Video Mining in Basketball Shot and Game Analysis

Laszlo Ratgeber*, Zdravko Ivankovic, Zoran Gojkovic***, Zoran Milosevic****, Branko Markoski *****, Anja Kostic – Zobenica ******

*University of Pécs, - Faculty of Health Sciences, Ret u. 4, 7623 Pécs, Hungary

** Preschool Teacher Training and Business Informatics College of Applied Studies "Sirmium", Zmaj Jovina 29, 22000 Sremska Mitrovica, Serbia

*** University of Novi Sad, Faculty of Medicine, Clinical center of Vojvodina, Clinic for orthopedic surgery and traumatology Novi Sad, Hajduk Veljkova 3, 21137 Novi Sad, Serbia

**** University of Novi Sad, Faculty of Sport and Physical Education, Lovcenksa 16, Novi Sad, Serbia

*****University of Novi Sad, Technical faculty "Mihajlo Pupin", Djure Djakovica BB, 23000 Zrenjanin, Serbia, markoni@uns.ac.rs

****** University of Novi Sad, Faculty of Techical siences, Trg Dositeja Obradovica 6, Novi Sad, Serbia

Abstract: The aim of this study is to analyze the footage of basketball games presented to viewers on television. It includes a wide range of activities from identifying players, determining their position, recognizing the ball, hoops, as well as analyzing the shots and determining shot efficacy. The player detection is based on mixture of non-oriented pictorial structures. The detection of body parts is performed by the Support Vector Machines (SVM) algorithm. This paper contains algorithms for detecting player positions of the court, ball position detection and determination of shot. It is achieved by detecting court position and applying spatial transformation. It also includes detection of shot, detection weather shot was successful and position from which shot was taken. All algorithms are tested in large number of frames from different basketball games.

Keywords: video mining; basketball; shot recognition

1 Introduction

The digital analysis of multimedia content is a steady growing technology, constantly progressing in recent years. The aim of this study is to analyze the footage of basketball games viewed through television stations. It includes a wide range of activities from identifying players, determining their position, recognizing the ball and hoops, as well as analyzing the shot and determining its efficacy.

The analysis process starts with the footage of a basketball game. Frames are extracted from the footage, where, according to the standard, there are thirty frames per second. During the analysis, it was concluded that it was not necessary to process every single frame. In this paper, every third frame, or ten frames per second, is used. In this way, the time of execution is significantly reduced without losing the accuracy, since no event does happens so fast that it was not recorded within the reduced set of frames. Such set represents the starting point for the application of algorithms that will allow analysis of the match and the collection of basic basketball parameters.

The identification of the hoops and the ball is done on the basis of the dominant color. The image is translated into HSV color model. Then, areas that have color in a given range are determined, based on the mean value for the color of the ball, or the hoops, from several frames. The areas created in this way are further analyzed in relation to the size (number of pixels) and the position, in order to introduce restrictions that will allow them to be more accurately identified.

An analysis of the position of the hoop and ball allows a further reduction of the set of images that will be further processed. The goal is to extract the shots to the basket, precisely from the moment of sending the ball to the moment when the ball passes through the hoop or until it bounces from it after an unsuccessful shot. Determining the position of the shot means to recognize the position of the player shooting the ball to the basket. In this research, identifying players and determining their position is done only at this single moment. In such a way, this process, which is the most time-consuming, is executed in a relatively small number of frames.

All the players on one team wear the same shirts, which represents a mitigating factor, from a detection point of view. However, there is a question which level of detection is needed. Basketball players can be treated as rectangular areas in the image where the head, torso, and limbs enter the observed rectangle. Understanding their behavior is reduced to the pure tracking of the rectangle as it moves through the image. This kind of approach in basketball can be interesting, because it enables us to monitor the actions of the team plays, in offence, as well as, the individual moves of the players depending on whether they play an offence or defense, and depending on the type of offence/defense they are playing. The disadvantage of this kind of display is the fact that a large number of data is

ignored. Players play defense in different ways (with lowered or raised arms), have different techniques of shooting, dribbling a ball, approaching the basket, etc. This work is focused on a kinematic detection in which the positions of the limbs are modeled at all times. A practically applicable algorithm should fulfill a certain set of requirements. It should simultaneously recognize several people, be resistant to partial blocking of the view and be efficient in computing. It is also necessary, not to be based on any backdrop-reducing techniques, since it is often necessary to watch a basketball player standing still, while the camera is moving.

In this paper, the aim is to detect players in basketball games, where broadcasting is mostly done in HD format. The algorithm of the mixture of non-oriented picture structures is applied over the pyramid of characteristics obtained by scaling and applying the HOG algorithm over the base image. By analyzing the games, we have come to the conclusion that only a part of the set of the scales used in the work by [1-2] is sufficient. This led to a time savings of almost 30%, without reducing the accuracy of the detection.

After identifying players, they need to be attributed to their associated teams. This is done by using the k-means clustering technique whose input is the shirt color saturation histograms, where each cluster is representing one team. In our case, a trained algorithm (trained for the color of the shirts of referees) can be used in all matches within the league, which is a significant reduction in the time required to prepare the algorithm for a new match.

In addition to detecting the players, this work also includes the collection of basic statistical parameters during the game itself. These parameters include shots to the basket, with differentiation of shots from different positions. The court itself is divided into eleven positions, which is the most common division in basketball, with wide application in a large number of competitions. In order to determine the positions from which the shot is made, the recognition of the court boundaries is expanded by determining the restricted zone (colored area under the basket). Analyzing the position of the ball in the adjacent frames is performed, providing information about the shot attempt and whether the shot was successful or not. Determining the position from which the shot was sent is done by means of a spatial transformation determined on the basis of detected court boundaries and actual court dimensions. The applied solution is quite robust and efficient, as shown by the experimental results obtained by applying the algorithm over the actual footage of basketball games.

2 Review of Relevant Research

Extracting information from video content is a very important current field of scientific research, especially in sporting events. The aim is to automatically harvest information from the recordings of sports matches concerning the actions

played, the successful and unsuccessful shots, the positions from which they were shot, the performance of individual players, etc. These topics are dealt with in a large number of scientific papers [3-7].

Through the development of technology, sports events have become available in digital form. By using a large number of tools that have been previously developed for text search, video and multimedia content search is becoming more and more common in sports. Automated game review methods are used to parse video content and translate it into a searchable form [8-9].

According to [8] the application of data mining in sports will face a number of challenges and obstacles over the next few years. The biggest obstacle will be overcoming the long-standing opposition, by certain sports organizations members, who are advocating the traditional way of acquiring knowledge. The same authors state that the application of data mining in sport is at a turning point and that a large number of possibilities are just waiting to be used. Some of these options will quickly lead to the desired results; while others it will take years or even decades. They also point out that the basic task is not to find the right way to collect data, but to determine which data should be collected and how to use them in the best way possible.

Piatetsky-Shapiro [10] points out that although statistical techniques are at the heart of data mining, they are used to distinguish between templates and other objects of interest such as the movements and tendencies of opponent players, in contrast to the noisy and useless data, enabling researchers and sports organizations to test hypotheses and make predictions based on the results obtained. Statistics by itself, does not explain relations; this is the purpose of data mining. Within a statistical research project, data mining evolved as a method for finding the reasons behind the relations. Using statistics, it is possible to find and measure of the strength of relations between variables. However, this statistical measure is not able to explain why the relation itself exists or what impact it may have in the future. Data mining provides tools for testing data and gathering further knowledge about cause and consequence relations. This is possible through interactive, iterative and /or research data analysis.

While the use of statistics in the decision-making process is definitely an improvement over the use of the instincts of a coach, manager or scout, statistics alone can easily go in the wrong direction without knowing the domain of the problem. The first part of the problem is determining the performance metric. A large number of existing sports metrics can be very easily used in an unsuitable way. A typical example of inaccuracy in data collection in basketball was given by [11] and cites an example of a rebound that represents how many times the player on defense acquired the ball after an unsuccessful opponent's shot. In order to record a rebound, teammates must block the opponent's players and keep them away from the basket. When blocking opponents, those basketball players are usually not able to catch the ball. However, their defense game makes them

equally important, in order to capture the ball. Regarding the way in which rebounds are recorded, it is clear that only a player who takes the ball is "rewarded" by rebound. The second part of the problem is finding interesting templates within the data. These templates can include the movements and tendencies of the opponent's team/player, detecting the start of the injury by monitoring the quality of the training, or predicting outcomes based on previous matches.

Data mining includes procedures for detecting hidden templates and creating new information from data storage. The storage may include well-structured and defined databases, such as statistical reports, or unstructured data in the form of video footage of whole games or some typical segments. Data mining activities, tools, technologies and human control are the essence of an area called "knowledge management" [12]. Knowledge management can give an organization an advantage over competition [13], and in particular a method for maintaining the continuity of knowledge in the organization. However, before raw data comes to the state of useful knowledge, it is necessary to examine all the levels among the data and knowledge as required by the DIKW (Data - Information – Knowledge - Wisdom) hierarchy [14]. DIKW hierarchy is a widely accepted concept on knowledge management stages. Every next level: data, information, knowledge and wisdom, builds on previous levels and provides increased awareness of the environment in which one can find meaning [15-16]. The DIKW principle contains a phase that serves to differentiate data from knowledge and to set the final limits that determine what are data, information, and knowledge.

2.1.2 General Approaches in Analyzing Sports Events

In object recognition and analysis of sporting events discrimination methods are dominating. Discriminatory training methods determine the model's parameters in order to minimize detection of algorithm errors over a set of training images. Such approaches directly optimize the decision-making limit between positive and negative examples. This is one of the reasons for the success of simple models trained by discriminating methods, such as Viola-Jones [17] and Dalal-Triggs [18] detectors. It is quite harder to train discrimination in partial modeling, although there are certain approaches [19-22].

One of the first solutions, that were successfully implemented in the detection and monitoring of players in sports games, was the BPF (Boosted Particle Filter). Okuma et al. [23] have used BPF to track players in hockey games. Cai et al. [24] have expanded BPF by introducing a two-part matching in order to link detection to target objects.

Our approach is very similar to [7], as they also deal with identifying basketball players from video material broadcasted via television. The difference is that they deal with the identification of players on the ground using the CRF (Conditional Random Fields) algorithm. After identifying objects that represent players, they

are trying to recognize who the player is, while our goal is to identify the position at which the player currently is. [6] are also committed to identifying basketball players with footage intended for TV viewers, but their principle is mainly based on the detection of the dominant color representing the court, while players are detected as objects of the different color on the court. The aim of their research is to recognize the situation when the player in the attack got an open position for a shot because no good defense was played against him. Lifang et al. [25] have created an algorithm by which it is possible to recognize on the footage from a basketball game whether the observed frame represents a shot from the default camera (a camera that records the side of the court at which the game is currently played) or whether it is a frame from one of auxiliary cameras (camera under the basket, on the backboard itself, etc.). Huang et al. [26] applied a SVT (Support Vector Tracking) algorithm in the analysis of basketball games in order to track the ball, players and mesh of the hoops. SVT integrates the SVM classifier along with optical stream-based monitoring. By using these detections, they are able to determine what is currently happening on the ground using BBN (Bayesian Belief Network). Different types of shots (close range, intermediate distance and long distance) are recognized, as well as the event of scoring. Zhu et al. [3] use an audio signal in addition to video one in the detection of events at basketball games. They analyze the applause of the audience and the referees' whistle to correlate some semantic clues.

Alahi et al. [5] deal with detecting and tracking players in basketball matches using a large number of synchronized cameras covering each part of the court. The player is monitored with each camera in particular, but also in the 3D environment. The presented algorithm is based on the spatial approximation of the points representing the player's locations. This research has a similar goal as ours, but the principle of research itself is different because of additional cameras that significantly facilitate the determination of the player's position. Daniyal et al. [27] also use multiple synchronized cameras to define the characteristics at both the object and frame level. Objects are detected using an algorithm that monitors color change, while information about the paths for each object is generated using a multi-frame matching technique. At the frame level, the total activity is considered, as well as, the probability that it is one of the defined events, the number of objects and the total score for all objects. These characteristics are used to obtain the total result using a multivariate Gaussian distribution. The best view camera is selected using the DBN (Dynamic Bayesian Network) algorithm. Perše et al. [28] use two cameras attached to the ceiling of the hall where the basketball match analysis is carried out. With these cameras they identify players and their trajectories. Based on the trajectory, they distinguish three types of game: defense, offence and time-out. After that, they are trying to do an analysis of what is currently happening on the floor, using elements such as starting formations, blocks and gestures.

Theron and Casares [29] also deal with analyzing the movements of basketball players in the court. For the purpose of monitoring, GPS (Global Positioning System) devices are used, which give the right position of the player in the near real time. The main goal of their work is the statistical and kinematic monitoring of the players due to physical activity during the match.

Wu et al. [30] examined the process of image reduction in the example of basketball, so it could successfully be displayed on devices such as mobile phones or tablets. The goal is to determine which parts of the image are important (such as the court) in order to display them in the highest resolution, while the less important parts (such as the audience) are rejected. The determination of the court is based on the dominant color principle.

3 Shots Analysis

As part of our previous research [31], the process of separating frames representing a shot to the basket and recognizing all players is explained. This paper presents algorithms that allow the completion of the process of collecting the basic statistical parameters. For recognized players, the exact position on the court is determined. Among them one player stands out who is shooting the ball to the basket and his exact position is given as the position of the shot. All players are attributed to their teams depending on the color of the shirt, in order to have statistics in relation to the team. The final stage is to determine the accuracy of the shot, i.e. whether the ball has passed through the hoop or not.

3.1 Determination of the Player's Position

When all players are detected, it is necessary to determine their current position on the court. The effect of shots from different positions on the team game was described in our previous research [32].

The division of the court into positions can be done in several ways. In professional literature related to basketball, it can be noted that there is no general solution to this problem. The reason can be found in the fact that the effectiveness of a player is generally observed in relation to their position. For this reason, a different division of court should be created for the canter, guard, forward, etc. Perše et al. [28] have created their own division of court in 14 areas using k-means clustering over recorded player positions. In our study we used a division that is used in most competitions: the court itself is divided into 11 areas.

3.2 Determination of Spatial Transformation

Taking into account the actual dimensions of the court in the NBA basketball league and five points recognized on the court (four points on the paint corners and one point in the visible court corner), we can determine a function that represents a spatial transformation. This spatial transformation performs the projection of the point $p = (x, y, 1)^T$ from the image coordinates to the point $p' = (x', y', 1)^T$ which is located within the real court coordinates. A pair of points (p, p') has the relation $p' = Hp$ where H represents a 3x3 transformation matrix having the following form:

$$H = \begin{bmatrix} a_1 & a_2 & b_1 \\ a_3 & a_4 & b_2 \\ c_1 & c_2 & 1 \end{bmatrix} \tag{1}$$

The previous equation has the following elements:

- $\begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix}$ represents a matrix determining the type of transformation to be applied: scaling, rotation, etc.
- $\begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$ represents a translation vector which simply moves the points
- $\begin{bmatrix} c_1 & c_2 \end{bmatrix}$ represents the projection vector.

The transformation of a point with x, y coordinates into a point with x', y' coordinates may be noted as follows:

$$\begin{bmatrix} a_1 & a_2 & b_1 \\ a_3 & a_4 & b_2 \\ c_1 & c_2 & 1 \end{bmatrix} \times \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} \tag{2}$$

Transformation by projection shows how the object changes when the angle of view is changed. This transformation allows the creation of a perspective distortion. To determine the parameters in the matrix, reference points were used in the corners of the restricted area and in cross-sections of the horizontal and vertical court boundaries of the terrain, as well as actual court measurements.

By determining the spatial transformation, we have actually set a function that will enable us to determine the exact position on the court at any point in the image. In this way we also determine the position of basketball players. In [28], observation of the court is performed with special cameras placed on the ceiling of the facility.

Their use does not require any mapping, as the position of players on the court can be directly determined. More cameras are also used in [5], [27], while GPS coordinates are used in [29]. All of this greatly facilitates the process of determining the position of players in relation to our research topic.

