \cdot SA_r \cdot TA_r$ (0.022). These three weighted functions are suitable for the accuracy improvement of other objective measures.

Contrary, on the FERIT-RTRK dataset, there is no improvement except for the SSIM measure. The gain achieved for this measure on this dataset is slightly worse than the gain achieved on the CSIQ dataset. According to the achieved gain, the multiplicative weighted functions stand out – $SA_d{\cdot}TA_d{\cdot}SA_r{\cdot}TA_r$ (0.013), $SA_d{\cdot}TA_d$ (0.01), $TA_d{\cdot}TA_r$ (0.008) and $SA_r{\cdot}TA_r$ (0.008). Objective measure SSIM is the only measure which has improvement on both used datasets and in all 18 used weighted functions.

Table 2

Improvement/deterioration of the performance of the objective measures on CSIQ and FERIT-RTRK datasets relative to the SROCC of all frames in all test sequences

Weight, w_i	PSNR		SSIM		GMSD		GMSM		VQ^{AB}	
	CSIQ	FERIT	CSIQ	FERIT	CSIQ	FERIT	CSIQ	FERIT	CSIQ	FERIT
SA_r	+ 0.011	- 0.004	+ 0.005	+ 0.003	- 0.003	- 0.002	+ 0.002	- 0.002	+ 0.006	- 0.003
SA_d	+ 0.008	- 0.006	+ 0.004	+ 0.006	- 0.008	- 0.003	0	- 0.004	+ 0.001	- 0.003
TA_r	+ 0.019	- 0.004	+ 0.011	+ 0.005	+ 0.001	- 0.001	+ 0.004	- 0.003	+ 0.012	- 0.002
TA_d	+ 0.019	- 0.005	+ 0.009	+ 0.003	- 0.001	- 0.001	+ 0.004	- 0.001	+ 0.009	- 0.007
$SA_r{\cdot}TA_r$	+ 0.024	- 0.012	+ 0.012	+ 0.008	+ 0.002	- 0.003	+ 0.002	- 0.003	+ 0.018	- 0.004
$SA_d{\cdot}TA_d$	+ 0.018	- 0.015	+ 0.010	+ 0.010	- 0.003	- 0.004	+ 0.001	- 0.003	+ 0.008	- 0.006
$\max(SA_r,TA_r)$	+ 0.014	- 0.003	+ 0.011	+ 0.005	+ 0.001	- 0.001	+ 0.005	- 0.003	+ 0.007	- 0.001
$\max(SA_d,TA_d)$	+ 0.012	- 0.004	+ 0.006	+ 0.006	- 0.004	- 0.002	+ 0.003	- 0.002	+ 0.004	- 0.003
SA_r+TA_r	+ 0.015	- 0.003	+ 0.009	+ 0.004	- 0.001	- 0.002	+ 0.005	- 0.002	+ 0.009	- 0.002
SA_d+TA_d	+ 0.011	- 0.003	+ 0.005	+ 0.005	- 0.003	0	+ 0.003	- 0.001	+ 0.005	- 0.003
SA_r+SA_d	+ 0.009	- 0.004	+ 0.005	+ 0.004	- 0.006	- 0.002	+ 0.001	- 0.002	+ 0.002	- 0.002
TA_r+TA_d	+ 0.020	- 0.004	+ 0.010	+ 0.004	- 0.001	- 0.002	+ 0.004	- 0.002	+ 0.011	- 0.003
$SA_r+SA_d+TA_r+TA_d$	+ 0.015	- 0.004	+ 0.008	+ 0.005	- 0.002	- 0.002	+ 0.005	- 0.001	+ 0.007	- 0.002
$\max(SA_d,SA_r)$	+ 0.011	- 0.004	+ 0.005	+ 0.003	- 0.003	- 0.002	+ 0.002	- 0.003	+ 0.006	- 0.003
$\max(TA_d,TA_r)$	+ 0.020	- 0.004	+ 0.011	+ 0.004	+ 0.001	- 0.002	+ 0.005	- 0.003	+ 0.012	- 0.002
$TA_d{\cdot}TA_r$	+ 0.030	- 0.011	+ 0.014	+ 0.008	0	- 0.005	+ 0.002	0	+ 0.014	- 0.006
$SA_d{\cdot}SA_r$	+ 0.010	- 0.015	+ 0.006	+ 0.008	- 0.006	- 0.004	0	- 0.006	+ 0.006	- 0.005
$SA_d{\cdot}TA_d{\cdot}SA_r{\cdot}TA_r$	+ 0.022	- 0.022	+ 0.019	+ 0.013	- 0.008	- 0.006	0	- 0.007	+ 0.011	- 0.008

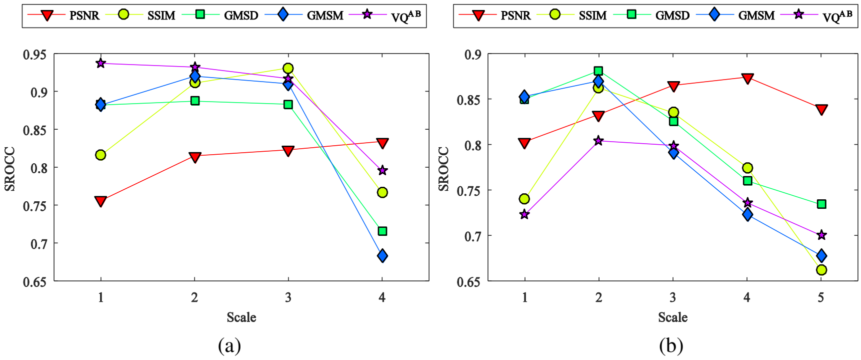
3.3 Multiscale Analysis

In further analysis, the objective quality of video sequences in different scales (resolutions) is evaluated. The level of agreement between subjective and objective quality scores is analyzed in five scales for the FERIT-RTRK dataset, and in four scales for CSIQ video dataset, because the resolution of videos from

FERIT-RTRK dataset is greater than videos from CSIQ dataset. The decimation of the original and test sequences was made with scaling factors 1/2, 1/4, 1/8, and 1/16. In addition to decimation, bicubic interpolation was performed. Scale 2 corresponds to scaling factor 1/2, while scale 5 corresponds to scaling factor 1/16 [20-22].

The values of the VQ^{AB} measure are obtained by averaging of the lowest 20% frames quality scores in this analysis, as suggested in [15]. For all other measures, 100% of the frames are used.

Figure 6 shows the dependence of the level of agreement between subjective and objective quality scores (SROCC) over different scales (resolutions) with both used datasets. Correlation is calculated on complete datasets. From Figure 6 it is noted that the performance of objective measures significantly depends on the scale in which the original and compressed video was compared. The choice of the optimal scale depends on the objective measure, too. In this way, the observation from [20] that the assessment of image/video quality in different resolutions provides more flexibility in incorporating the variations of viewing conditions was confirmed.