Before determining the position of players in the court, it is necessary to determine the position of the players in the picture. In the work [31], we explained how to detect players. However, we now need one point per player that will represent the place where he is located on the court. The X coordinate of this point is determined using the rectangle that marks the entire player. Our X coordinate is located in the middle between the two X coordinates representing the left and right edges of the rectangle. The Y coordinate should represent the feet of the basketball player, because the required point is actually the place where the basketball player is standing on the court. For this reason, we took into account the rectangles representing the ankles of basketball players (areas 14 and 26) to determine this value. The requested Y coordinate is the mean Y-coordinate of the lower edges of these rectangles. Player detection and position determination is shown in Figure 1.

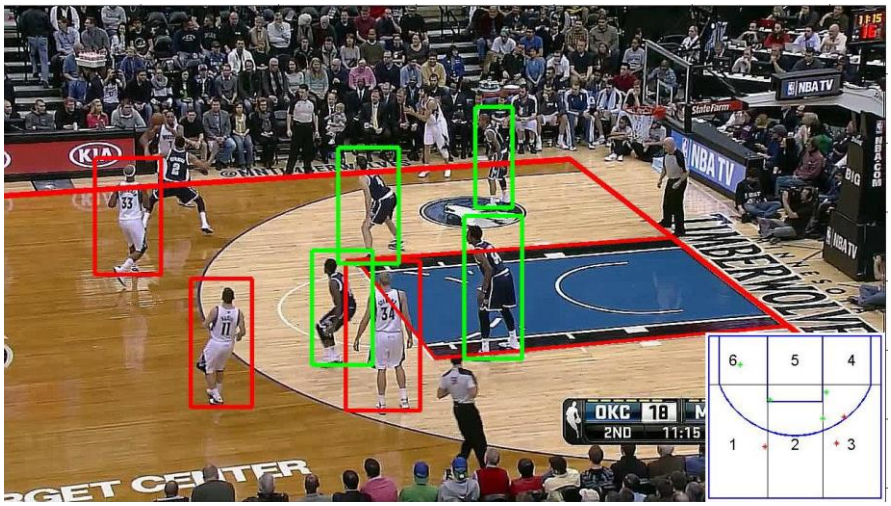


Figure 1
Determining positions of detected players

3.3 Recognizing the Position at the Moment of the Shot

The first step is to determine the position of the ball. By analyzing the video as explained in the previous chapters, frames are selected from the moment of sending the ball to the basket until the moment when ball passes through the hoop or bounces off the hoop. The moment of the shot represents the first frame in this

series, so it is necessary to determine in it the position of the player who sends the shot. The requested player is the player who is the closest to the ball and the analysis starts from determining the position of the ball.

To locate the ball, the image is first transferred to the HSV color model. After that, the area on the image containing the color in the required range is determined. If several areas contain the required color, their analysis is made in relation to the position and size, on the basis of which the ball is located.

The next step is to determine the size of the area within which players will be recognized. In this study, the size of the area is 225 × 425 pixels, with the starting point being 75 pixels left (right) and upwards relative to the ball position. Within these areas, several players can be identified. The goal is to determine one player, who is shooting the ball to the basket. In order to determine it, an algorithm consisting of two steps is used:

1) All objects whose head is located between the ball and the hoop are rejected
2) The object sought is the object whose head is the closest to the ball

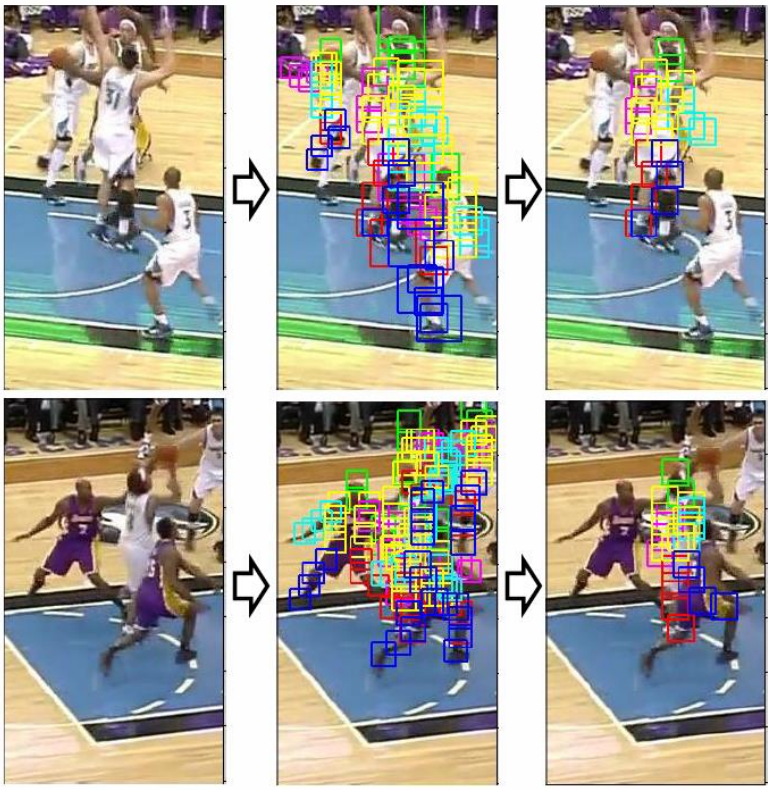


Figure 2
Recognition of a player shooting the ball

An example of the execution of this algorithm is shown in Figure 2. From the pictures it can be seen that from a large number of potential detections, the algorithm determines one based on the previous limitations.

After identifying a player, it is necessary to determine his location in relation to the court. The point that represents his location is between his feet. By applying spatial transformation, this point is transferred to the coordinates of the court.

3.4 Attribution of Players to Teams

After the completion of the detection and removal process of the unattainable objects, all recognized objects represent basketball players. Among them are the players of both teams and in order for further processing, it is necessary to detect which player belongs to which team.

The first step is the selection of the clustering area. This is actually the selection of those parts that are covered by the shirt and shorts. Based on these parts, one area that contains them is created and it will be used in further processing.

After creating the area, it is translated into HSV color space for two reasons. Firstly, because of the fact that the H (hue) component of this color system represents a color. By determining the range that represents the parquet and player's skin, these pixels can be removed from further processing. The other reason is that each team in the championship has two sets of shirts. One set is "light" and the other is "dark". Both sets are created according to the colors of the club, but also in accordance with the aforementioned rule. In any match, the teams are in different types of shirts (one is in a light set and the other in a dark one). If the S (saturation) component from the HSV model is observed, it can be noted that the light set has a low level of saturation, while the dark set has a high level of saturation.

A 100-bit histogram is calculated, from which the five peaks with the highest value are then selected. A similar principle, in determining the color of the court, was used by Wu, et al. (2012), but in this case, the H component was used and one or two of the most influential peaks.

Five peaks from the histogram of each player enter the process of clustering by the k-means principle. Objects belonging to the first cluster represent the players of one team, and objects belonging to the other cluster represent the players of the other team. By applying this approach, it is not necessary to re-train the algorithm for each new team and the new color of the shirts. An example of the separation of players by teams is shown in Figure 3.

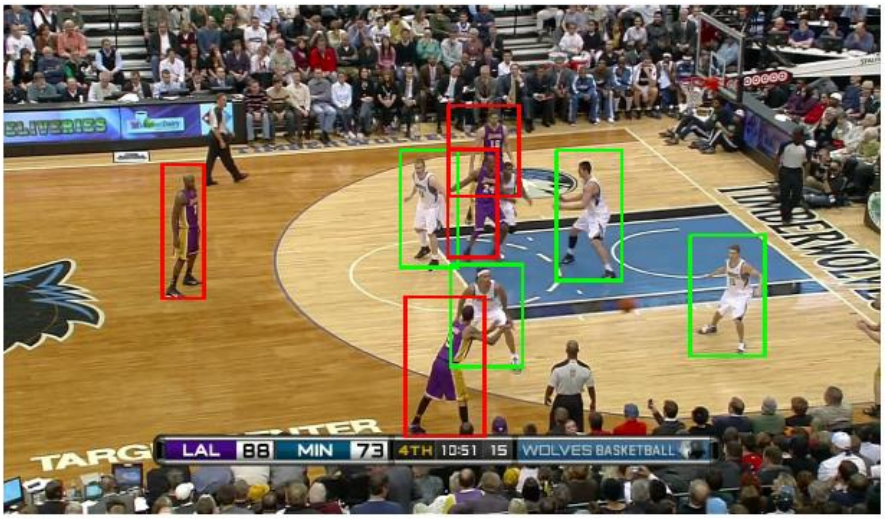


Figure 3
Separation of players in teams

4 Experimental Results

Testing was done over a quarter of a basketball game. During a quarter, the algorithm takes into account over 10,000 frames, so this period is quite sufficient to evaluate the performance of algorithms, such as determining the frames representing the court, the game under the basket, or the shot to the basket. The algorithm for frame extraction is trained specifically for each new game, and the results obtained over one quarter can be expected in other matches too. The identification of the ball and the hoop is based on their color, which is identical in all matches. Determining the accuracy of the shot depends only on the recognition of the ball and the hoop, so here again, the same results can be expected in other matches.

Some algorithms, such as those for identifying the court boundaries, identifying players, separating players by teams and determining their position, are not trained specifically for each new match. The color of the court and the marking line, as well as the color of the player's shirts, vary in different matches. For the purpose of realistic assessment of these algorithms, their testing was carried out on a special set consisting of ten different games. Each game had ten frames selected, which means that the algorithms have been tested over a set of hundred frames. Within the set, a total of 973 players were shown, so 973 players needed to be identified and their position determined, as well as the team they are playing for.

During the determination of the court boundaries, a total of five lines were determined (horizontal and vertical court boundaries, as well as two horizontal and one vertical restriction zone boundaries), so in this set it was necessary to recognize total of 500 lines.

Test sets used one quarter of a basketball game and a hundred frames from ten different games which are sufficient to evaluate the actual capabilities of the created algorithms, so there should not be any significant deviations in the results obtained if the algorithm was applied over the new matches.

4.1 Identification of the Hoop and the Ball

Identification of the hoop and ball is done based on their color. The hoop is recognized as a red area located near the point that represents the intersection of the horizontal and vertical court boundaries. The accuracy of hoop identification on all frames representing the game under the basket is shown in Table 1.

Table 1
Accuracy of hoop identification (frames representing the game under the basket)

Number of frames	Correct	Incorrect	Percent
5605	5194	411	92.67

Regarding the hoop recognition, in 411 frames hoop was not properly recognized. This was mostly in frames when the camera shows players leaving the given side of the court after an unsuccessful shot, so the hoop is located at the very corner of the image. In such situations, the position of the basket is not of great importance, and thus relatively large number of incorrect identifications can be ignored.

During the game, we are especially interested in frames that represent a shot to the basket. These are frames from sending the ball to the basket, up to five frames after the ball leaves the area above the basket. The accuracy of hoop recognition is particularly important in these frames and is shown in Table 2.

Table 2
Accuracy of hoop recognition (frames representing a shot to the basket)

Number of frames	Correct	Incorrect	Percent
883	880	3	99.66

The ball is recognized as an area of color orange, and it can be found at almost any location in the frame. Exceptions are areas beyond the basket that surely represent the audience and where the ball cannot be during the game. When recognizing a ball, it may happen that it is obstructed by a player who holds it, and therefore it is not visible. The accuracy of the ball recognition in frames representing a shot to the basket is shown in Table 3.

Table 3
Accuracy of ball recognition

Number of frames	Correct	Incorrect	Not recognized	Visibility percent
883	809	7	67	91.62

From the table it can be seen that from 883 frames, the ball is not visible in 67 frames, and it is not accurately recognized at 7 more, i.e. the visibility percent is 91.62%. Regarding the accuracy of recognition, the algorithm correctly recognized the ball at 809 frames, while the ball is incorrectly recognized in 7 frames. Thus the percentage of accuracy is 99.14%.

4.2 Allocation of Players per Teams

When the player identification process is completed, the goal is to separate them into teams based on the color of their shirt. For this the saturation component of the color of the shirts applied in histograms was used. Such principle of separation of players is applicable in most basketball games and in most leagues. One of its biggest advantages is that the algorithm does not have to be re-trained for every new match and a new color of shirts. The algorithm was tested on the same test set that was also used in player recognition, consisting of 100 frames from ten different NBA basketball games. The results of the players' separation based on the color of the shirts are shown in Table 4. From the total of 748 recognized players, the algorithm correctly classified 691 players, which is an accuracy of 92.38%. The algorithm showed more accuracy for players wearing light shirts (96.34%), compared to players in dark shirts (88.21%). The reason for this difference lies in the fact that dark shirts often have a light area on the side.

Table 4
Separating players by teams

	Classified correctly	Classified incorrectly	Percent
Light shirts	369	14	96.34
Dark shirts	322	43	88.21
Total	691	57	92.38

4.3 Determining the Player's Position

Once the boundaries of the court and players are identified, it is possible to determine their court position. During the game there are many borderline cases. For example, if a player is between two positions, the decision of the algorithm to assign one of them to a player can be considered both correct and incorrect. Also, if a player is on the three-point line, the algorithm is not sufficiently accurate to

recognize whether the foot is "pinching" the line or not. In this paper, the position is considered correct if the algorithm has allocated the player to any of the two areas between which he is located.

Determining the player's position has been tested over a set of one hundred frames from ten different matches. The results of determining the positions of players are shown in Table 5. From 748 players, the correct position is determined for 724 ones, which is an accuracy of 96.79%. There are two reasons for 24 incorrectly determined positions. The first is in defining the boundaries of the restricted area. If the boundary is not correctly defined, the starting points for the transformation matrix are also incorrect. As a consequence, points in the frames representing player positions are not correctly mapped to real positions on the court. Another reason is the accuracy of the recognition of the player's ankles. In some cases, the joints are recognized above or below their actual positions. Then the mapping is done in accordance with the recognized positions of the ankles, which can lead to the allocation of the player to an incorrect area on the court.

Table 5
Accuracy of allocating player's position on court areas

Area	Correct	Incorrect	Percent
Area 1-3	60	0	100
Area 1-2	60	6	90.91
Area 2-3	42	0	100
Area 2-2	148	7	95.48
Area 3-3	58	0	100
Area 3-2	62	2	96.87
Area 4-3	13	0	100
Area 4-2	74	1	98.67
Area 5-2	102	6	94.44
Area 6-3	14	0	100
Area 6-2	91	2	97.85
Total	724	24	96.79

4.4 Determination of the Shot Position

In the process of testing the algorithm for determination of the shot position, one quarter of the basketball game was used, within which the algorithm recognized a total of 43 shots to the basket. By analyzing the first frames, the data presented in Table 6 were obtained. The table shows that the location is correctly recognized for 36 shots, which is 83.72% accuracy. Examples of accurate determinations are shown in Figure 4. The position of the player at the moment of the shot is shown by a green square, while the position of the ball is shown as a red square.

Table 6
Accuracy of determination of the shot position

Total number of shots	Correct	Incorrect	Percent
43	36	7	83.72

Figure 4
Player shooting position detections

During the analysis, the algorithm failed to detect the player's position in seven frames. This has mostly happened in situations where a player who does not shoot the ball to the basket is recognized, and therefore the correct shot position on the court could not be determined. Due to the resolution of the images, the algorithm is sometimes not able to recognize which of the two players, located near the ball, sends a shot. It also happened that the court boundaries are not correctly recognized so the algorithm does not map the recognized position to the correct location on the court.

4.5 Determining the Accuracy of the Shot

The accuracy of the shot is determined based on whether the ball has passed through the area under the basket after it has been found in the area above the basket. Testing was carried out over 43 recognized shots. Determining the accuracy is given in Table 7. The table shows that testing was done with a different number of frames that are being considered after the ball is found in the area under the basket.

Testing showed that the optimal number of frames is five, when an accuracy of 88.37% is obtained. A smaller number of frames have a greater number of

incorrect recognitions in situations where the points are scored; in some situations, the ball is still not visible in the area under the basket, so the algorithm describes the shot as unsuccessful. When more than five frames are used, the algorithm has a higher number of incorrect identifications in situations where no points are scored. This happens because the player who catches the ball after a failed shot, comes in the area under the basket, and the algorithm recognizes the ball in the required area and describes the shot as successful.

Table 6
Accuracy determining of the shot outcome depending on the number of frames

Three frames after the exit of ball from the basket area								
Score			No score			Total		
Correct	Incorrect	%	Correct	Incorrect	%	Correct	Incorrect	%
13	6	68.1	20	4	83.33	33	10	76.74
Four frames after the exit of ball from the basket area								
Score			No score			Total		
Correct	Incorrect	%	Correct	Incorrect	%	Correct	Incorrect	%
15	4	78.95	20	4	83.33	38	5	81.4
Five frames after the exit of ball from the basket area								
Score			No score			Total		
Correct	Incorrect	%	Correct	Incorrect	%	Correct	Incorrect	%
18	1	94.74	20	4	83.33	38	5	88.37
Six frames after the exit of ball from the basket area								
Score			No score			Total		
Correct	Incorrect	%	Correct	Incorrect	%	Correct	Incorrect	%
18	1	94.74	18	6	75	36	7	83.72

Conclusion

The automatic keeping of basketball statistics is the ultimate goal of this research. It would allow a completely objective recording of the event, as there is a danger that different statisticians will characterize the same event differently (for example, after a missed shot, if the ball first touches the floor, and then one of the players takes it, whether it is a rebound or turnover?). In addition, during a game, numerous activities are taking place that are omitted by classical statistics. Creating such a system, would potentially allow all of these events to be recorded.

The presented algorithms are the first step in creating automatic statistics management. Their basic contribution is robustness and applicability over large samples of footage, from basketball games. The research brings new approaches in the court determination process (using the Canny algorithm), the separation of players by teams (based on the saturation component calculated over areas covered by their shirts) and determining the exact position of players (using spatial transformations). A player-part-based player identification approach has

also been modified, and the search area has been reduced based on part size analysis performed over complete training and testing sets.

Automated determination of the player's position in the court is done using the footage broadcasted to viewers through television stations. This type of footage provides a look from just one camera, at any moment of observation, making the detection process more difficult. The first step involves decomposing the image to frames. An analysis is performed on frames in order to separate the game itself and discard the frames that display audience images, interviews with celebrities, announcements of events, etc. In this research, the goal is to collect data on shots as one of the basic statistical parameters.

In addition to the reduced set of frames (which represents approximately a half of the starting set), an algorithm for the separation of shots is applied. Shot is represented as a group of frames from the moment of sending the ball to the basket, up to five frames after the ball leaves the area above the basket. Therefore, it is necessary to implement the recognition of the hoop and ball. This recognition is based on their color within the HSV color model.

The first frame in each group of frames representing the shot is used to determine the position from which the shot is sent. In order to do that, it is necessary to recognize the player who shoots the ball to the basket and determine his position on the court.

For the purpose of detecting a player, a model based on an unoriented mix of parts was used. This approach provides a general framework for modeling the relations between occurrences among mixes of parts, as well as for spatial relations between part locations. In this way, players are detected in the court, as well as the positions of their body parts (arms, legs, head, and torso). Determining the position of arms and legs, in a very dynamic sport such as basketball, is of great importance.

During the detection process, a number of incorrect positive recognitions occurred. These were mostly people in the audience and referees. The paper presents solutions to remove these objects, so that only players are selected, because determining their position is precisely the purpose of this research. Using the Canny algorithm for determining the edges, the court boundaries were defined as well as the area that represents the restriction zone. By mapping points from these areas into real coordinates, a spatial transformation function is obtained that allows us to determine where the player is in relation to the court. The position of the player who sends the shot to the basket represents the location from which the shot was sent.

The accuracy of the shot is determined on the basis of a simple algorithm that checks whether the ball, after the area above the hoop, was also present in the area under the hoop. If it is, we understand that a shot is successful; otherwise we say that the shot was unsuccessful.

The space for further research is wide. Here we present only the first step in creating a system for automatic keeping of basketball statistics. Its advantage in comparison to other systems is that it uses images from just one camera. When analyzing the opposing team and preparing for new matches, coaches most often have exactly this kind of footage. In order to create a practically applicable solution, it is necessary to record other statistical parameters such as rebounds, assists, turnovers, blocks, etc.

It would be particularly useful to further develop the research itself in the direction of qualitative analysis of the game and recognizing the templates in the game of a particular team. Such research should include the movement of players, and their recognition in each frame would be done. In this way, a player base could be created based on their characteristics, which are determined automatically by the created algorithm. Such a system would be very useful to coaches and scouts, as they could quickly find the profile of the players that fit their team.

References

[1] Felzenszwalb, P., Girshick, R., McAllester, D. & Ramanan, D., 2010. Object detection with discriminatively trained part based models. IEEE Transactions on Pattern Analysis and Machine Intelligence, 32(9), pp. 1627-1645

[2] Yang, Y. & Ramanan, D., 2011. Articulated pose estimation with flexible mixture of parts. Computer Vision and Pattern Recognition

[3] Zhu, X. et al., 2005. Video data mining: Semantic indexing and event detection from the association perspective. IEEE Transactions on Knowledge and Data Engineering, 17(5), pp. 665-677

[4] Huang, C., Shih, H. & Chen, C., 2006. Shot and scoring events identification of basketball videos. s.l., s.n., pp. 1885-1888

[5] Alahi, A., Boursier, Y., Jacques, L. & Vandergheynst, P., 2009. Sport players detection and tracking with a mixed network of planar and omnidirectional cameras. s.l., s.n., pp. 1-8

[6] Chang, M., Tien, M. & Wu, J., 2009. WOW: wild open warning for broadcast basketball video based on player trajectory. s.l., s.n., pp. 821-824

[7] Lu, W., Ting, J., Murphy, K. & Little, J., 2011. Identifying Players in Broadcast Sports Videos using Conditional Random Fields. s.l., s.n., pp. 3249-3256

[8] Schumaker, R., Soliman, O. & Chen, H., 2010. Sports data mining. s.l.:Springer

[9] Lewis, M., 2003. Moneyball. New York: W.W.Norton & Company

[10] Piatetsky-Shapiro, G., 2011. [Online] Available at: http://www.kdnuggets.com/faq/difference-data-mining-statistics.htm

[11] Ballard, C., 2005. Measure of success. Sports Illustrated

[12] Devenport, T. & Prusak, L., 1998. Working knowledge. s.l.:Harvard Business School Press

[13] Lahti, R. & Bayerlein, M., 2000. Knowledge transfer and management consulting: a look at the firm. Business Horizonts, 43(1), pp. 65-74

[14] Ackoff, R., 1989. From data to wisdom. Journal of applied systems analysis, Volume 16, pp. 3-9

[15] Carlisle, J., 2006. Escaping the veil of Maya - widsdom and the organization. Koloa Kauai, HI, s.n

[16] Chen, H., 2006. Intelligence and security informatics for international security: information sharing and data mining. s.l.:Springer

[17] Viola, P. & Jones, M., 2004. Robust real-time face detection. International Journal of Computer Vision, 57(2), pp. 137-154

[18] Dalal, N. & Triggs, B., 2005. Histograms of oriented gradients for human detection. s.l., s.n., pp. 886-893

[19] Bar-Hillel, A. & Weinshall, D., 2008. Efficient learning of relational object class models. International Journal of Computer Vision, 77(1), pp. 175-198

[20] Holub, A. & Perona, P., 2005. A discriminative framework for modeling object classes. s.l., s.n

[21] Quattoni, A. et al., 2007. Hidden conditional random fields. IEEE Transactions on Pattern Analysis and Machine Intelligence, 29(10), pp. 1848-1852

[22] Ramanan, D. & Sminchisescu, C., 2006. Training deformable models for localization. Computer Vision and Pattern Recognition, Volume 1, pp. 206-213

[23] Okuma, K. et al., 2004. A boosted particle filter: Multitarget detectionand tracking. s.l., s.n

[24] Cai, Y., de Freitas, N. & Little, J., 2006. Robust visual tracking for multiple targets. s.l., s.n

[25] Lifang, W., Xiuli, H., Hao, C. & Wei, S., 2007. Shot segmentation and classification in basketball videos. s.l., s.n., pp. 539-542

[26] Huang, C., Shih, H. & Chen, C., 2006. Shot and scoring events identification of basketball videos. s.l., s.n., pp. 1885-1888

[27] Daniyal, F., Taj, M. & Cavallaro, A., 2010. Content and task-based view selection from multiple video streams. Multimedia Tools and Applications, Volume 46, pp. 235-258

[28] Perše, M. et al., 2009. A trajectory-based analysis of coordinated team activity in a basketball game. Computer Vision and Image Understanding, 113(5), pp. 612-621

[29] Theron, R. & Casares, L., 2010. Visual Analysis of Time-Motion in Basketball Games. s.l., s.n., pp. 196-207

[30] Wu, L. et al., 2012. Semantic aware sport image resizing jointly using seam carving and warping. Multimedia Tools and Applications

[31] Ivankovic. Z. Rackovic M. Ivkovic M. 2014. Automatic player position detection in basketball games, Multimedia Tools and Applications, 72(3), pp. 2741-2767

[32] Ivankovic, Z. et al., 2010. Appliance of Neural Networks in Basketball Scouting. Acta Polytechnica Hungarica, 7(4), pp. 167-180

Transport Habits and Preferences of Generations — Does it Matter, Regarding the State of The Art?

Anita Kolnhofer-Derecskei, Regina Zs. Reicher, Ágnes Szeghegyi

Óbuda University Keleti Faculty of Business and Management, Institute of Enterprise Management Tavaszmező u. 17, 1084 Budapest, Hungary
derecskei.anita@kgk.uni-obuda.hu; reicher.regina@kgk.uni-obuda.hu; szeghegyi.agnes@kgk.uni-obuda.hu

Abstract: Every single day we spend one hour, on average, with travelling and this value has not changed for decades. According to the Hungarian timescale statistics, approximately one hour per day, on average, has been spent on travelling for the last 30 years. The world, however, has changed a lot in 30 years and one the best examples for this is the quick sequence of generations. Currently, there are at least four generations at the same time. The current study briefly introduces each generation, then discusses the differences and preferences in the travelling habits of generations who are present in the labor market. The aim of this study is to give a structured preliminary research plan, based on the state of the art. Therefore, the problems of further empirical research is reasoned and a well-structured research plan can be specified. Later, this conceptual model helps us to study and understand travel habits and preferences of various generations.

Keywords: travel habits; generations; preference; literature review; research plan

1 Introduction

According to the national travel timescale statistics, the time spent on travelling has not decreased since 1986/87, but the composition of the traffic has considerably changed and the distribution of travel time within society has also been irregular. [11] According to the 1986/87 timescale of KSH (Hungarian Central Statistical Office) we travelled 61.8 minutes per day on average; the survey indicated 59.4 minutes in 1999/2000 and 65.2 minutes in 2009/2010 spent on travelling. [10]

When we are talking about travelling, it is not equal to urban transport or commuting, it consists of more, because holiday trips and journeys should be also taken into account. Urban transport covers mostly, the public transport system in a

city and commuting can be defined as "travelling from home to the workplace" [8]. For example, a business trip happens out of the city, but the aim differs. In this paper everyday travelling choices were examined regarding the secondary data, therefore, all expressions were used as similar. Several things can affect the time spent on travelling: demographic background of users; typical features of the given settlement or the social status. [22] Fleischer and Tir [10] analyzed the six explanatory variables together (age, gender, activity, qualification, county, and settlement-category) and they found that it could be explained only through 10% of the heterogeneity of the time-use values. It was the age and the activity status that influenced the time-use pattern best. Therefore, our travels can be influenced by several factors; the present study will focus on the differences by age. The age is examined from two aspects: (1) generation characteristics and (2) age characteristics. Fleischer and Tir [11] analyzed the Hungarian data and reached the conclusion that during the sixty-year period from the age of 20 to the age of 79 the time spent on travel is decreasing by one minute by approximately one age-year, from about 90 minutes to 30 minutes. (Fig. 1)

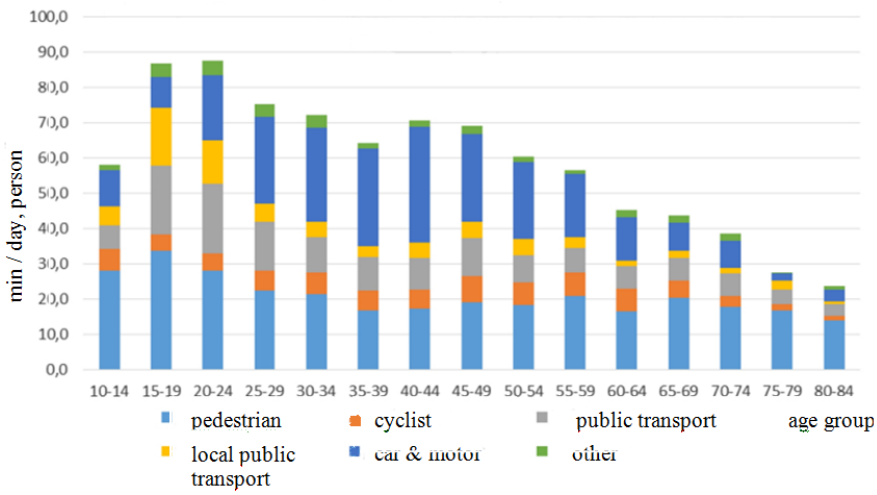


Figure 1

Time spent with travel (minute/day, head) in the population aged 18-84 according to age groups and modes of travel, 2009/2010 [10]

Examining the chosen means of travel (how the referred authors called them), the most striking difference can be observed in case of car and motorbike travel. While men aged 40-44 would spend 43.5 minutes per day on average in a car (it is 21.6 minutes in case of women); it is reduced to 18.2 minutes for 60-64 year old men (and 7.6 minutes for women of the same age group). This can also be explained by the commuting to work, because the working-age starts after school-leaving age and lasts until retirement. Although mobility and willingness to travel has increased, the objective of everyday travel is still commuting to work [13].

Travelling to work is the greatest item for the employees on weekdays. Therefore, the time-use within the work week is different on work-days and weekdays. Although daily commuting and long travels are decreasing the free time in case of each generation, but there are some activities, where this condition does not apply. Moreover, among those aged 17-29, the daily commuters spend even more time with social leisure activities than their non-commuter fellows, who work where they live [23]. The daily commuters make up for the time lost during weekdays on the weekends and they spend more time with looking after their children over the weekend than those working near their residences. The travel destinations in the weekends include shopping centers, visits to family and friend, as well as doing sports or engaging in cultural activities. [23] Regarding travel destinations, there is a huge difference between travelling on weekdays and during a weekend.

The travel habits, of course, can be characterized not only by the time spent on it but also by the chosen means of transport and this latter is justified by the travel distance. Regarding public transport and considering the number of passengers carried in the domestic long-distance passenger transport, the share of bus traffic was dominant (77% in 2017) and the share of rail passenger traffic was 23%. In international passenger transport, 49% of passengers travelled by air, 32% by rail and 20% by bus. The public transport within a settlement meant bus transport (55% is the share of passengers carried); streetcar (22.7%), tube and underground (14.5%), trolleybus (4.8%) and HÉV Suburban Railway Service (3%). 40% of travels considered without returning home was commuting. Regarding the means and methods of transport, 27% of passengers named public transport on a national level; 38% travelled by car and 17% travelled by bicycle. [15]

In terms of individual travel, the sharing of cars is still a strongly determinant factor and this tendency is even growing. Parallel with this, the number of automobiles registered for the first time in Hungary is increasing year by year. [21] (Fig. 2) It does not mean that the car use is also exponentially increasing at this pace; it is much more that travelling by car has become part of a mixed, modal travelling lifestyle; in other words, the passenger is changing the means of transport flexibly. Together with this, parking has also become an issue in all of the dense population areas. The European Union drafted an action plan to solve parking issues several years ago and researchers have been working on multiple, real-time information systems, in order to find a solution for this expanding and urgent problem. [28] Owning a car, however, still has a strong impact.