(a)

(b)

Figure 6

Rank order correlation (SROCC) between subjective and objective quality scores in different scales on: (a) CSIQ dataset and (b) FERIT-RTRK dataset

The highest degree of agreement for CSIQ dataset is between subjective and VQ^{AB} objective quality scores (original resolution, SROCC=0.936). Applying the GMSD objective measure provides the highest degree of agreement of quality scores on the FERIT-RTRK dataset (scale 2, SROCC=0.881).

In both video datasets, it is noticeable the performance improvement of the SSIM objective quality assessment measure, comparing with the original resolution; the comparison for the FERIT-RTRK dataset is carried out in scale 2 (Figure 6 (b) shows the SSIM's SROCC jump from 0.741 to 0.862), and for the CSIQ video dataset in scale 3 (Figure 6 (a) shows the SSIM's SROCC jump from 0.816 to 0.931).

Observing the results of the PSNR measure, a higher correlation with subjective scores is obtained by analyzing lower resolutions, while other objective measures have better performance in higher resolutions. In this way, it can be concluded that in the comparison of the signals in higher scales (lower resolution), the analysis of the energy preservation of the signal is important, while in the lower scales (higher resolution), the analysis of the preservation of the signal structure is more important. Furthermore, on FERIT-RTRK dataset it can be seen that the degree of agreement between subjective and PSNR objective quality scores in scale 4 is close to the results of the best GMSD measure (0.874 vs. 0.881).

Conclusions

Three approaches for improving the performance of objective quality assessment measures are presented in the paper. The presented approaches are quality guided temporal pooling, information content-weighted temporal pooling, and multiscale analysis. The five objective video quality assessment measures and two publicly available video datasets with H.264, H.265, and MPEG-2 compressed video contents are used in the analysis. It has been shown that the performance of objective measures significantly depends on the choice of the dataset, which makes it necessary to use more reference video datasets in video quality analyzes. Since these datasets contain a relatively small number of test signals with H.264 and H.265 compressed contents (60+72 sequences), it can be concluded that there is a need for new datasets, which will contain a significantly larger number of compressed test signals.

In addition, from the analysis it can be concluded that the greatest potential for improving the performance of objective measures on both datasets has the multiscale approach, where the improvement depends on the choice of an objective measure. By applying this approach, the improvement of accuracy prediction achieved through the correlation of ranks was up to 0.12 (SSIM objective measure on both analyzed datasets). Quality guided temporal pooling, implemented through the use of the lowest quality scores, on the CSIQ dataset has led to the improvement of the performance of objective measures (rank correlation increased by up to 0.05), while on the FERIT-RTRK dataset the performance with such integration is in the level of performance without pooling. Information content-weighted temporal pooling does not give significant improvement (rank correlation increased by up to 0.03 on CSIQ dataset), and in this case, except for SSIM, information content-weighted temporal pooling did not lead to an improvement in the results of objective measures on the FERIT-RTRK dataset. The lack of improvement in the results of objective measures using temporal pooling on a FERIT-RTRK dataset is probably due to the format of the delivered sequences – the original and test sequences were delivered in a compressed mp4 format.

As these three approaches were analyzed separately, in further work we will analyze their combined effect in objective video quality assessment.

Acknowledgment

This research has been a part of the project No. VA-TT/3/20-22 supported by the Ministry of Defence, Republic of Serbia.

References

[1] Bovik, A. C.: Automatic prediction of perceptual image and video quality, Proceedings of the IEEE, 2013, Vol. 101, No. 9, pp. 2008-2024

[2] Baig, M. A., Moinuddin, A. A., Khan, E., Ghanbari, M.: Image fidelity estimation from received embedded bitstream, Signal, Image and Video Processing, 2020, Vol. 14, No. 3, pp. 465-472

[3] Yu, X., Bampis, C. G., Gupta, P., Bovik, A. C.: Predicting the quality of images compressed after distortion in two steps, IEEE Transactions on Image Processing, 2019, Vol. 28, No. 12, pp. 5757-5770

[4] Bampis, C. G., Li, Z., Moorthy, A. K., Katsavounidis, I., Aaron, A., Bovik, A. C.: Study of temporal effects on subjective video quality of experience, IEEE Transactions on Image Processing, 2017, Vol. 26, No. 11, pp. 5217-5231

[5] Miličević, Z., Bojković, Z., Rao, K. R.: HEVC vs. H.264/AVC standard approach to coder's performance evaluation, WSEAS Transactions on Signal Processing, 2015, Vol. 11, pp. 272-279

[6] Sullivan, G. J., Ohm, J.-R., Han, W.-J., Wiegand, T.: Overview of the high efficiency video coding (HEVC) standard, IEEE Transactions on Circuits and Systems for Video Technology, 2012, Vol. 22, No. 12, pp. 1649-1668

[7] Tan, T. K., Weerakkody, R., Mrak, M., Ramzan, N., Baroncini, V., Ohm, J.-R., Sullivan, G. J.: Video quality evaluation methodology and verification testing of HEVC compression performance, IEEE Transactions on Circuits and Systems for Video Technology, 2016, Vol. 26, No. 1, pp. 76-90

[8] Sullivan, G. J.: Video coding standards progress report: Joint video experts team launches the versatile video coding project, SMPTE Motion Imaging Journal, 2018, Vol. 127, No. 8, pp. 94-98

[9] Takamura, S.: Versatile video coding: A next-generation video coding standard, NTT Technical Review, 2019, Vol. 17, No. 6, pp. 49-52

[10] ITU Tutorial: Objective perceptual assessment of video quality – Full reference television, Geneva, ITU, 2004

[11] Wang, Z., Bovik, A. C.: Mean squared error: Love it or leave it? A new look at signal fidelity measures, IEEE Signal Processing Magazine, 2009, Vol. 26, No. 1, pp. 98-117

[12] Jakšić, B., Gara, B., Petrović, M., Spalević, P., Lazić, Lj.: Analysis of the impact of front and back light on image compression with SPIHT method

during realization of the chroma key effect in virtual TV studio, Acta Polytechnica Hungarica, 2015, Vol. 12, No. 2, pp. 71-88

[13] Wang, Z., Bovik, A. C., Sheikh, H. R., Simoncelli, E. P.: Image quality assessment: From error visibility to structural similarity, IEEE Transactions on Image Processing, 2004, Vol. 13, No. 4, pp. 600-612

[14] Xue, W., Zhang, L., Mou, X., Bovik, A. C.: Gradient magnitude similarity deviation: A highly efficient perceptual image quality index, IEEE Transactions on Image Processing, 2014, Vol. 23, No. 2, pp. 684-695

[15] Petrović, V., Bondžulić, B.: Objective assessment of surveillance video quality, Sensor Signal Processing for Defence, London, UK, IET, 2012, pp. 1-5

[16] Bajčinovci, V., Vranješ, M., Babić, D., Kovačević, B.: Subjective and objective quality assessment of MPEG-2, H.264 and H.265 video, International Symposium ELMAR, Zadar, Croatia, IEEE, 2017, pp. 73-77