According to a representative survey carried out in 2016 in Hungary, 44% of households had at least one car. Out of them 39% owned one car, while 5% of the households had two or more vehicles [9]. On the basis of statistics, the individual transport – with all its advantages, disadvantages and risks - has become the main method of mobility besides public transport. The authors explain this with suburbanization processes, spread of corporate cars and the cost efficiency, because some would say „if we have a car, only the fuel means extra cost when we travel". [13, p. 187]

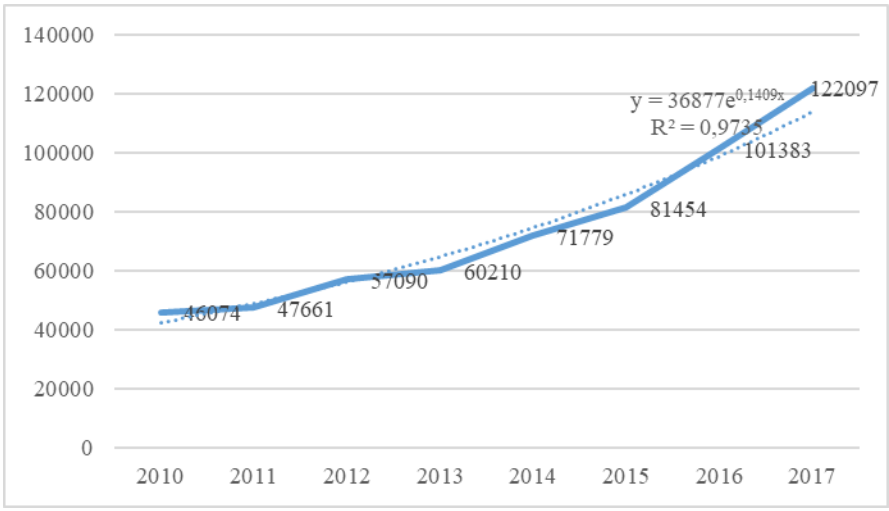


Figure 2
Number of road vehicles per year registered for the first time in Hungary (pcs.) (Source: [21])

And the fuel prices have increased proportionally to a lesser degree than the public transport tickets, the price of which grew 64-fold between 1990 and 2013, although this feeling is changing over time. The data of the Hungarian Central Statistical Office from 2013 contradict to this statement: the Hungarian population spent 732 billion HUF on transport in 2012; 71% of this amount (519 billion HUF) was spent on car use including motorway and parking fees but excluding taxes, insurance costs and costs of car purchase. 212 billion HUF was spent on public transport. Comparing these values with the distance taken by the individual means of transport, it can be concluded that the cost of car use (25 HUF/passenger kilometer) is significantly higher than the average cost of public transport (15 HUF/passenger kilometer). But car owners and car users must be separated also. Regarding the different ages of the drivers and passengers, the younger and elder generations belong to the second part. In case of car use there is another wider issue to be considered, namely the state and length of local and national road network. This was almost 207 thousand kilometers in 2016, and it had not changed significantly compared to 2010. Since 2010, the length of bike paths and motorways has grown the most (by 56% and 44% respectively), while the length of roads belonging to the category of other types of roads and the length of footpaths has decreased (by 1.4% and 1.0% respectively). The length of motorways in 2016 (1168 km) was more than in 2010 by 9.5%. [16]

The expenditures of households spent on transport are also affected by the type of the settlement. While on average 20% of Hungarian households spend only on means of public transport, this ratio is 29% among the households in Budapest. Only 17% of the citizens in the capital city spent money on cars as opposed to the

average Hungarian ratio, which was 28%. This, of course, did not mean that less people had cars; it rather referred to the mixed transport (while 25% of the Hungarian households spent on public transport as well as car use, this ratio is 30% in Budapest). [15]

HUF 8984 per head and per month was spent on travelling and transport in 2018 on average; the greatest proportion of which was fuel purchase, with HUF 6209 per month. This type of expenditure increased by 6.4% on current prices compared to the previous year. [21] The cost of travel and transport is the third largest item in the monthly expenditures of households; it was 10.8% of total consumption expenditure. Compared to year 2000 (volume index, year 2000 = 100%) the volume change of consumption expenditures of households was 180.8% in terms of travel costs; regarding the structure of consumption expenditure per head it was 11.5% in the households. [18] Regarding this value, however, there has been a gradual decline because the ratio of money spent on travelling has been gradually decreasing in the household consumptions. In the first half of 2017 the average spending per head was HUF 8928 per month. [19]

Regarding the bicycle usage that is for local public transport, it is used first of all by those living in smaller towns or villages, where there is access to local transport. An average Budapest resident spends 27 minutes on this mode daily, 4-5 times more than the capacity of a village or small town resident to avail of this mode. Contrarily, in case of long distance public transport, village residents spend almost three times more time on this mode than the residents of Budapest or of the county centers do. [10] Although there are more and more cyclists in the capital city as well and the bicycle sharing system (MolBubi) is also growing, the present study does not further discuss this means of transport.

In this chapter we try to organize all those factors which influence travel and commute habits. Fleischer and Tir [10] provided a multidimensional model which characterized two secondary important factors. The first was the effects of different demographic, social backgrounds on the transport time-use variability. The second was the specialties in the time-use character of the different transport modes. Regarding the first one they found that the social background (i.e. six explanatory variables together like age, gender, activity, qualification, county, and settlement-category) explained only 10% of the heterogeneity of the time-use values. The age and the activity status influenced the time-use pattern best. As for the age groups: between age 20 and 80 years the average daily transport time use decreased from 90 minutes to 30 minutes. Looking at the activity, 60% of the time-use of the population is produced by the 51.7% employed people.

2 Generations

Despite of the fact that age and activity influence transport and commute preferences, the passengers are not divided into separate age groups in the available Hungarian statistics, even though the international references extensively discuss the comparison of travel habits and preferences by generations. The definition of generation according to the glossary of definitions by the Hungarian Central Statistical Office (KSH) says that generation is a specific type of population cohort: it means a group of people who were born in the same year. As the members of a generation should live through demographically important events (for example getting a degree, marriage, birth, employment, death etc.) interlocked this way time and frequency of occurrence of these events are comparable with factors affecting in time. In other words, the impact of social and historical background, in which the given generation is growing up, is significant, but all the generations go through the same stages of life and more or less they have to face the same challenges during this journey.

As it was discussed above, the working age population travels the most. Currently, there are three generations on the labor market at the same time. The generational differences are markedly visible in the field of HR. [25]

Our previous work discussed the HR relevant differences among generations in detail [14] and their travelling habits were also briefly reviewed. [26] Table 1 below provides a summary of the latest composition of the national labor market on the basis of the Hungarian Central Statistical Office (KSH) database, as well as references and own research outcomes.

In case of individual generations, that historical, social and cultural background should be highlighted, in which the members of the generation of the given age (cohorts) were raised. In other words, the changes of the world generate the generations. The borderlines between generations are not nearly as sharp; it is also proven by the existence of the so-called intermediate (cuspars) generation. In addition to this, the individual birth years may differ by regions and countries; moreover, each generation can be further divided by life stages. Let's just consider that the younger members of Y generation still live with their parents and study, while the older members work and some of them even founded a family or at least leads an independent household apart from their parents. Hereinafter, several examples prove the above by analyzing international professional literature sources. The biggest emphasis in research is given to the Y generation or Millennials. "Millennials are also living through times of economic dislocation and technological change. History shows that the combination of technological change, such as the advent of smartphone technology, television, or radio; combined with macro forces that shape behaviors, such as the Great Recession, the Great Depression, or World War II can lead to societal change that can last generations.

Table 1
Description of individual generations

Generation	Year of birth	Short description [3]	Historical background they were raised in	Number of population (thousands of persons)	Number of economically active people (thousands of persons)	Number of the employed (thousands of persons)
Y generation	1982-1996 (20-35 years old)*	delay in marriage, childbearing and other life events high adoption of technology credited to have higher preference for urban areas delay in driver's licensing compared to previous generation	global economic crisis; strong globalization; free movement mobile and smart devices and applications; Internet is part of everyday life; greater mobility (mobility-encouraging services, e.g. Erasmus program or products e.g. low-cost flights); Influencers and bucket list.	1769.6	1296.6	1224.2
generation X	1967-1982 (35-50 years old*	active workers often live with children telecommute more often increased adoption of e-commerce	change of regime; Western impact, globalization (less obstacles in front of travels, transition to EU); new technologies, personal computers, mobile devices, world wide web, travel magazines, exhibitions, programs.	2254.6	2013.3	1946.9
Baby Boom	Before 1966 (50-65 years old)	transitioning into retirement higher income generation (* not in Hungary) increased amount of discretionary funds for leisure trips less need for space in residential location	socialist era, closed economy; no chance or only limited and strictly restricted chance for travel, e.g. strong customs, currency limit) landline phones, radio and television; written communication (through mail); stores for purchases in foreign currency, exclusive foreign goods, dissident acquaintances.	1910	1220.9	1176.4

*Age is adapted to the data of KSH classification (year 2017); year of birth is according to the references

It is in this context that Millennials, with their relative propensity for urban lifestyle components (whether they live in cities or in suburbs), dexterity with technology, while starting careers during economically constrained times can leave a lasting impact on society. In fact, they are already driving trends." [1]

It is very interesting that in the field of age grouping, the KSH (Hungarian Central Statistical Office) only classifies by generations in one study (different years of birth were indicated in the classification). This study analyzed the internet use habits. It means that the description of individual generations can be closely related to the technological changes. Currently the internet use is part of a modern lifestyle in Hungary; according to the KSH [20] 69.9% of respondents connects to the Internet every day, while 20%, several times per week. When they are asked what they generally think about Internet, 56.9% of the population declared that they regard it indispensable and only 20% say that they could live happily without it. Primarily, the members of generation Z are web connected and feel it is indispensable (83.7%) while older people, the members of the baby boom generation, could tolerate the absence of internet with less trouble. Therefore, the generations we examined can be distinguished on the basis of their relation to IT and the World Wide Web. This paper shows that internet use habits are related to generational differences, because this factor differs slightly among each generation, and influences their behavior, even though, this factor must be taken in consideration.

The "online" generation would like to be connected and in real-time, anywhere and anytime. For this, they need appropriate IT support and the use of both their hands; therefore, they cannot split their attention to driving a car. This need is also obvious in the purchase of transport services. In addition to comfort and cost efficiency, the environmental consciousness further strongly justifies the preference of new (mixed and public) means of transport. Csigéné et al. have confirmed this in their study on a Hungarian sample. Their conclusions show the interests of Millennials on sustainable consumption and eco-labeling. [7]

International Results

Although there have not been any Hungarian studies in transport and commuting research dealing exclusively with generation differences; International Professional sources in English, refer to this topic several times. Unfortunately, the targeted topic (i.e. generational differences regarding transport habits) was not studied, typically the age differences were underlined, but mostly these sources were from the characteristics of the labor market (e.g. commuting as a part of transport). This study focuses on the generational differences. The latest most relevant study was examined by Circella et al. – prominent representatives of the topic – have been doing one of the most extensive researches [5]. They studied Millennials' choices, through the analysis of a comprehensive secondary dataset

and approximately 2400 residents of California, including both Millennials and members of Generation X.

They have concluded that Millennials are increasingly reported to behave, and travel, differently from previous generations at the same stage in life. Among the observed changes, they postpone the time they obtain a driver's license, often live in urban locations and do not own a car, drive less if they own one, and use alternative travel modes more often. Millennials' current choices are expected to be a sum of lifecycle, period and generational effects: their current behaviors are not necessarily going to continue as Millennials grow older and transition to more stable life stages. Millennials tend to live in areas that have the lowest levels of accessibility by non-car modes. This sharply contrasts the residential location of independent Millennials who are more often found to live in locations with higher accessibility. Central locations are more conducive to the adoption of greener and non-auto commute modes (and/or may reinforce the propensity of young adults to use such modes or to adopt multimodal travel). In addition to this [2] the Generation Y cohort generally have lower rates of driver licensing, vehicle registration and car ownership, in addition to their increased rates of public transport usage. These trends have been observed in many countries around the world, including the USA, Australia, Canada, Japan, the UK and many other European counties.

Urban mobility literature has also been reviewed and processed by Cost et al. [6] According to their conclusions, generation Y twice as willing to ride a bicycle than the older generation; three times as willing to choose shared transport (e.g. Uber) and five times is happier to use public transport to commute to school or to work. Gen X'ers rely heavily on the use of cars for their commute. Many older Millennials who live in urban areas actually report that they do plan to purchase a new vehicle in the near future, but they are less likely to be mono-drivers and more likely to be multimodal commuters, even if they live in neighborhoods that are less supportive of such behaviors. Millennials often report reducing their use of transit 15 or the amount of walking or biking as the result of the use of Uber/Lyft or other shared mobility solutions that mean of travel is not straightforward.

Lavieri et al. [24] examined the driver's license holding, vehicle ownership, and commute mode choice of the millennial generation. According to their results, parenthood is associated with an increase in driver's license holding and personal vehicle ownership; Parents are likely to express a greater pro-car attitude than non-parents, a finding that is consistent with expectations. Parents need the flexibility afforded by a personal automobile to transport their children, in addition to fulfilling their own travel needs and hence, vehicle ownership is higher in households with children. Through these mixed-use developments, where Millennials can work, play, and shop within short distances could help foster the continued use of non-motorized modes of transportation. Moreover, Millennials who are more technology-dependent exhibit lower levels of vehicle ownership and

usage, and higher levels of non-motorized mode and transit use for commuting. [24]. Now, associates of these young people are using transit more than in past years. [12] Millennials are multimodal; they choose the best transportation mode (driving, transit, bike or walking) based on the trip they are planning to take. [1] Public transportation options are considered the best for digital socializing and among the most likely to connect the user with their communities. Transit also allows Millennials to work as they travel. They justify their choices by clear competitive advantage. Reasons and motivations for transportation choices are pragmatic, with 46% stating that a need to save money drives their choices; 46% note convenience, 44% want exercise and 35% say they live in a community where it just makes more sense to use other transit.

Shearmur [29] divides the members of generation Y according to their travelling habits. The individual groups travel depending on their labor market situation and occupation because work can be performed across the city, on the move or on the fly. The three groups are as follows: (1) hyper-mobile knowledge-related jobs independent artist, or a successful businessperson who prefer a wide variety of urban locations (2) semi-mobile: dog walkers, hairstylists who come to clients' homes, house cleaners – these occupations have no fixed place-of-work, and do not have the same locational freedom (3) and ordinary fixed place-of-work. [29]

We have found an interesting research in the studies of Bösehans and Walker, who specifically concentrated on bus transport and identified different types of bus users. In this sample, 88% of bus users lived within a four mile radius of the university campus, supporting the notion that bus trips can be sufficiently short to be undertaken by either walking or cycling. [2]

Further research has been done in Great Montreal [12] and New Zealand with focus group depth interviews and online surveys. [27] The conclusions are very similar. It seems that the changing generations and the new habits and attitudes have an impact on transportation as well.

3 Research Plan

3.1 Research Questions

Based on the aforementioned facts, we would like to describe our research plan. We are interested in the demographical impacts (highlighted the age) on transport and commute habits. First, the methodological facts are listed: (1) mostly preferred and used method is diary technique, with it, a longer time period can be studied, providing a possibility to compare weekday and weekend activities. In addition this method results in qualitative and quantitative data. The Hungarian

Central Statistical office also uses diary study techniques. (2) Our population will emerge from the capital city's passengers, who mostly commute to their schools or workplaces. It should be noted that our research is going to involve only people living in Budapest and its suburbs. There are two reasons for this: (1) the previous chapter also highlighted the impact of urbanization and (2) all the means of transport used in Hungary can be found here [17].

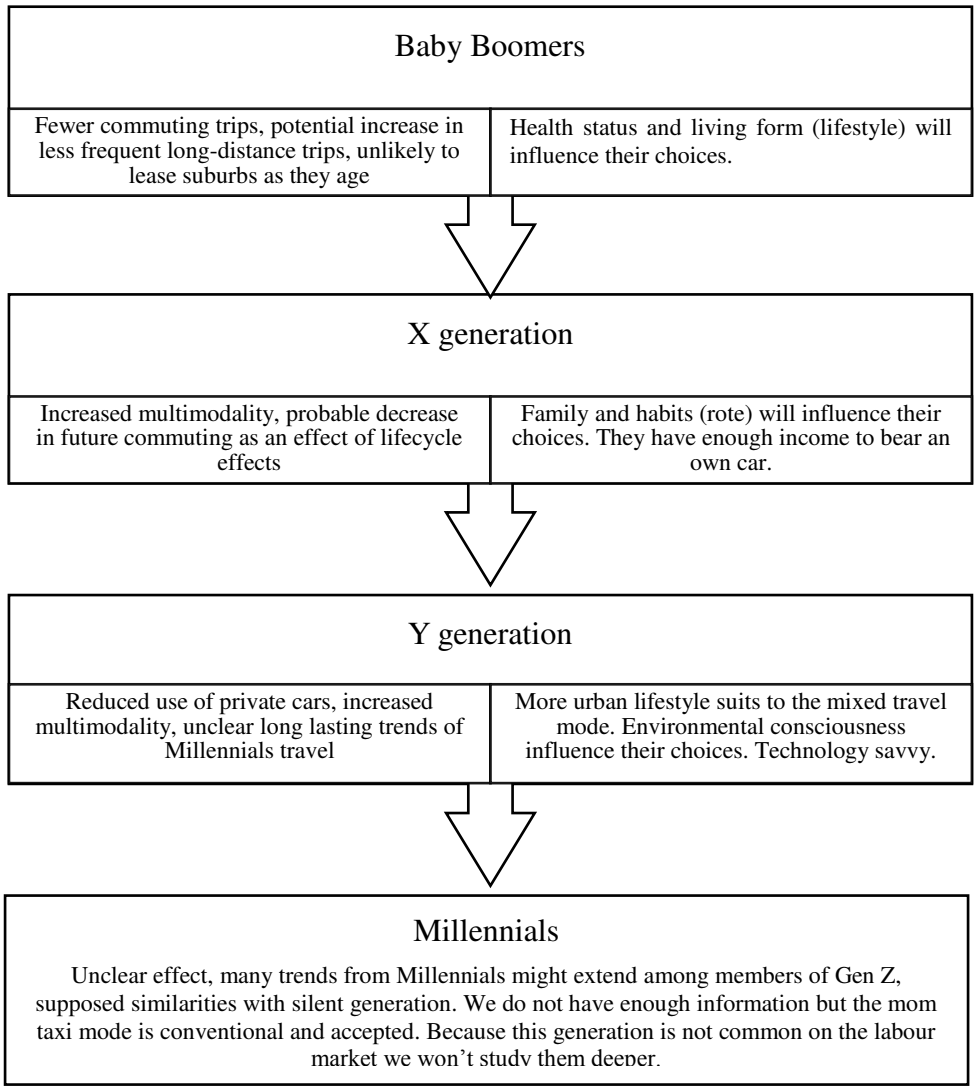


Figure 3
Impact on travel demand

The age-related and generational impact definitely showed up in the transport habits and preferences that we examined. Fig. 3 compares this tendency, which has also been discussed in international references [3] and our own preliminary hypothesis.

In sum, we would like to study the following topics and areas: transport habits and preferences among different Hungarian generations' members. We focus on the following three generations (1) Baby Boomers (2) generation X (3) Millennials or Gen Y. Finally the Fig. 4 shows a possible conceptual model about the measure-wanted factors.

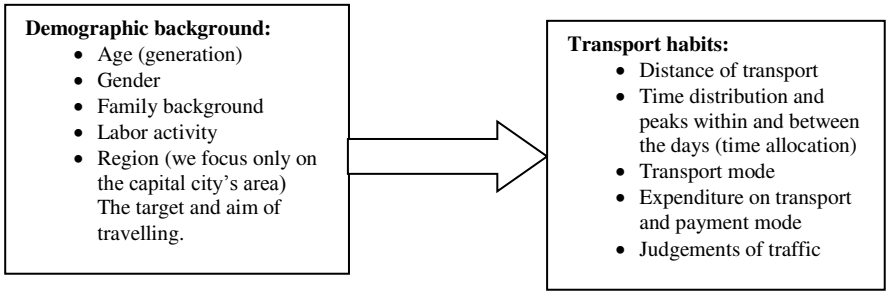


Figure 4
Conceptual model of further research

Although the demographic background is an important influencer, we focus only on age factor, using generation scales, because different generations are used to and socialize in different life environments. At the same time each generation goes through the same life periods. We are interested in the impact of the generational differences. Focusing on this, the following factors could be foreseen as influencers: (1) time spend and distance traveling (2) peaks in time distributions like (a) travels during certain times of days in a week and (b) travel times in one particular day (3) feeling and judgment (experiences) of the traffic (4) expenditures (a) amount of expenditure and (b) mode of payment (5) last, but not least, the preferred mode of transport.

Our preliminary hypothesis can be the following:

> H1. Age (grouped into generations) will influence users commute and transport habits.

Here we anticipate the impact from factor (1) to factor (4), because these factors are determined by the aim and target of travelling.

> H2. Life period and labor activity have a stronger impact on preferred transport mode.

Here we anticipate the impact of factor (5) which has an indirect impact on H1.

This last factor is the most interesting, we are interested in quantitative and qualitative data, as well, because reasons and ways of thinking given by the various generations' members, are meaningful and useful.

Conclusions

Understanding the differences among generations would take us closer to the planning of a future means of transport. The working-age population travels the most, but this population cannot be regarded as homogenous. The KSH analysis made on a Hungarian sample highlighted this fact [23], saying that "the proportion of commuters is significantly higher among employed men, youth, people with lower school qualifications and those living in villages, than those among women, older people, people with higher school qualifications and those living in towns." The characteristics of daily travel, however, are not uniform; there are considerable differences by the size of settlements, social, economic and demographic features, seasons, days of the week, time of the day, aim of travel or the chosen means of transport. For example, people living in Budapest use public transport as opposed to a car, in significantly higher proportion than the national average. [15]

As Circella et al. saw the new generation [5], Millennials are in a "changing" stage of their lifecycle, in which they are building the basis for their future life, family and work career. They will contribute to create new households and influence future travel patterns in many ways. Millennials are more technologically oriented than their older peers. They use social network platforms, internet and smartphone apps more often to perform certain activities and engage more often in travel multitasking. They show a stronger commitment to protecting the environment and are less opposed to increases in gas taxes to provide funding for public transportation [4].

Although the impact of generation characteristics is not negligible, the impact of age specific features is much stronger. Moreover, the two go hand in hand. It is clear that the typical age characteristics, in the case of generations, are shifted. The daily time scale also indicates the shifted life stages; as discussed by Lakatos [23]: "women give birth to their children at an increasingly older age, thus the 17-29-year-old people spend only 17 minutes per day on average on looking after their children, while the 30-49-year-old people spend twice as much, 33 minutes.

After all this, the question is: what affects the mobility of generation Y? The answer surveyed on an American sample [3], can be summarized as follows, in Table 2.

In summary, it can be concluded that the generational differences can be observed, but the age characteristics or the geographical demographic features have a much stronger impact on whether this trend is temporary, sustained or growing. Lavieri et al. has reached the same conclusion: Millennials seem to become more auto-oriented as they age and gain economic resources. [24]

Table 2
Factors affecting the travel preferences of generation Y [3]

Economic	**Auto cost**	**Technology**	**Demographic Changes**
Recession Unemployment	Gasoline Auto insurance Auto repairs Driver's education Other fees	Communication technology Transportation technology (Uber)	Delayed marriage Fewer children Boomerang effect
Residential Location	**Cultural**	**Regulatory Changes**	**Alternative Modes**
More likely to move to and live in cities	Environmentalists Less materialistic	Graduated driver's licensing Texting while driving laws	Better transit Improved infrastructure for walking, biking Sharing commute

The aim of this study is to draft further empirical, primary research, which enables the testing of our preliminary hypothesis but; the limitation of our study should also be mentioned. In case of the above classified high number of secondary data, the different generation limits, the limits given for the members of individual generations in the year of their birth may cause problems. Although it is understandable, because the classification limits (due to the definition) refer to the given social cultural impacts.

Acknowledgement

 Supported By the ÚNKP-17-4/I. New National Excellence Program of the Ministry of Human Capacities. The research was carried out in the frames of BrainBayCentrum.

References

[1] Association, A. P. T., 2015. *Millennials and Mobility.* [https://www.apta.com/resources/reportsandpublications/Documents/APTA-Millennials-and-Mobility.pdf, downloaded 10/06/2018]

[2] Bösehans, G. & Walker, I., N. D. 2016. 'Daily Drags' and 'Wannabe Walkers' – Identifying dissatisfied public transport users who might travel more actively and sustainably. *Journal of Transport & Health, Volume 3, Issue 3, pp. 395-403*

[3] Circella, G., 2015. *FACTORS AFFECTING PASSENGER TRAVEL DEMAND IN THE UNITED STATES,* California: Planning Horizons Seminar. White Paper Draft. [http://www.dot.ca.gov/hq/tpp/offices/owd/horizons_files/NCST_WP_Travel_Demand_Draft.pdf, downloaded 10/06/2018]

[4] Circella, G., 2017. *What Affects Millennials' Mobility? PART II: The Impact of Residential Location, Individual Preferences and Lifestyles on*

Young Adults' Travel Behavior in California, California: National Center for Sustainable Transportation (NCST). Research Report. [https://steps.ucdavis.edu/wp-content/uploads/2017/10/CIERCELLA-FULTOIN-PART-2-2017-UCD-ITS-RR-17-05-2.pdf, downloaded 10/06/2018]

[5] Circella, G. & Berliner, R. & Lee, Y. & Handy, S., L. & Alemi, F. & Tiedeman, K. & Fulton, L. & Mokhtarian, P., L. 2017. *The Multimodal Behavior of Millennials: Exploring Differences in Travel Choices between Young Adults and Gen Xers in California*

[6] Costa, P. B., Morais Neto, G. C. & Bertoldec, A. I., 2017. URBAN MOBILITY INDEXES: A BRIEF REVIEW OF THE LITERATURE. *Transportation Research Procedia,* pp. 3645-3655

[7] Csigéné, N. & N., Görög, G., & Harazin, P. & Baranyi, P. R., 2015. „FUTURE GENERATIONS"AND SUSTAINABLE CONSUMPTION. *Economics and Sociology,* Vol. 8, No. 4, pp. 207-224

[8] Davis, D. & Dutzik, T., 2012. *Transportation and the New Generation. Why Young People Are Driving Less and What It Means for Transportation Policy,* USA: Frontier Group

[9] Ficzere, F., 2016. Bosch a magyar autóhasználati szokásokról. / Bosch about the Hungarian Car Usage Habits. http://www.aoeportal.hu/index.php/infokozpont/hirek/365-bosch-kutatas-a-gepkocsik-eletkora-atlagosan-11-ev-magyarorszagon [online, downloaded on 06/2018]

[10] Fleischer, T. & Tir, M., 2016. The transport in our time-budget. *Regional Statistics,* pp. 54-94

[11] Fleischer, T. & Tir, M., 2018. *Hazai közlekedési időmérleg elemzés. / Domestic transport Time Use Research.* Siófok, Budapesti és Pest Megyei Mérnöki Kamara, pp. 81-86

[12] Grimsrud, M. & El-Geneidy, A., 2014. Transit to eternal youth: lifecycle and generational trends in Greater Montreal public transport mode share. *Transportation,* 41. pp. 1-19

[13] Kiss, P. J. & Szalkai, G., 2018 58(2). Az ingázás mobilitási jellemzői a legutóbbi népszámlálások adatai alapján. / Commute characteristics based on the census. *Területi Statisztika,* pp. 177-199

[14] Kolnhofer Derecskei, A., Reicher, R. & Szeghegyi, Á., 2017. The X and Y Generations' Characteristics Comparison. *Acta Polytechnica Hungarica,* pp. Vol. 14, No. 8, 107-125

[15] KSH, 2013. June 27. A lakossági közösségi és egyéni közlekedési jellemzői. / Public and individual transports' description. 2012. *Statisztikai Tükör,* VII. 47

[16] KSH, 2016a. A települések infrastrukturális ellátottsága. / Settlement's infrastructure. 2016. *Statisztikai Tükör*

[17] KSH, 2016b. *Az ingázás kiemelt célpontjai. / Accentuated Target of transport.* Budapest: Központi Statisztikai Hivatal

[18] KSH, 2017a. *A háztartások fogyasztása. / Domestic household consumptions. 2017. I. (preliminary data).* KSH

[19] KSH, 2017b. *Kiadási radar adatok. / Expenditures radar data.* Budapest: Központi Statisztikai Hivatal

[20] KSH, 2018a. A háztartások fogyasztása, 2017 (előzetes adatok). / Domestic household consumptions. 2017 (preliminary data) *Statisztikai Tükör*

[21] KSH, 2018b. Szállítási teljesítmények, közúti közlekedési balesetek. / Transport records, Transport Accidents on Public roads. 2018. I. quarter. *Statisztikai Tükör*

[22] Kukely, G. & Aba, A. & Fleischer, T., 2017. New framework for monitoring urban mobility in European cities. *Transportation Research Procedia ,* pp. 155-162

[23] Lakatos, M., 2013. *A foglalkoztatottak időfelhasználása az ingázás és a munkába járás idejének tükrében. /Employers' time use depending on commutes to the workplaces* . Budapest: KSH Műhelytanulmányok 13

[24] Lavieri, P. S. & Garikapati, V. M. & Bhat, C. R. & Pendyala, R. M., 2017. *An Investigation of Heterogenity in Vehicle Ownership and Usage for the Millennial Genartion.* 96th Annual Meeting of the Transportation Research Board

[25] Reeves, T. C. & Oh, E., 2008. Generational Differences. In: *Handbook of Research on Educational Communications.* Springer, pp. 295-303

[26] Reicher, R. & Kolnhofer Derecskei, A., 2018. *Generációs különbségek a hétköznapi közlekedésben. /Generational differences in everyday transport.* Siófok, Budapesti és Pest megyei Mérnöki Kamara, 7380

[27] Rive, G; &Thomas, J; & Jones, C; & Frith, B; & Chang, J., 2015. *Public transport and the next generation June 2015,* NZ Transport Agency research report

[28] Sándor Zs. P. & Csiszár Cs., 2013. Development Stages of Intelligent Parking Information Systems for Trucks *Acta Polytechnica Hungarica,* pp: Vol. 10, No. 4, 2013

[29] Shearmur, R., 2016. *The Millennial urban space-economy: dissolving workplaces and the de-localization of economic value creation,* McGill School of Urban Planning: Working Paper

Mobile Banking Authentication Based on Cryptographically Secured Iris Biometrics

Nemanja Maček[1], Saša Adamović[2], Milan Milosavljević[3], Miloš Jovanović[4], Milan Gnjatović[5], Branimir Trenkić[6]

[1,6] School of Electrical and Computer Engineering of Applied Studies, 283 Vojvode Stepe st., Beograd, Serbia, e-mails: {nmacek, btrenkic}@viser.edu.rs

[2,3,4] Singidunum University, 32 Danijelova st., Beograd, Serbia, e-mails: {sadamovic, mmilosavljevic}@singidunum.ac.rs, milos.jovanovic.10@singimail.rs

[5] Faculty of Technical Sciences, University of Novi Sad, 6 Trg Dositeja Obradovića st., Novi Sad, Serbia, e-mail: milangnjatovic@uns.ac.rs

Abstract: This paper[1] presents an approach to designing secure modular authentication framework based on iris biometrics and its' implementation into mobile banking scenario. The system consists of multiple clients and an authentication server. Client, a smartphone with accompanying application, is used to capture biometrics, manage auxiliary data and create and store encrypted cancelable templates. Bank's authentication server manages encryption keys and provides the template verification service. Proposed system keeps biometric templates encrypted or at least cancelable during all stages of storage, transmission and verification. As templates are stored on clients in encrypted form and decryption keys reside on bank's authentication server, original plaintext templates are unavailable to an adversary if the phone gets lost or stolen. The system employs public key cryptography and pseudorandom number generator on small-sized templates, thus not suffering from severe computational costs like systems that employ homomorphic encryption. System is also general, as it does do not depend on specific cryptographic algorithms. Having in mind that modern smartphones have iris scanners or at least high-quality front cameras, and that no severe computational drawbacks exist, one may conclude that the proposed authentication framework is highly applicable in mobile banking authentication.