[17] Vu, P. V., Chandler, D. M.: ViS3: An algorithm for video quality assessment via analysis of spatial and spatiotemporal slices, Journal of Electronic Imaging, 2014, Vol. 23, No. 1, pp. 1-24

[18] Moorthy, A. K., Bovik, A. C.: Visual importance pooling for image quality assessment, IEEE Journal of Selected Topics in Signal Processing, 2009, Vol. 3, No. 2, pp. 193-201

[19] Duanmu, Z., Ma, K., Wang, Z.: Quality-of-experience for adaptive streaming videos: An expectation confirmation theory motivated approach, IEEE Transactions on Image Processing, 2018, Vol. 27, No. 12, pp. 6135-6146

[20] Wang, Z., Simoncelli, E. P., Bovik, A. C.: Multi-scale structural similarity for image quality assessment, The Thirty-Seventh Asilomar Conference on Signals, Systems and Computers, Pacific Grove, USA, IEEE, 2003, pp. 1398-1402

[21] Bondžulić, B., Petrović, V., Mitrović, S., Pavlović, B., Andrić, M.: Visual attention pooling and understanding the structural similarity index in multi-scale analysis, Optica Applicata, 2014, Vol. 44, No. 2, pp. 267-283

[22] Pavlović, B., Bondžulić, B., Stojanović, N., Novčić, M., Bujaković, D.: Comments on objective quality assessment of MPEG-2, H.264 and H.265 videos, New Trends in Signal Processing, Demanovska Dolina, Slovakia, IEEE, 2018, pp. 156-163

A Highly Reliable, Modular, Redundant and Self-Monitoring PSU Architecture

Bertalan Beszédes, Károly Széll, György Györök

Alba Regia Technical Faculty, Óbuda University
Budai út 45, H-8000 Székesfehérvár, Hungary
{beszedes.bertalan; szell.karoly; gyorok.gyorgy}@amk.uni-obuda.hu

Abstract: The production of highly reliable, electronic devices is a source for significant environmental emissions and energy consumption. A modern, cost-effective and modular design can enhance product maintainability and lifetime. Many end-users would certainly be willing to devote more resources (money) for a device they use, if, in return, they could extend the life of the device. This paper introduces the architecture for a high-reliability, modular, end-user-configurable, redundant power supply, based on these principles.

Keywords: redundant; robust; self-monitoring; embedded system; modular PSU; high reliable

1 Introduction

Despite the arrival of many renewable energy sources, the vast majority of the world's energy supply is still provided by fossil fuels, the extraction and use involves the release of large amounts of greenhouse gases into the atmosphere. Therefore, improving energy efficiency is currently also the most effective means of trying to fight climate change. While the world's energy efficiency is improving [1], so is the amount of wasted energy, decreasing. As a result of the growth of Earth's population and global economy, humanity's total energy consumption (global primary energy demand) continues to rise. And because of the predominance of fossil fuels, polluting energy sources in the global energy mix, it is also leading to an increase in greenhouse gas emissions, which have recently grown at the fastest pace since 2013.

Technological advances have made it possible to increase energy efficiency, which significantly reduces emissions while increasing energy use. Energy efficiency could be improved at a much higher rate, with technologies that are already available, but rarely used [2] [3].

According to the International Energy Agency (IEA), the potential is enormous, and by taking advantage of it alone, we could stop the rise of greenhouse gas emissions after 2020. However, according to the latest surveys, the world is moving further and further away from this goal [4] [5].

According to the International Energy Agency's Efficient World Strategy (EWS), at least a 3% improvement would be needed, each year, to meet global climate and sustainability goals. The 3% has only been realized once, in 2015 and the pace has slowed gradually, thereafter.

By using cost-effective technologies [6], the pace of energy efficiency improvement can be increased to a much higher level. Design principles like modular device design, maintainability and traceability can serve this goal. Digitalization and remote supervision systems are closely linked to these features [7].

The continuous development of end-user technological capabilities also supports the need for modular, easy-to-maintain designs [8]. There is a niche market for manufacturers to ensure user-friendliness and repair-ability of civil and industrial devices - in exchange for some extra cost. The authors hope that the modular design and the installation and repair manuals of the devices will come back into vogue.

In terms of robustness of such systems successful results have been achieved in the field of fault diagnosis and fault tolerant techniques [9] - [11]. Basically, we can define the following categories for fault diagnosis methods:

- Model-based [12]–[14]
  The outputs of the system-model and the outputs of the real system is compared
- Signal-based [15]–[18]
  A diagnostic decision is made based on the measured signal
- Knowledge-based [19]–[21]
  A large volume of historic data is needed
- Hybrid and active [22]–[24]
  Combination of the previous methods based on their advantages

The basic fault types are as follows [25] [26]:

- Actuator fault
- Sensor fault
- Plant fault

In the literature, several practical solutions can be found [27] [28].

2 Architecture

2.1. Power Supply Unit

2.1.1. Modular Power Supply Unit

Figure 1 shows a block diagram of a battery-powered modular power supply (PSU) controlled by a microcontroller (MCU). The external source of energy can be the electricity grid or renewable energy source. The battery charger is responsible for properly charging and discharging the energy storage unit. The function of a DC/DC converter is to ensure the voltage level of the battery or external power source to meet customer needs.

In case of AC external power source or a load that requires AC power supply, the proper battery charger and AC/DC, DC/AC or AC/AC converter has to be applied.

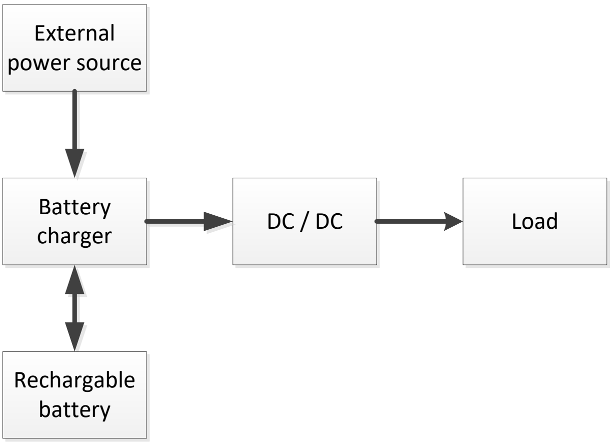


Figure 1
Modular power supply structure

2.1.2. Redundant Power Supply Unit

For single-redundant PSU (see Figure 2), if one of the PSU fails, the backup PSU takes over the task. This means that only one PSU is working at a time and that it supplies 100% of the required electricity for the powered system. In this case, the backup power supply is out of service or under test. This design ensures that the power supply is fault tolerant. This mode is also called, hot-stand-by mode.

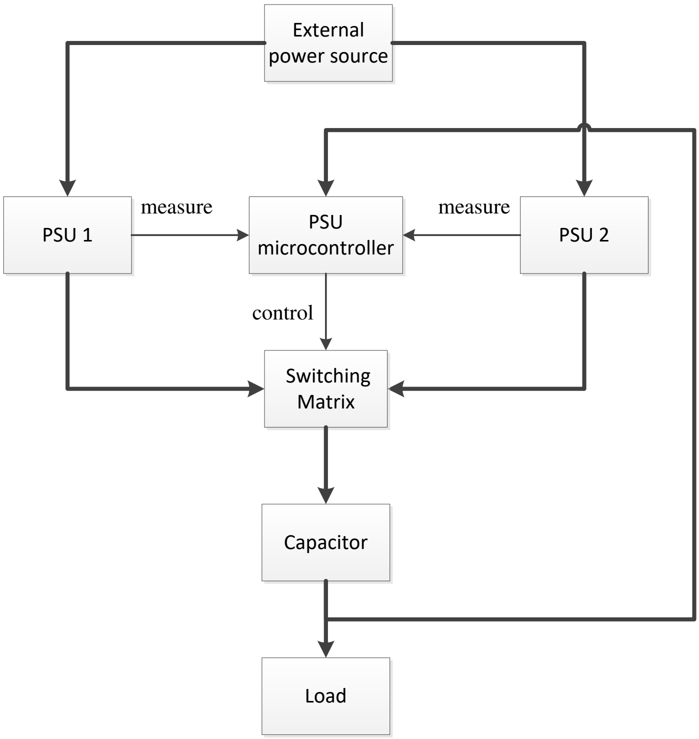


Figure 2
Redundant power supply structure with supervisor MCU

Another solution is the load-sharing mode, where power supplies share the load power. If there are more than two power supplies in the system and one is out of service for failure, replacement, or testing, the remaining power supplies will share the total load current equally.

For example, if there are four redundant PSUs in the power supply system and one of them, for the above mentioned reasons, goes out of service, the power supplies that are still in operation, will distribute the load, so, for example, the single units would provide 33% instead of 25% of the load current. All power supplies would be able to provide full load current if left alone in the power supply system.

2.1.3. Redundant Modular PSU

On the Figure 3, it can be seen, the capacitor supplies power to the microcontroller, which supervises the PSU, the connected load, and other optional modules, when the power supply is temporarily interrupted due to replacement of the redundant modules.

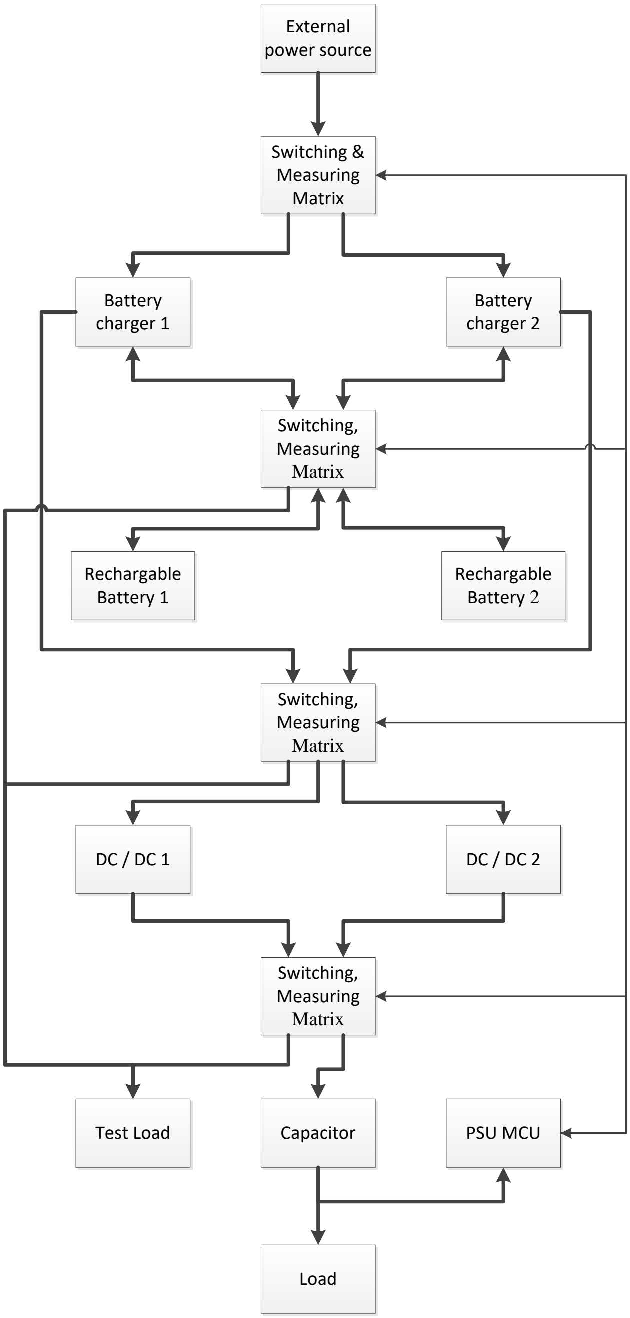


Figure 3
Simplified redundant modular power supply unit structure

2.1.4. Microcontroller

Power supplies usually include a microcontroller that affects the power supply and provides measurement, logging, communication, etc. functions as well.

In the case of modular power supplies, the power supply control microcontroller and the tightly-coupled components may be provided as separate modules or may be part of the motherboard.

For redundant power supplies or redundant modular power supplies, the microcontroller for power supply control extends to monitoring, testing, evaluating the power supplies or power supply modules, and controlling switching matrices.

2.1.5. Switching Matrix

The microcontroller also controls the switching matrices that connect the power modules. The function of the switching matrices is to provide energy flow between the power modules in a reconfigurable manner. The switching matrices can be used to connect and disconnect redundant power modules, including replacing redundant power modules.

When replacing power modules, switching multiple redundant power modules in a way that would lead to a short circuit should be avoided. The first step is to disconnect the currently active module, after which the redundant module can be turned on. The process results in a short-term power line break, but with the addition of energy storage devices (buffer capacitors), the power supply is continuously maintained, and transient-low switching is possible. In the experimental setup galvanically isolated relay modules were used. In case of the end product it is recommended to use modern technology based semiconductor switching elements [29]–[31].

2.2. Measuring Method

In redundant fault tolerant systems, some basic features are needed for proper operation. In hot-stand-by mode or load-sharing mode, the embedded monitoring system must be able to notify a supervisor and a control system of the actual status or failure of power supplies and modules.

The embedded monitoring system must be able to monitor power supplies and modules. Depending on the various aspects of error detection, this may occur using normal operational load or dummy load. Both the normal operation and the test operation must be measured with active and standby power modules. This can be done by swapping the power modules or by intermittently disabling them, as in a test procedure.

Redundant fault tolerant power supply systems must be provided with continuous power supply even when defective modules are replaced. The hot-plug-in function ensures smooth operation of the powered system.

2.2.1. Module Measurement

During the measurement of the modular power supply, the module's efficiency, temperature, input and output voltage and current values and waveforms must be monitored. The measured values should be transmitted to the microcontroller for further processing and storage.