Keywords: mobile banking; authentication; biometrics; iris; cryptography

[1] This paper is an altered version of [19] presented at the 9th International Conference on Business Information Security (BISEC), in Belgrade, Serbia, 2017. The modifications include the following: application of the proposed framework in mobile banking, detailed description of iris feature extraction, experimental evaluation with highly-realistic dataset and more detailed security evaluation of the system.

1 Introduction

Mobile banking is a service provided to customers by a financial institution that allows financial transactions to be conducted using a mobile device (a smartphone or a tablet) and accompanying software, usually provided by the same institution. Having said that , one may conclude that mobile banking is one of the most security-sensitive tasks performed by a typical smartphone user [1]. Although many financial institutions offer their mobile banking services "with peace of mind" [2], there is not a bulletproof solution providing users with 100% security guarantee. There are several security aspects regarding financial transactions conducted via mobile devices that should be addressed: physical security of the device, security of the application running, authentication of the user and the device to the service provider (bank's authentication server), encryption of data being transmitted and data that will be stored on device for later analysis by the customer. This paper addresses the authentication of the user and the device to the service provider. Variety of authentication methods are implemented in mobile banking today, all having their upsides and downsides. As an example, passwords are the easiest method to implement, but customers that employ passwords to mobile banking authentication are at risk of fraud or theft. Major companies have identified the need for strong security countermeasures and they are producing new hand-held devices with built-in biometric scanners. According to Gartner, over 30% of mobile devices are currently using biometrics, and banks should see that as an opportunity to secure their customers and transactions rather than a barrier to adoption [3].

Biometric authentication is the process of establishing user identity based on physiological or behavioral qualities of the person [4, 5]. Biometrics may be addressed as an ultimate authentication solution: users do not need to remember passwords or carry tokens and biometric traits are distinctive and non-revocable in nature [6], thus offering non-repudiation [7]. Like any personal information, biometric templates can be intercepted, stolen, replayed or altered if unsecured biometric device is connected to a network or if a skilled adversary gains physical access to a device which does not employ anti-forensic techniques that would prevent extraction of sensitive data (i.e. unprotected templates). Brief surveys of attacks on biometric authentication systems are given in [8, 9]. Due to non-revocability of biometric data aforementioned attacks may lead to identity theft. Having said that, it becomes clear that biometric systems operate with sensitive personal information and that template security and privacy are important issues one should address while designing authentication systems. To counterfeit identity theft, one should not rely on post-mortem misuse identification [10] – it should be prevented with technological countermeasures that provide strong template security and user's privacy protection. Additionally, the performance of the biometric system should be downgraded to the reasonable level after introducing these countermeasures to the system, i.e. they are expected not to degrade the

verification accuracy to unacceptable level or introduce severe computational costs or storage requirements.

2 Approaches to Biometric Template Protection

One approach to biometric template security and privacy is cancelable biometrics. Two main categories of cancelable biometrics can be distinguished: intentional distortion of biometric features with non-invertible transforms [11], such as block permutation of iris texture, and biometric salting. Cancelable biometrics that employs non-invertible transforms is based on application of the same transformation to a given biometric sample during enrollment and verification. There are a large number of non-invertible transforms for variety of biometric modalities, and some of them operate with the key. Having said that, each compromised template can easily be revoked and another transformation can be applied during re-enrollment; if transformation operates with the key, only the key is changed. Examples of cancelable transforms applicable to fingerprint and iris are given in [12] and [13], respectively. However, non-invertible transforms may be partially reversible and they usually degrade overall verification accuracy, thus they are not a fail-safe solution to a biometric template protection problem. Biometric salting refers to transformations of biometric templates that are selected to be invertible, where any transformation is considered to be an approach to biometric salting even if templates have been extracted in a way that it is not feasible to reconstruct the original biometric signal [14]. Although biometric salting does not degrade the verification accuracy, non-invertible transfors provide higher level of security. Hence, biometric salting is not a fail-safe solution to the problem either.

Another approach to providing biometric template security and privacy is the application of homomorphic encryption schemes [15, 16]. Homomorphic encryption refers to cryptographic algorithms that allow some computations to be performed in the encrypted domain. Research on homomorphic encryption algorithms that support both addition and multiplication based on lattice encryption was expected to provide novelties in biometric template security [17], but no results were reported in relevant literature. Although applicable in theory (e.g. homomorphic encryption appears to be suitable for application in systems that employ bitwise XOR to calculate Hamming distance during verification between two binary iris templates), there are two reasons why it is not practical: the encrypted template is large and the system is computationally expensive. Reader may consult [16] for more details.

The main contribution of this paper is a secure modular authentication system based on iris biometrics applicable to mobile banking. An approach presented in this paper employs public key cryptography, pseudorandom number generators

and cancelable biometrics. Non-invertible transformation operates with the key stored on a device's trusted storage. The system does not suffer from the drawbacks of homomorphic encryption as cryptographic operations are not computationally expensive and no large templates are created. Biometric templates are encrypted or at least cancelable during all stages of operation (excluding feature extraction) resulting in a system prone to a variety of attacks. Having in mind that the system satisfies requirements set to a cryptographically secured biometric system that provides strong privacy protection listed in [10], and that devices with iris scanners are emerging technology, we can conclude that this modular system is suitable for implementation in mobile banking.

3 Counterfeiting Attacks with Modular Architecture

Biometric authentication systems consist of four modules: sensor, feature extractor, matcher and template database. If these modules reside on one device, authentication system is vulnerable to variety of attacks [18]. These include fake biometrics, replay attack, attack on the feature extraction module, attack on the channel between feature extractor and matcher, compromising the database, attack on the communication channel between template database and the matcher and overriding the result declared by the matcher module. Some of these attacks are easy to execute if the system is not properly designed. For example, if the system does not employ liveness detector, it is easy to perform sensor attack with fake biometrics. To prevent execution of aforementioned attacks, entire system is split into two high-level modules, residing on two devices. Additionally, both cancelable biometrics and strong cryptographic protection are introduced to the system. Modular system now consists of multiple clients (devices used to capture biometrics, manage auxiliary data and create encrypted cancelable templates) and an authentication server (device that manages encryption keys and verifies cancelable templates). As proposed system deals with the iris biometrics, which employs XOR operation to verify a person, a cancelable transform that partially reassembles one-time-pad cypher (the key is employed more than once, but is of the same length as plaintext) is used.

Aside from cryptographic security, system is expected to provide strong privacy protection, resulting in following set of requirements: (1) biometric templates remain encrypted or at least cancelable during all stages of storage, transmission and verification (e.g. authentication server should never obtain plaintext templates,) and (2) no client is allowed to access private keys stored on authentication server as it may compromise the security of the stored templates. Further, system should be resilient to a template substitution and all low level attacks, it should not suffer from severe computational drawbacks and cryptographic countermeasures should not degrade overall accuracy (i.e. they

should not introduce additional false acceptance or false rejection rates to the system).

4 Proposed Modular Authentication Framework

A framework for modular authentication systems based on conventional XOR biometrics, such as iris, is presented in this section [19]. Conventional XOR biometrics is based on Hamming distance calculation between templates obtained during enrollment and verification phases. Hamming distance is chosen as verification metrics as it is suitable for application of one-time-pad partially based non-invertible transforms of the template, i.e. simple XOR operation with the non-invertible transform key of the same length as the original template. This method of biometric template protection guards the end user from identity theft and allows the user to easily re-enroll with another key, if any suspiction about the key being compromised occurs.

During the enrollment phase, the user provides numeric user ID and non-invertible transform key K_t to the client. Let H(x) denote the hash function (one may select solution-specific), ID the identity of the user and K_{priv}, K_{pub} the private and the public key, respectively. Hash of the user ID is calculated on the client and sent to the authentication server. Authentication server generates a keypair (K_{priv}, K_{pub}), stores the private key with hash of user ID (H(id), K_{priv}) and sends public key to the client. Client obtains biometrics, creates a binary template b_0, and generates cancelable binary template $b = K_t \oplus b_0$. Client generates random seed s_0 and encrypts it with the public key: $s_E = E(s_0, K_{pub})$, where E denotes the encryption operation. Any public-key encryption algorithm that suffice the principles behind the information theory and strong cryptography can be used. Client generates a keystream s = PRNG (s_0) using pseudorandom number generator and given seed, where PRNG denotes applicable pseudorandom number generator. Client calculates $s \oplus b$, stores (H(id), s_E, $s \oplus b$) and discards the rest of the data.

During the verification phase, the user provides numeric user ID and non-invertible transform key K_t to the client. Client obtains biometrics, creates a template b_0' and generates cancelable binary template $b' = K_t \oplus b_0'$. Client calculates user ID hash and retrieves values s_E and ($s \oplus b$) from stored record (H(id), s_E, $s \oplus b$) with the corresponding user ID hash. Client calculates $s \oplus b \oplus b'$ and sends it with the encrypted seed s_E to the authentication server. Hash of the user ID calculated on the client is sent to the authentication server. Authentication server retrieves private key from stored record (H(id), K_{priv}) with the corresponding user ID hash. Let D denote the decryption operation. Authentication server decrypts the seed by doing $s_0 = D(s_E , K_{priv})$ and generates the keystream: s' = PRNG (s_0). As pseudorandom number generator is deterministic and same seed is used to generate keystreams both in enrollment and

verification phases, generated keystreams s and s' will be identical, i.e. $s = s$'. Aside from this, the same non-invertible transform key K_t is used is both phases. Thus, server calculates $s \oplus b \oplus s' \oplus b' = b \oplus b' = K_t \oplus b_0 \oplus K_t \oplus b_0' = b_0 \oplus b_0'$ and compares the Hamming distance between templates b_0 and b_0' with the threshold. According to that result, the decision is made (user is genuine or imposter) and sent back to the client. One should note that, although the result of comparison is the Hamming distance between original, unaltered templates obtained via feature extractor, server makes the calculation using cancelable templates generated with the non-invertable transform key.

4.1 Security Evaluation of the Proposed Framework

Security of the system may be summarized as follows. Templates are encrypted or at least cancelable during all stages of storage, transmission and verification, and the client is not allowed to access private keys stored on authentication server, which satisfies the conditions set for an ideal biometric system. System employs two factor authentication thus making an imposter with helper data virtually impossible to claim as genuine user. If templates stored on a client are somehow compromised, re-enrollment with another transform key and encryption key-pair will remediate the situation. Substitution attacks cannot be performed, as the public key is discarded at the end of enrollment. As an adversary cannot recreate the keystream s from the encrypted seed s_E and the public key, system is resilient to most of the attacks on the biometric encryption systems.

5 Implementation in the Mobile Banking Authentication Scenario

Authentication server resides in the bank. As authentication server stores encryption keys, it is logical that encrypted templates reside on the client. This prevents an attacker who obtains illegal access to authentication server to decrypt templates. The client is a mobile device (smartphone or a tablet) with an iris scanner. Additional software that provides feature extraction and cryptographic operations is installed on the client (as an additional application provided by the bank). Non-invertible transform key is stored on the device. User obtains this key from the bank as an output of true random number generator; the length of the key must be equal to the length of iris template as the XOR of original template and the key is performed straight after the feature extraction. User is allowed to wipe both the key and the data stored during enrollment phase both locally, if he suspects the data is somehow compromised, and remotely, if the device gets stolen. The bank is allowed to do remote data wiping also, if the authentication server is somehow compromised.

During the enrollment phase the system operates as depicted in Figure 1.

- Client-side application calculates hash of the devices' IMEI and sends it to the authentication server. Devices' IMEI is hashed to protect user's privacy – hashing prevents plaintext transmission between the client and the server as well as the storage of user-sensitive plaintext data on the server-side.
- Server generates a private-public keypair (K_{priv}, K_{pub}), stores the private key with hash of IMEI (H(IMEI), K_{priv}) and sends public key to the mobile device.
- User provides iris biometrics to the mobile device. Client-side application creates a binary iris template b_0 (as explained in section 5.1 of this paper) and generates cancelable binary template $b = K_t \oplus b_0$ using non-invertible transformation key stored on the device. Client-side application further generates random seed s_0 and encrypts it with the public key: $s_E = E(s_0, K_{pub})$. Application generates a keystream s = PRNG (s_0) using pseudorandom number generator and given seed, calculates $s \oplus b$, stores values (s_E, $s \oplus b$) on the device and discards the rest of the data.

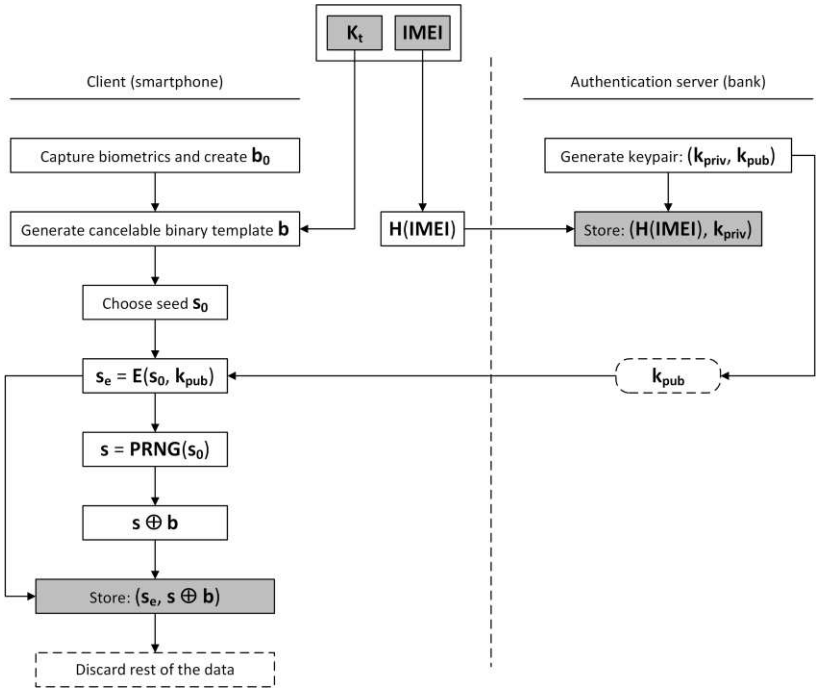


Figure 1
Enrollment phase

During the verification phase the system operates as depicted in Figure 2.

- Hash of the device IMEI is calculated on the client-side application and sent to the authentication server.
- User provides biometrics to the mobile device. Client-side application creates binary iris template b_0' and generates cancelable binary template $b' = K_t \oplus b_0'$. Application retrieves values s_E and $(s \oplus b)$, calculates $s \oplus b \oplus b'$ and sends it with the encrypted seed s_E and hash of the devices' IMEI to the authentication server.
- Server retrieves private key from stored record (H(IMEI), K_{priv}) with the corresponding device IMEI hash, decrypts the seed with the private key by doing $s_0 = D(s_E , K_{priv})$ and generates the keystream: $s' = PRNG(s_0)$. As stated in section 4, due to deterministic nature of PRNGs, same seeds will produce identical keystreams s and s', and the same key K_t is used during enrollment and verification. Authentication server further calculates $s \oplus b \oplus s' \oplus b' = b \oplus b' = K_t \oplus b_0 \oplus K_t \oplus b_0' = b_0 \oplus b_0'$. As with the framework, the result of comparison is the Hamming distance between original, unaltered templates, but the server makes the calculation using cancelable templates (thus having no access to original ones, nor to the non-invertible transform key).

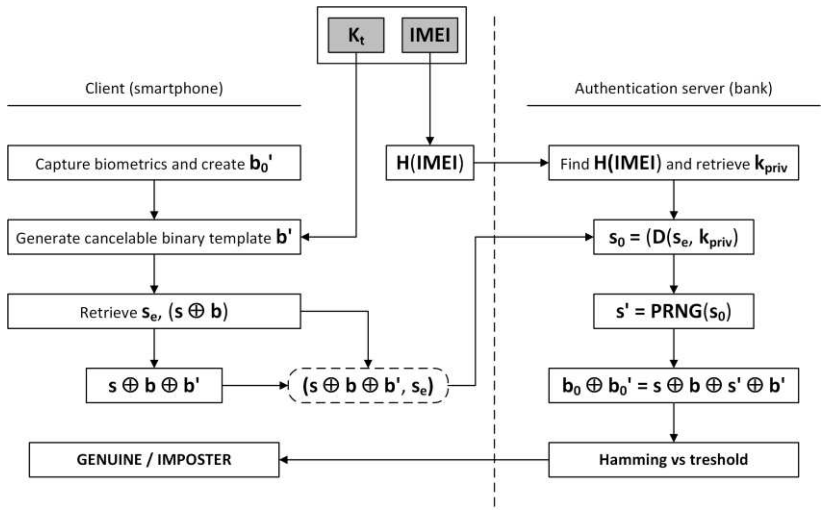


Figure 2
Verification phase

Pseudocodes for enrollment and verification phases are listed below.