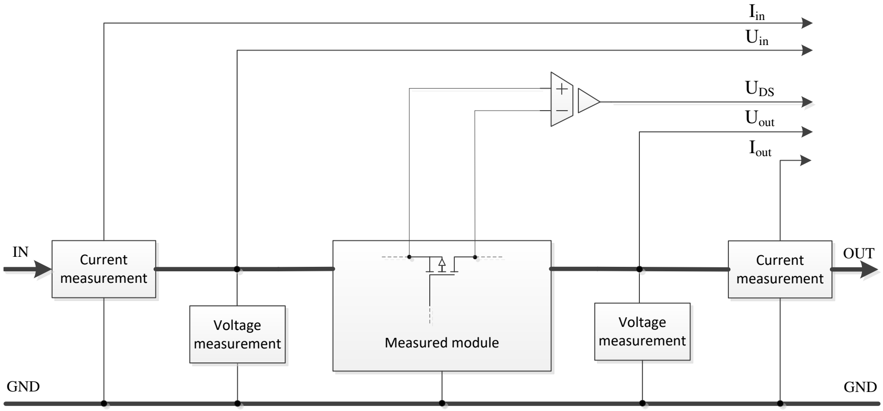


Figure 4
Module measurement scheme

Voltage measurement, if higher than the reference voltage of the analog digital converter of the microcontroller, is carried out by a voltage divider made of high precision and stable elements. There is even the possibility of using an optocoupler, but the brightness degradation of its built-in semiconductor LED, typically in the infrared range, can over time falsify the measurement.

Current measurement can be accomplished by measuring or calculating the differential voltage across Shunt resistors with lower cost. In order to reduce the number of test cables or to measure higher current values, it is also possible to use hall sensors, which have the disadvantage of higher costs.

For lifetime prediction, the voltage difference of the semiconductor drain-source shown in Figure 4 is also measured when the MOSFET is open. When the MOSFET is closed, the drain-source voltage can easily be higher than the input of the operational amplifier used to measure the voltage difference, so the measurement must be constructed with a galvanically isolated operational amplifier.

2.2.2. Analog Measurements

The measurement and evaluation of the power modules can be accomplished by using integrated hardware elements, which support the measurement. They can also detect overcurrent, overvoltage, voltage drop, voltage fluctuation, instability, voltage loss and other anomalies.

The configuration provides additional options for controlling the output voltage and for detecting common differences listed below:

- After a positive voltage spike, the output voltage returns to normal
- After a positive voltage spike, the output voltage will remain above normal (see Figure 5)
- The output voltage remains stably higher than normal, after a voltage surge
- After a negative voltage spike, the output voltage returns to normal (see Figure 6)
- After a negative voltage spike, the output voltage remains below normal
- The output voltage gradually decreases
- The output voltage fluctuates below normal level
- The output voltage remains stable below a normal level after a voltage drop (see Figure 7)
- The output voltage becomes unstable (see Figure 8)

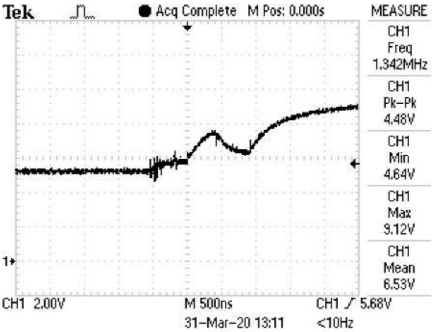


Figure 5
A positive voltage spike, and the output voltage will remain above normal level

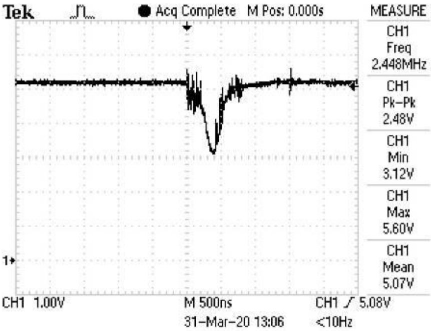
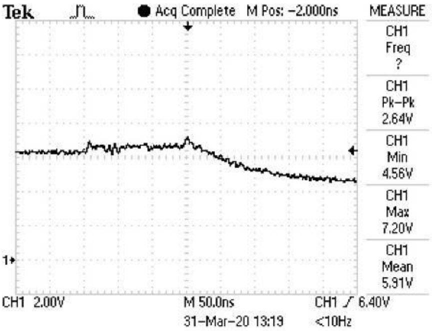


Figure 6
Negative voltage spike the output voltage will remain normal



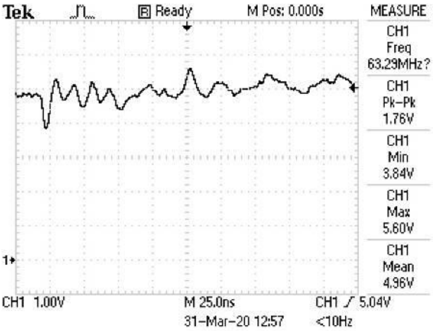


Figure 7
The output voltage remains stable below a normal level after a voltage drop

Figure 8
Unstable output voltage

2.2.3. Multiplexing Analog Lines

The microcontroller that monitors the power supply and controls the switching matrices has a limited number of terminals and is required to expand due to the large number of measurement and control signals. Extending the number of control outputs is easily accomplished with a serial I/O extender IC. The analog signals to be measured are coupled via an analog multiplexer to the ADC terminals of the microcontroller.

Choosing a more advanced microcontroller eliminates the need for external hardware, with sufficient software (multiplexing the input terminals to the ADC peripheral) to implement the solution presented. If the microcontroller has multiple internal ADC peripherals, it is recommended to measure the same analog signals with the same ADC modules - to avoid measurement errors due to differences in measurement peripherals. If the microcontroller has one internal ADC periphery, a sample and hold (SH) circuit is necessary to be used.

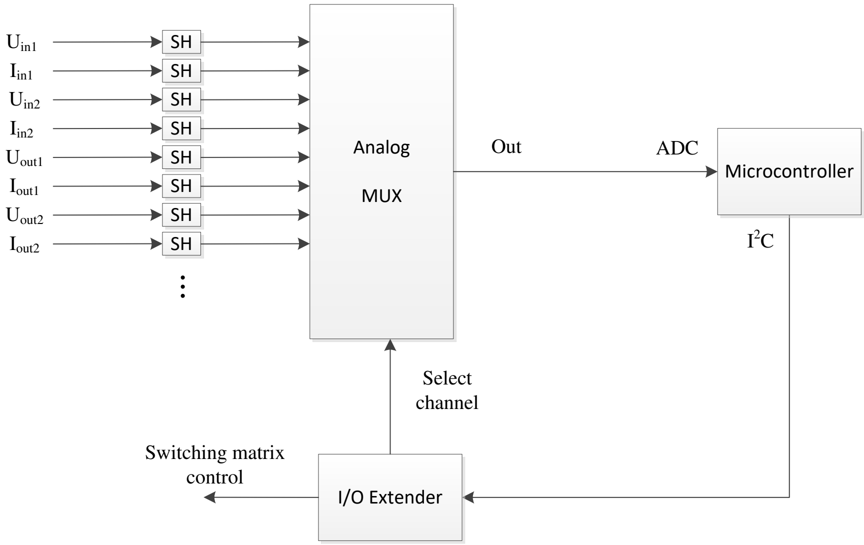


Figure 9

Hardware measurement and control scheme

2.2.4. External Interrupt Subsystem

Most of the time, the microcontroller controlling the power supply is in sleep mode. During sleep mode or when executing an instruction, you may not be able to detect unexpected voltage fluctuations in the power supply. Due to the architecture of the interrupt request system, which is designed as an external hybrid circuit, it is able to continuously monitor the output voltage state and, for example, to request an interrupt from the microcontroller in case of voltage fluctuation outside the specified limit.