User enrollment algorithm

INPUT: b – plaintext biometric template, K_t – transform key

OUTPUT: $s \oplus b$ – encrypted biometric template, s_E – encrypted seed

Client:

1. send (H(IMEI))
2. $b = K_t \oplus b_0$
3. random (s_0); s = PRNG (s_0)
4. get (K_{pub}); s_E = E(s_0, K_{pub})
5. store (s_E, $s \oplus b$)

Server:

1. get (H(IMEI))
2. generate (K_{priv}, K_{pub})
3. send (K_{pub}); store (H(IMEI), K_{priv})

User verification algorithm

INPUT: b' – plaintext biometric template, K_t – transform key, t - threshold

OUTPUT: decision

Client:

1. send (H(IMEI))
2. $b' = K_t \oplus b_0'$
3. send ($s \oplus b \oplus b'$, s_E)
4. get (decision)

Server:

1. get (H(IMEI), $s \oplus b \oplus b'$, s_E)
2. s_0 = D(s_E , K_{priv}); $s' = PRNG (s_0)$
3. $b_0 \oplus b_0' = s \oplus b \oplus s' \oplus b'$
4. if ($b_0 \oplus b_0'$) < t then decision = "genuine" else decision = "imposter"
5. send (decision)

5.2 Generating Binary Iris Templates

For more details on iris feature extraction methods reader may consult [20]. Before the template is generated from extracted features, acquired iris image must be pre-processed. Outer radius of iris patterns and pupils are first localized with Hough transform that involves a canny edge detector to generate an edge map. Poorly localized iris will result in unsuccessful segmentation and poor reproducibility of the template, which further results in high false rejection rates. Hough transform identifies positions of circles and ellipses [21] – it locates contours in an *n*-dimensional space by examining whether they lie on curves of a specified shape. Hough transform for outer iris and pupil boundaries and a set of *n* recovered edge points (x_i, y_i) is defined by:

$$H\left(x_c, y_c, r\right) = \sum_{i=1}^{n} h\left(x_i, y_i, x_c, y_c, \mathrm{r}\right), \tag{1}$$

$$h\left(x_i, y_i, x_c, y_c, r\right) = \begin{cases} 1, \left(x_i - x_c\right)^2 + \left(y_i - y_c\right)^2 - r^2 = 0 \\ 0, \left(x_i - x_c\right)^2 + \left(y_i - y_c\right)^2 - r^2 \neq 0 \end{cases}. \tag{2}$$

The circle (x_c, y_c, r) through each edge point (x_i, y_i) is defined as:

$$\left(x_i - x_c\right)^2 + \left(y_i - y_c\right)^2 = r^2 . \tag{3}$$

The triplet that maximizes H (x_c, y_c, r) is common to the greatest number of edge points and is a reasonable choice to represent the contour of interest [22]. Once an iris image is localized, regions of interests are defined and it is transformed into fixed-size rectangular image. The normalization process employs Daugman's rubber sheet model that remaps the iris image $I(x, y)$ from Cartesian to polar coordinates [20]:

$$I\left(x\left(r,\theta\right), y\left(r,\theta\right)\right) \rightarrow I(r,\theta) . \tag{4}$$

Parameter r is on the interval [0, 1] and θ is the angle [0, 2π]. If iris and pupil boundary points along θ are denoted as (x_i, y_i) and (x_p, y_p), respectively, the transformation is performed according to:

$$x\left(r,\theta\right) = \left(1-r\right) x_p\left(\theta\right) + r x_i\left(\theta\right), \tag{5}$$

$$y\left(r,\theta\right) = \left(1-r\right) y_p\left(\theta\right) + r y_i\left(\theta\right). \tag{6}$$

The rubber sheet model does not compensate rotational inconsistencies, but it takes into account pupil dilation size inconsistencies in order to produce a normalized representation with constant dimensions [23] set by angular and radial resolution. Angular resolution is set by number of radial lines generated around

the iris region, while radial resolution refers to the number of data points in the radial direction.

Although various extraction methods are reported in the literature, discriminant features are extracted from a normalized iris using conventional method based on Gabor filtering. This method is validated as suitable feature extraction method in various researches presented by other authors. Normalized image is broken into a number of 1-D signals that are convolved with 1-D Gabor wavelets. The frequency response of 1-D log-Gabor filter [24] is given by:

$$G(f) = \exp\left(-\left(\log \frac{f}{f_0} \right)^2 \Big/ 2\left(\log \frac{\sigma}{f_0} \right)^2 \right), \tag{7}$$

where f_0 denotes center frequency, and σ the bandwidth of the filter. Phase quantization is applied to four levels on filtering outputs (each filter produces two bits of data for each phasor) and the quantized phase data is used to encode an iris pattern into a bit-wise biometric template. The number of bits in the biometric template depends on angular and radial resolution and the number of used filters, while the template entropy depends on the number of used filters, their center frequencies and the parameters of the modulating Gaussian.

5.3 Performance Evaluation of the Proposed System

Performance of the proposed system depends on various factors, such as the quality of the camera and illumination, as well as the parameters of employed feature extraction algorithms. It is very important to state that in our mobile banking scenario the accuracy of the system does not depend on the cryptographic protection and cancelable biometrics – they introduce no additional false acceptance or false rejection rates. Majority of the experiments on iris verification reported in the literature employ CASIA-Iris database, collected by the Chinese Academy of Sciences' Institute of Automation [25]. However, in order to get the realistic picture on how iris verification works with smartphones, a custom dataset is created using Huawei P10 Lite front camera. Images were subsequently processed in MATLAB (version R2016a). The iris image dataset used in our experiments consists of 210 gray-scale samples from 10 subjects obtained outdoors and indoors with different illumination. Each iris image is normalized into an 8-bit 240x20 pixel image, and a 1-D log-Gabor filter with σ=0.5 and 12 pixel center wavelength is subsequently applied, resulting in a 9600 bit template. These parameters were found to provide high local entropy and optimum encoding [26, 27]. One randomly selected outdoor image for each subject is used to enroll the user and all images are used to verify them. Results of experimental evaluation are given in Table 1.

Table 1
Experimental evaluation on realistic dataset (iris images captured by smartphone's front camera)

Scene	FAR	FRR
Low threshold (reducing FAR)		
Outdoors (daylight)	0 %	2 %
Indoors (normal illumination)	0 %	2 %
Indoors (medium illumination)	0 %	4 %
Indoors (poor illumination)	0 %	18 %
High threshold (reducing FRR)		
Outdoors (daylight)	0 %	0 %
Indoors (normal illumination)	0 %	0 %
Indoors (medium illumination)	2 %	2 %
Indoors (poor illumination)	6 %	4 %

Although verification with low threshold values fails indoors with poor illumination, this is not something we consider to be the drawback, as user is allowed to retry. The real problem occurs if the threshold is high, as user may still be verified as genuine, even if larger number of bits differ between two templates. This results in occurrence of false acceptance with medium illumination (less than 450 lumens, approximately one 9-11 watts compact fluorescent lamp illuminating 25 square meters sized room), or poor illumination (less than 200 lumens, i.e. one 3-5 watts compact fluorescent lamp illuminating the room of the same size). In other words, if the treshold is to high and the illumination is inappropriate, system enters the danger zone and is no more applicable to the mobile banking due to occurrence of false acceptance rates. Outside of that zone, it operates stable and may only require additional authentication attempt(s). Having said that, it is necessary to keep the verification threshold as low as possible to avoid false acceptance.

The concrete threshold depends on the camera used to capture iris image, and it should be set on the client-side application automaticaly (if the pre-calculated optimal threshold for the concrete device exists in records on devices previously used for that purpose) or by bank's authorized officer, during the first enrollment (if pre-calculated data does not exist for the concrete model). The later one should employ several captures of user's iris, a set of irises belonging to different persons and decidability as the metric, which takes into account the mean and standard deviation of the intra-class and inter-class distributions. The overall decidability of iris recognition is revealed by comparing Hamming distance distributions for same versus for different irises [20]. Users should not be allowed to set this value by themselves.

Another issue of iris verification system is the presence of contact lenses. Contact lenses, particularly textured ones, obfuscate the natural iris patterns, thus

presenting a challenge to the iris verification. Effects of contact lenses on iris verification systems were analyzed in [28, 29]. Yadav et al. [29] presented lens detection algorithm that can be used to reduce the effect of contact lenses, stating that their approach outperforms other lens detection algorithms and provides improved iris recognition performance.

5.4 Security Evaluation of the Proposed System

Regarding security of the proposed mobile banking authentication solution, same conclusions can be made as with the framework it is built upon. Templates are encrypted or at least cancelable during all stages of operation, and the mobile device is not allowed to access private keys stored on authentication server. Authentication server has no access to the transform keys and cancelable templates created on the mobile device during enrollment. If the phone is stolen, an adversary cannot claim as legitimate user as the system is prone to all attacks listed in [18] as well as to hill-climbing [30], non-randomness [31], re-usability [32], blended substitution [33] and linkage attack [34].

Although some key-exchange protocols may be introduced to the system, the most secure way to distribute non-invertible transform key is to make the user obtain it directly from the bank, as it eliminates chances of identity spoof (which may cause further social engineering attacks). Both the user and the bank are allowed to remotely wipe all stored data (including the key) if the phone gets lost or stolen. If the device is somehow returned to the owner, he or she may retrieve new non-invertible key from the bank and undergo re-enrollment procedure. During the re-enrollment, user will provide new biometric sample to the device, client-side application will generate new seed for pseudorandom number generator, and bank's authentication server will generate new private-public keypair that will further be used to encrypt and decrypt the seed.

One should note that physical access to the device does not allow an adversary to retrieve the non-invertible transformation key. Latest smartphone models shipped with biometric sensors that operate with several modalities (e.g. fingerprint, face and iris) running an Android operating system (version 6 or higher) include countermeasures that prevent physical acquisition of sensitive data, even with the state of the art forensic tools and devices. This fact originating from digital forensics provides us with sufficient level of security when certain amount of sensitive information is stored on the device, such as this non-invertible transform key.

The data being transmitted over the network (reassembling the Alice-Bob scenario used to explain cryptographic protocols) and stored on smartphones and the bank server is depicted in Figure 3. According to Figure 3, the following values are transmitted: user-specific public key (just one time, during enrollment), hash of the users' IMEI (during enrollment and during each verification), and the XOR of

enrolled and verification cancelable templates with the keystream (during each verification): $s \oplus b \oplus b'$. We could not identify any possible weakness that would allow an adversary to extract information from the transmitted data. One thing, however, is very important to state – the choice of poor pseudorandom number generator may lead to small leakage of information when $s \oplus b \oplus b'$ is transmitted from client to the server. Hence, one should use the cryptographically strong generator that is highly entropic in the information theory sense.

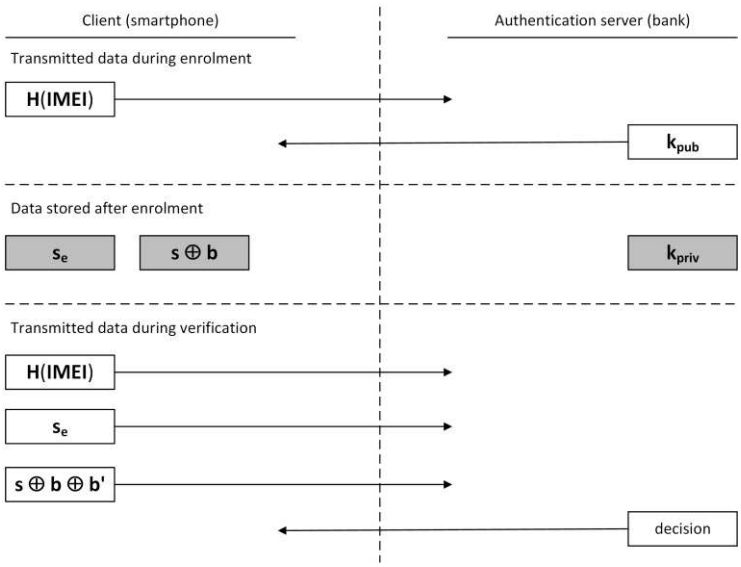


Figure 3
Data being transmitted and data being stored

Additionally, one should note that the security of the entire system depends also on the security of iris recognition subsystem. Although iris scanners should be hard to trick into false acceptance, a group of hackers managed to do so with the Galaxy S8 iris-based authentication; hardware required to complete the attack cost less than the smartphone itself [35]. If the phone does not employ fake iris countermeasures, similar scenarios may occur – for example, data extracted from selfies found on the stolen phone with performant cameras may be used to obtain fake iris images. Several approaches have been proposed to detect fake irises. An approach to iris contact lens detection based on deep image representations [36] uses a convolutional network to build a deep image representation and an additional fully connected single layer with softmax regression for classification. Sinha et al.'s iris liveness detection approach [37] employs Flash and motion detection of natural eye in order to detect the liveliness of real iris images before matching from stored templates, thus significantly increasing security and

reliability of the system. A solution suitable for implementation in mobile devices was proposed by Gragnaniello, et al. [38]; a fast and accurate technique to detect printed-iris attacks is based on the local binary pattern descriptor (LBP). Their algorithm encompasses three steps: computation of the high-pass image residual, feature extraction based on a suitable LBP descriptor and classification with support vector machines with a linear kernel. According to authors, the detection performance is extremely promising, despite the very low complexity.

Conclusions

An implementation of modular authentication system based on iris biometrics into mobile banking scenario was presented in this paper. Strong cryptography that is not bound to a specific public key algorithm or pseudorandom number generator and bitwise XOR cancelable biometrics were introduced to the modular system in order to prevent execution of number of attacks on classical biometric and biometric encryption systems. Employed cryptographic countermeasures do not degrade the verification accuracy and do not introduce severe computational costs. According to security evaluation of the system, results of the experiments conducted with realistic dataset, and the fact that devices with iris scanners are emerging technology, we conclude that this modular architecture is highly applicable in mobile banking scenario. The only drawback of the proposed modular authentication framework is it's limitation to biometric modalities that are verified by calculating Hamming distance. Although it is applicable to iris and, conditionally, fingerprint [33], we will focus our further research into developing authentication systems that can employ other biometric modalities, e.g. systems based on speaker recognition and face recognition. Additionally we will evaluate the application of fake iris detection approaches presented in [37] and [38] in our system in order to raise overall level of security, as well as the application of lens detection algorithm [29] to reduce the effect of contact lenses and increase verification accuracy if subjects wear lenses.