Unlike the block diagram shown in Figure 1, the use of a built-in comparator as the internal periphery of the microcontroller controlling the power supply is recommended, in which case the reference voltage can be set as a register content by software. The internal comparator peripheral of the microcontroller also has the ability to request an interrupt (see Figure 10).

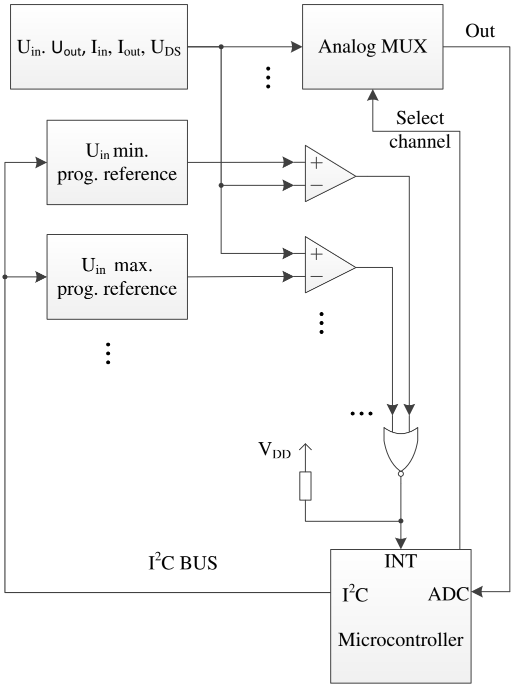


Figure 10

Block diagram of the external interrupt system

3 Realization

3.1. Main Program

The monitoring system calculates the primary and secondary power of the modules by measuring the input and output voltages and currents of the power supply modules. Based on these measurements, it calculates the efficiency of the power supply modules. If this value is below a predetermined level, or based on several measurements, it can be determined from the stored results that the condition of the unit is deteriorating (the efficiency value drops below a certain level), the monitoring system will send an error message to the monitoring system and jumps to a subroutine.

The monitoring system must process a significant amount of data. The cost-effective microcontroller-observed monitoring system can measure only one point at a time, and the analog-to-digital conversion takes a finitely long time. Power modules have relatively large number of measurement points. The frequency of each measurement and the accuracy of analog-to-digital conversion (measurement time) can be dynamically changed for efficient operation.

In case the error of a measurement point shows only a slight deviation from the ideal, it can be checked at a lower frequency and with less accuracy, so monitoring the measurement point takes only a small amount of time in the cycle. If the error of the measuring point increases, for example, on the basis of a look-up table, it is recommended to increase the frequency and the accuracy of the measurement. If the rate of change of the error at the measurement point increases, it is recommended to further increase the measurement frequency and accuracy of the measurement point to monitor the change. Measured and stored data can be used to perform fault prediction functions.

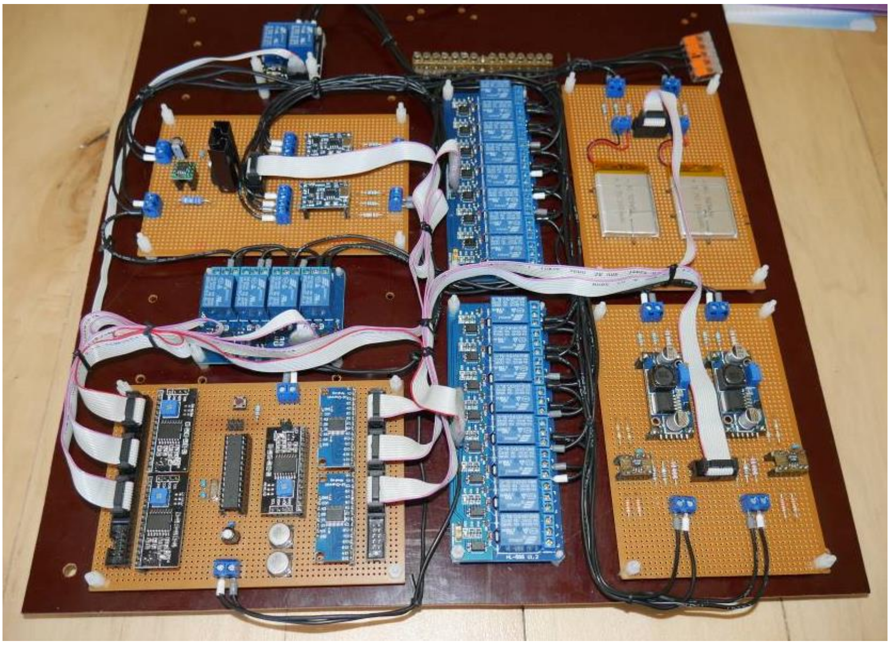


Figure 11
The model circuit

3.2. Operation Modes

The software supports user settings. The user has the ability to select the mode of operation and to weigh the following considerations (see Figures 11 and 12).

In case the user wants to maximize the life of the device, because maintenance is difficult, access to the device is difficult or the goal is to reduce the carbon footprint, you choose Swapping mode with reduced maximum charging and discharging currents. The swapping modules subroutine switches between redundant elements primarily based on the efficiency of the modules, but it can also change based on the temperature of the modules (the temperature of the active module increases), thus saving parts from heat stress and faster aging.

With advanced user settings, it is possible to fine-tune the above-mentioned parameters, customize reference levels, hysteresis values and timings.

If higher reliability is the main goal, it is recommended to activate Simultaneously running mode. In this case, the redundant modules run at 50-50% load. The heat load is higher than in the previous case. If one module fails, the other takes over 100% of the load with a smaller transient.

This type of control is recommended for powering easily maintainable equipment. In the case of a module failure, the failure of one of the modules increases the likelihood of a failure of the module remaining in the system, and in the case of the failure of the remaining 100% load module there are no spare modules in the system. The 100% load should only be tolerated by the module remaining in the system until maintenance, so we may use a lower power margin than in Swapping or Backup mode.

The third mode of operation is a classic Backup Mode. The module that builds the primary power supply will operate until it fails, after which the backup module will take over the load. In this case, the backup module may remain reliable for a long time due to higher performance margin compared to the Simultaneously running mode, although it does not include a spare element after the failure of the backup module.

This control mode is used to drive the primary active module till failure or till a predetermined degradation of efficiency, thus the long-term power consumption of this control mode will be the highest.

In all three cases, especially in the Backup Mode, it is important to periodically test the modules and determine their functionality. Measurements can be made with the programmatically variable load or with relatively short operation of the inactive module, until reaching the operating temperature, to perform measurements during normal operation.

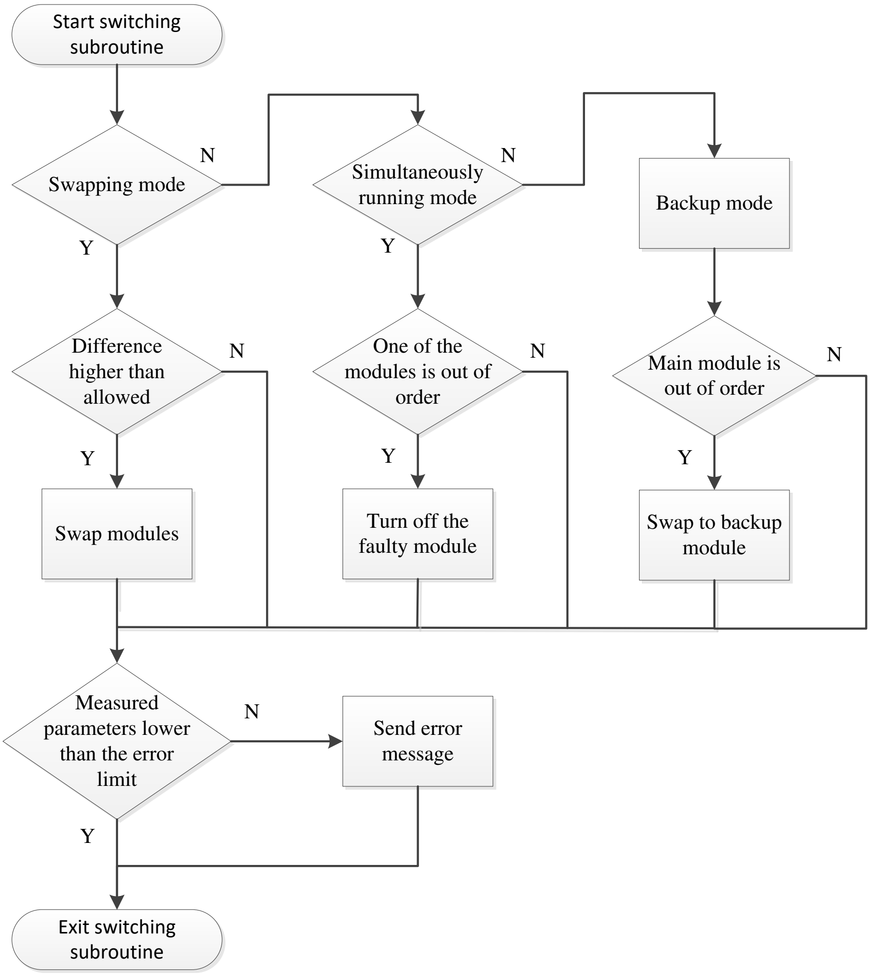

Start switching subroutine
Swapping mode
N
Y
Simultaneously running mode
N
Y
Backup mode
Difference higher than allowed
N
Y
One of the modules is out of order
N
Y
Main module is out of order
N
Y
Swap modules
Turn off the faulty module
Swap to backup module
Measured parameters lower than the error limit
N
Y
Send error message
Exit switching subroutine

Figure 12

Mode selector algorithm

Conclusions

The presented redundant power supply, greatly increases the reliability of the device. A cost-effective solution that monitors itself and a modular architecture that greatly enhances upgrades and maintainability, which can significantly increase the life of the equipment, thus, reducing global emissions/waste. The Authors believe that the presented system architecture, can be successfully implemented for both Civil and Industrial applications and especially for applications that demand a high level of reliability.

Acknowledgement

The Authors wish to thank the support to the Arconic Foundation. Bertalan Beszédes thankfully acknowledge the financial support by the ÚNKP-19-3 New National Excellence Program of the Ministry for Innovation and Technology, GINOP-2.2.1-15-2018-00015 and GINOP-2.2.1-15-2017-00098 projects. Károly Széll thankfully acknowledge the financial support of this work by the Hungarian State and the European Union under the EFOP-3.6.1-16-2016-00010 and GINOP-2.2.1-15-2017-00073 projects.

References

[1] M. Kolcun, A. Gawlak, M. Kornatka, and Z. Čonka, "Active and Reactive Power Losses in Distribution Transformers," *Acta Polytech. Hung.*, Vol. 17, No. 1, pp. 161-174, 2020, doi: 10.12700/APH.17.1.2020.1.9

[2] "Energy Efficiency 2019," IEA, Paris, 2019. Accessed: Mar. 26, 2020 [Online] Available: https://webstore.iea.org/download/direct/2891

[3] "Energy Efficiency Indicators 2019," IEA, Paris, 2019. Accessed: Mar. 26, 2020 [Online] Available: https://webstore.iea.org/download/direct/2707?fileName=Energy_Efficiency_Indicators_2019_Highlights.pdf

[4] *World Energy Balances 2019*. Paris: IEA, 2019

[5] *World Energy Statistics 2019*. Paris: IEA, 2019

[6] G. Györök, *Programozható analóg áramkörök mikrovezérlő környezetben*, Vol. 1, Székesfehérvár: Óbudai Egyetem, 2013

[7] É. Hajnal, "Big Data Overview and Connected Research at Óbuda University Alba Regia Technical Faculty," in *AIS 2018 - 13th International Symposium on Applied Informatics and Related Area*, 2018, pp. 1-4

[8] T.-M. I. Băjenescu and M. I. Bazu, *Reliability of Electronic Components: A Practical Guide to Electronic Systems Manufacturing*. Berlin Heidelberg: Springer-Verlag, 1999

[9] Z. Gao, C. Cecati, and S. X. Ding, "A Survey of Fault Diagnosis and Fault-Tolerant Techniques—Part I: Fault Diagnosis With Model-Based and Signal-Based Approaches," *IEEE Trans. Ind. Electron.*, Vol. 62, No. 6, pp. 3757-3767, Jun. 2015, doi: 10.1109/TIE.2015.2417501

[10] Z. Gao, C. Cecati, and S. X. Ding, "A Survey of Fault Diagnosis and Fault-Tolerant Techniques—Part II: Fault Diagnosis With Knowledge-Based and Hybrid/Active Approaches," *IEEE Trans. Ind. Electron.*, Vol. 62, No. 6, pp. 3768-3774, Jun. 2015, doi: 10.1109/TIE.2015.2419013

[11] W. Kong, Y. Luo, Z. Qin, Y. Qi, and X. Lian, "Comprehensive Fault Diagnosis and Fault-Tolerant Protection of In-Vehicle Intelligent Electric

Power Supply Network," *IEEE Trans. Veh. Technol.*, Vol. 68, No. 11, pp. 10453-10464, Nov. 2019, doi: 10.1109/TVT.2019.2921784

[12] J. Hu, J. Wang, J. Zeng, and X. Zhong, "Model-Based Temperature Sensor Fault Detection and Fault-Tolerant Control of Urea-Selective Catalyst Reduction Control Systems," *Energies*, Vol. 11, No. 7, p. 1800, Jul. 2018, doi: 10.3390/en11071800

[13] P. Szcześniak, G. Tadra, and Z. Fedyczak, "Model Predictive Control of Hybrid Transformer with Matrix Converter," *Acta Polytech. Hung.*, Vol. 17, No. 1, pp. 25-40, 2020, doi: 10.12700/APH.17.1.2020.1.2

[14] L. R. Neukirchner, A. Magyar, A. Fodor, N. D. Kutasi, and A. Kelemen, "Constrained Predictive Control of Three-PhaseBuck Rectifiers," *Acta Polytech. Hung.*, Vol. 17, No. 1, pp. 41-60, 2020, doi: 10.12700/APH.17.1.2020.1.3

[15] H. Chen and S. Lu, "Fault Diagnosis Digital Method for Power Transistors in Power Converters of Switched Reluctance Motors," *IEEE Trans. Ind. Electron.*, Vol. 60, No. 2, pp. 749-763, Feb. 2013, doi: 10.1109/TIE.2012.2207661

[16] N. M. A. Freire, J. O. Estima, and A. J. Marques Cardoso, "Open-Circuit Fault Diagnosis in PMSG Drives for Wind Turbine Applications," *IEEE Trans. Ind. Electron.*, Vol. 60, No. 9, pp. 3957-3967, Sep. 2013, doi: 10.1109/TIE.2012.2207655

[17] M. Shahbazi, E. Jamshidpour, P. Poure, S. Saadate, and M. R. Zolghadri, "Open- and Short-Circuit Switch Fault Diagnosis for Nonisolated DC–DC Converters Using Field Programmable Gate Array," *IEEE Trans. Ind. Electron.*, Vol. 60, No. 9, pp. 4136-4146, Sep. 2013, doi: 10.1109/TIE.2012.2224078

[18] A. M. Stanisavljević, V. A. Katić, B. P. Dumnić, and B. P. Popadić, "A Comprehensive Overview of Digital Signal Processing Methods for Voltage Disturbance Detection and Analysis in Modern Distribution Grids with Distributed Generation," *Acta Polytech. Hung.*, Vol. 16, No. 5, Aug. 2019, doi: 10.12700/APH.16.5.2019.5.8

[19] X. Xiang, C. Yu, and Q. Zhang, "On intelligent risk analysis and critical decision of underwater robotic vehicle," *Ocean Eng.*, Vol. 140, pp. 453-465, Aug. 2017, doi: 10.1016/j.oceaneng.2017.06.020

[20] R. Ghimire, C. Zhang, and K. R. Pattipati, "A Rough Set-Theory-Based Fault-Diagnosis Method for an Electric Power-Steering System," *IEEEASME Trans. Mechatron.*, 2018, doi: 10.1109/TMECH.2018.2863119

[21] P. Lezanski and M. Pilacinska, "The dominance-based rough set approach to cylindrical plunge grinding process diagnosis," *J. Intell. Manuf.*, Vol. 29, No. 5, pp. 989-1004, Jun. 2018, doi: 10.1007/s10845-016-1230-1

[22] A. Soualhi, G. Clerc, and H. Razik, “Detection and Diagnosis of Faults in Induction Motor Using an Improved Artificial Ant Clustering Technique,” *IEEE Trans. Ind. Electron.*, Vol. 60, No. 9, pp. 4053-4062, Sep. 2013, doi: 10.1109/TIE.2012.2230598

[23] B. M. Ebrahimi, M. Javan Roshtkhari, J. Faiz, and S. V. Khatami, “Advanced Eccentricity Fault Recognition in Permanent Magnet Synchronous Motors Using Stator Current Signature Analysis,” *IEEE Trans. Ind. Electron.*, Vol. 61, No. 4, pp. 2041-2052, Apr. 2014, doi: 10.1109/TIE.2013.2263777

[24] J. K. Scott, G. R. Marseglia, L. Magni, R. D. Braatz, and D. M. Raimondo, “A hybrid stochastic-deterministic input design method for active fault diagnosis,” in *52nd IEEE Conference on Decision and Control*, Dec. 2013, pp. 5656-5661, doi: 10.1109/CDC.2013.6760780

[25] Y. Song and B. Wang, “Survey on Reliability of Power Electronic Systems,” *IEEE Trans. Power Electron.*, Vol. 28, No. 1, pp. 591-604, Jan. 2013, doi: 10.1109/TPEL.2012.2192503

[26] S. Yang, A. Bryant, P. Mawby, D. Xiang, L. Ran, and P. Tavner, “An Industry-Based Survey of Reliability in Power Electronic Converters,” *IEEE Trans. Ind. Appl.*, Vol. 47, No. 3, pp. 1441-1451, May 2011, doi: 10.1109/TIA.2011.2124436.

[27] G. Györök, “The FPAA realization of analog robust electronic circuit,” in *2009 IEEE International Conference on Computational Cybernetics (ICCC)*, Jan. 2009, pp. 179-183, doi: 10.1109/ICCCYB.2009.5393941

[28] P. Holcsik, J. Pálfi, Z. Čonka, and M. Avornicului, “A Theoretical Approach to The Implementation of Low-Voltage Smart Switch Boards,” *Acta Polytech. Hung.*, Vol. 16, No. 4, Jul. 2019, doi: 10.12700/APH.16.4.2019.4.7

[29] G. T. Orosz, A. Sulyok, G. Gergely, S. Gurbán, and M. Menyhard, “Calculation of the Surface Excitation Parameter for Si and Ge from Measured Electron Backscattered Spectra by Means of a Monte-Carlo Simulation,” *Microsc. Microanal.*, Vol. 9, No. 4, pp. 343-348, Aug. 2003, doi: 10.1017/S1431927603030241

[30] G. T. Orosz *et al.*, “Experimental determination of electron inelastic scattering cross-sections in Si, Ge and III-V semiconductors,” *Vacuum*, Vol. 71, No. 1, pp. 147-152, May 2003, doi: 10.1016/S0042-207X(02)00729-7

[31] G. Gergely *et al.*, “Surface excitation correction of the inelastic mean free path in selected conducting polymers,” *Appl. Surf. Sci.*, Vol. 252, No. 14, pp. 4982-4989, May 2006, doi: 10.1016/j.apsusc.2005.07.017