References

[1] M. Mannan, P. C. Van Oorschot: Security and Usability – The Gap in Real-World Online Banking. NSPW'07, North Conway, NH, USA, Sep. 18-21, 2007

[2] Y. S. Lee, N. H. Kim, H. Lim, H. Jo, H. J. Lee: Online banking authentication system using mobile-OTP with QR-code. In Proc. 5th International Conference on Computer Sciences and Convergence Information Technology (ICCIT), November 2010, pp. 644-648, IEEE

[3] C. Stamford: Gartner Says 30 Percent of Organizations Will Use Biometric Authentication for Mobile Devices by 2016. February 4, 2014, available online, last time visited April 2018

[4] A. K. Jain, A. Ross, S. Prabhakar: An Introduction to Biometric Recognition. IEEE Transactions on Circuits and Systems for Video Technology, Vol. 14, pp. 4-20, 2004

[5] A. K. Jain, A. Ross: Introduction to Biometrics. In Handbook of Biometrics, A. Jain et al. (Eds), Springer, 2008

[6] Y. C. Feng, P. C. Yuen, A. K. Jain: A Hybrid Approach for Face Template Protection. In Proceedings of SPIE Conference of Biometric Technology for Human Identification, Orlando, USA, Vol. 6944, pp. 325, 2008

[7] P. Balakumar, R. Venkatesan: A Survey on Biometrics-based Cryptographic Key Generation Schemes. International Journal of Computer Science and Information Technology & Security, Vol. 2, No. 1, pp. 80-85, 2012

[8] A. K. Jain, K. Nandakumar, A. Nagar: Biometric Template Security. EURASIP J. Adv. Signal Process, 2008:1-17, 2008

[9] J. Galbally, C. McCool, J. Fierrez, S. Marcel, J. Ortega-Garcia: On the Vulnerability of Face Verification Systems to Hill-Climbing Attacks. Pattern Recognition, 43(3) pp. 1027-1038, 2010

[10] A. Stoianov: Cryptographically secure biometrics. In SPIE Defense, Security, and Sensing, International Society for Optics and Photonics, 2010

[11] N. Maček, B. Đorđević, J. Gavrilović, K. Lalović: An Approach to Robust Biometric Key Generation System Design. Acta Polytechnica Hungarica, Vol. 12, No. 8, pp. 43-60, 2015

[12] N. K. Ratha, S. Chikkerur, J. H. Connell, R. M. Bolle: Generating Cancelable Fingerprint Templates. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 29(4), pp. 561-572, 2007

[13] J. Zuo, N. K. Ratha, J. H. Connell: Cancelable iris biometric. In Pattern Recognition, ICPR 2008, 19th International Conference on (pp. 1-4), IEEE, 2008

[14] C. Rathgeb, A. Uhl: A survey on biometric cryptosystems and cancelable biometrics, EURASIP Journal on Information Security 2011, 2011:3, open access, no pagination

[15] J. Bringer, H. Chabanne: An authentication protocol with encrypted biometric data. In International Conference on Cryptology in Africa, pp. 109-124, Springer Berlin Heidelberg, 2008

[16] B. Schoenmakers, P. Tuyls: Computationally secure authentication with noisy data. Chapter 9 in P. Tuyls, B. Škorić, T. Kevenaar, eds., Security with Noisy Data: Private Biometrics, Secure Key Storage and Anti-Counterfeiting, Springer-Verlag, London, pp. 141-149, 2007

[17] C. Gentry: Fully Homomorphic Encryption Using Ideal Lattices. 41st ACM Symposium on Theory of Computing (STOC), pp. 169-178, 2009

[18] R. Jain, C. Kant: Attacks on Biometric Systems – An Overview. International Journal of Advances in Scientific Research, 1(07), pp. 283-288, 2015

[19] N. Maček, M. Milosavljević, I. Franc, M. Bogdanoski, M. Gnjatović, B.Trenkić. Secure Modular Authentication Systems Based on Conventional XOR Biometrics. In Proc. of the 9th Int. Conf. on Business Information Security (BISEC2017), Belgrade, October 18th, 2017, pp. 27-32

[20] J. Daugman: How iris recognition works. Circuits and Systems for Video Technology, IEEE Transactions on, 14(1) pp. 21-30, 2004

[21] D. J. Kerbyson, T. J. Atherton: Circle Detection using Hough Transform Filters. Fifth International Conference on Image Processing and its Applications, Edinburgh, UK, 04 – 06 July 1995, pp. 370-374

[22] R. P. Wildes: Iris Recognition – an Emerging Biometric Technology. Proceedings of the IEEE, 85(9) pp. 1348-1363, 1997

[23] G. Amoli, N. Thapliyal, N. Sethi: Iris Preprocessing. International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 2, No. 6, pp. 301-304, 2012

[24] D. J. Field: Relations between the Statistics of Natural Images and the Response Properties of Cortical Cells. Journal of the Optical Society of America, Vol. 4, No. 12, 1987

[25] Biometrics Ideal Test, http://biometrics.idealtest.org

[26] S. Adamović, M. Milosavljević: Information Analysis of Iris Biometrics for the Needs of Cryptology Key Extraction. Serbian Journal of Electrical Engineering, Vol. 10, No. 1, pp. 1-12, 2013

[27] S. Adamović, M. Milosavljević, M. Veinović, M. Šarac, A. Jevremović: Fuzzy commitment scheme for generation of cryptographic keys based on iris biometrics. IET Biometrics, Vol. 6, No. 2, pp. 89-96, 2017

[28] J. S. Doyle, K. W. Bowyer, P. J. Flynn: Variation in accuracy of textured contact lens detection based on sensor and lens pattern. In Biometrics: Theory, Applications and Systems (BTAS), 2013 IEEE Sixth International Conference on, pp. 1-7, 2013, IEEE

[29] D. Yada, N. Kohli, J. S. Doyle, R. Singh, M. Vatsa, K. W. Bowyer: Unraveling the effect of textured contact lenses on iris recognition. IEEE Transactions on Information Forensics and Security, Vol. 9, No. 5, pp. 851-862, 2014

[30] A. Adler: Vulnerabilities in Biometric Encryption Systems. LNCS, Springer 3546, pp. 1100-1109, 2005

[31] E.-C. Chang, R. Shen and F. W. Teo: Finding the Original Point Set Hidden among Chaff. In Proc. ACM Symp. ASIACCS'06, Taipei, Taiwan, pp. 182-188, 2006

[32] X. Boyen: Reusable cryptographic fuzzy extractors. In Proc. 11th ACM Conf. CCS, Washington, DC, pp. 82-91, 2004

[33] W. J. Scheirer, T. E. Boult: Cracking Fuzzy Vaults And Biometric Encryption. Biometric Consortium Conference, Baltimore, September 2007

[34] A. Cavoukian, A. Stoianov: Biometric Encryption – The New Breed of Untraceable Biometrics. In N.V Boulgouris et al., eds., Biometrics: fundamentals, theory, and systems, Wiley-IEEE Press, pp. 655-718, 2009

[35] D. Goodin: Breaking the iris scanner lock in Samsung's Galaxy S8 is laughably easy. Ars Technica, May 23, 2017, available online, last time visited December 2018

[36] P. Silva, E. Luz, R. Baeta, H. Pedrini, A. X. Falcao, D. Menotti, D: An approach to iris contact lens detection based on deep image representations. In Graphics, Patterns and Images (SIBGRAPI), 28th SIBGRAPI Conference on, 2015, pp. 157-164, IEEE

[37] V. K. Sinha, A. K. Gupta, Manish Mahajan: Detecting fake iris in iris biometric system. Digital Investigation, Vol. 25, pp. 97-104, 2018

[38] D. Gragnaniello, C. Sansone, L. Verdoliva: Iris liveness detection for mobile devices based on local descriptors. Pattern Recognition Letters, Vol. 57, pp. 81-87, 2015

[39] S. Barzut, M. Milosavljević: Jedan metod formiranja XOR biometrije otisaka prstiju Gaborovom filtracijom. In Sinteza 2014 – Impact of the Internet on Business Activities in Serbia and Worldwide, Belgrade, Singidunum University, Serbia, pp. 610-615, 2014

Hierarchical Spiral Discovery Networks for Multi-Layered Exploration-Exploitation Tradeoffs

Adam B. Csapo

Department of Informatics, Széchenyi István University
Győr, Hungary
csapo.adam@sze.hu

Abstract: The Spiral Discovery Network (SDN) was recently proposed as a tool for automated parametric optimization based on the Spiral Discovery Method. SDN can be seen as a heuristic optimization approach that offers tradeoffs between exploration and exploitation without having recourse to explicit gradient-based feedback information (unlike classical neural networks) and without requiring hand-coded representations of metaheuristic constructs such as genotypes (unlike genetic algorithms). In this paper, the properties of the SDN model are further explored, and two extensions to the model are proposed. The first extension corrects a shortcoming of the original model and has to do with the assignment of credit among different output components based on the most recent performance of the model at any given time. The second extension consists of using multiple SDN cells in a hierarchical architecture, which enables a fuller and more effective exploration of the parametric space. The improvements provided by the two extensions are validated on the same set of simulations discussed in earlier work.

Keywords: Spiral Discovery Network; Spiral Discovery Method; Non-convex optimization; Exploration versus exploitation

1 Introduction

The debate on whether evolutionary methods or neural networks are better suited to tackle problems in artificial intelligence is strongly related to the classical debate on nature versus nurture [5]. Clearly, both classes of approach are important components of human intelligence, and both have their pros and cons from a computational point of view. Evolutionary methods are well suited to finding good candidate solutions in large parametric spaces, however they often rely on the availability of hand-coded genotypes that are expected to yield useful results as the classical operations of recombination and genetic mutation are applied to them – a constraint that can be considered as arbitrary and limiting given that it is not directly relevant to the problem domain itself. At the same time, neural networks are known to be capable of automatically generating high-dimensional input encodings that are often useful

for the problem at hand, thus alleviating the need for hand-coded inputs; however, their applicability rests on the assumption that their performance can be evaluated using a loss function that is effectively computable and differentiable.

In cases where the loss function associated with a problem is unknown or difficult to compute and / or differentiate, the problem of assigning credit to the many individual components involved in the functionality of a neural network – or other parametric model – becomes intractable (for more on the credit assignment problem see e.g. [14, 9]). In such cases, methods based on coarser-grained feedback, such as reinforcement learning or direct search in weight space are often used [14]. The Spiral Discovery Network (SDN) model can be considered as an alternative approach with the following properties [3, 4]:

- Instead of relying on evolutionary operations such as recombination and mutation, SDN operates directly within the search space, and so the requirement of creating useful, hand-coded input encodings is alleviated;
- To compensate for the lack of (genetically-motivated) operations for generating new candidate solutions, the model explores the search space along a parametric hyper-spiral structure;
- The application of this hyper-spiral structure in turn implicitly generates differential feedback information (besides the coarse-grained feedback obtained via each candidate solution), allowing the model to adapt its behavior in subsequent cycles of exploration.

The hierarchical SDN model proposed in this paper goes a step further by searching both in the parametric space of the problem domain and the space of the model parameter domain, while allowing for the performance of the latter to be informed by the performance of the former. Generally speaking, hierarchical models offer improvements over flat architectures by enabling different parts of a problem to be tackled at different levels, and for partial solutions found in lower levels to be reused higher up in the architecture. As will be shown through a simulation example, information gleaned through individual cycles of SDN exploration can be used to find better hyper-parameters for the model in the hierarchical extension proposed in this paper. This in turn allows the model to keep finding improvements to earlier candidate solutions, without getting stuck in local minima.

The paper is structured as follows. Section 2 gives an overview of the theoretical background behind parametric optimization, in order to further highlight the motivations behind SDN. Section 3 briefly recapitulates the simplest (non-hierarchical) version of SDN proposed in [3]. Section 4 provides a brief analysis on the operation of the original model based on a simulation introduced in an earlier publication. Based on the analysis, two extensions to the model are proposed in Sections 5 and 6. Finally, the results of the paper are summarized in the Conclusions section.

2 Theoretical Motivations behind SDN

Non-convex optimization is a broad field of mathematics that has applications in engineering tasks where the goal is to find sufficiently good solutions on high-dimensional parametric manifolds. One of the most relevant application examples for non-convex optimization today is finding useful architectures for (deep) neural networks or other kinds of graphical models. The usual way to solve such challenges is to search in iterations based on the gradient of a globally defined loss function – an approach referred to as gradient descent [11].

The general idea of gradient descent can be highly successful on parametric landscapes that are associated with a clearly defined cost function, and contain no more than a small number local minima in terms of that function. However, as soon as the value of a cost function becomes difficult to interpret, or the cost function becomes so intractable that it is computationally difficult to determine its gradients, and / or it produces an intractably large number of local minima, the naive solution of gradient-based iterative optimization often starts to break down.

The problem of dealing with local minima can be addressed to some degree by finding good trade-offs between exploration and exploitation, i.e. by modifying the gradient descent approach slightly to counteract situations where the optimization process might slow down or stop. This approach is reflected in a host of existing solutions. One fruitful idea was to experiment with the scaling factor of the gradient – for example by making it adaptive to changes in sign via the concept of "momentum" [13, 12, 15], or by making it specific to the different dimensions in the parameter space [6, 8]. Other ideas include the normalization of inputs across layers and batches (specifically in training neural network models) [7], or simply adding noise to the gradients [10].

Despite the availability of such ingenious solutions, the requirement of a loss function that is defined globally, easy to compute and also differentiate may be too limiting in certain applications. The original motivation behind the spiral discovery approach that is further extended in this paper came from the problem domain of designing useful multimodal user interfaces [2, 1]. In applications where the performance of a system has to be tuned based on user feedback, it is obviously difficult to obtain feedback at a high input resolution (this would be tedious work for users providing the feedback) as well as at high output resolution (providing consistent evaluations over a period of time is difficult for human subjects). The original Spiral Discovery Method and its various extensions – including the ones proposed in this paper – render such processes of optimization more tractable by limiting the number and resolution of feedbacks necessary, and by compensating for the lack of detailed information through the structure of parametric discovery. It is worth noting that this approach fits well into the framework of interactive evolutionary computation [16, 17].

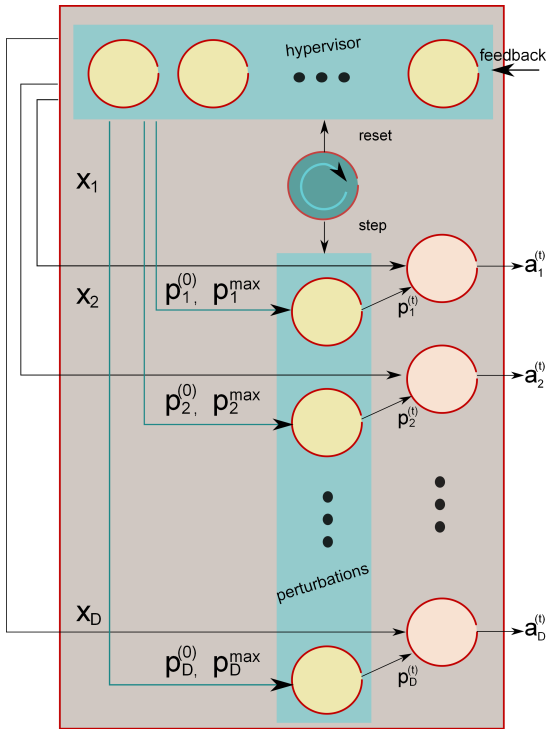


Figure 1
Neural network inspired formulation of an SDN cell.

3 The SDN Cell

As introduced earlier in [3], SDM can be formulated in simpler terms than the original tensor algebra based formulation (i.e. in [2]) using a recurrent model referred to as the Spiral Discovery Network cell. Such a cell consists of:

- A **timer** that functions as a modulo counter for updating the state of the cell at discrete time steps
- A **perturbation module** that determines the direction in which, and the extent to which the slope of exploration is to be modified at each time step
- A **hypervisor module** that refreshes the hyperparameters of the perturbation module based on feedback signals

A graphical representation of an SDN cell and its modules is shown in Figure 1. The updated activation at time t is:

$$\mathbf{a}^{(t)} = (t * step\_sz + 1)\mathbf{x} + \mathbf{p}^{(t)} \tag{1}$$

where:

$$\begin{aligned} \mathbf{p}^{(t)} &= \mathbf{p}^{(t-1)} + sgn(cycle\_dir^{(t)}) \cdot \frac{\mathbf{p}^{max} - \mathbf{p}^{(0)}}{cycle\_len} \\ cycle\_dir^{(t)} &= \begin{cases} 1 & \text{, if } \frac{cycle\_len}{4} \leq t < \frac{3*cycle\_len}{4} \\ -1 & \text{otherwise} \end{cases} \end{aligned} \tag{2}$$

Generally speaking, the state of the SDN cell is updated in a series of timesteps which together constitute optimization cycles. In the update equations, $\mathbf{x}$ refers to the (normalized) principal component vector – the general direction in the parametric space that is being explored by the cell, while $\mathbf{p}$ refers to the perturbation vector that is added to the principal component. The relationship between the two is governed by the hyperparameter $step\_sz$. The contribution of $\mathbf{x}$ is incremented by $step\_sz$ at each timestep to ensure that the path of parametric discovery expands in the general direction of the principal component (hence, $step\_sz$ represents the degree of *exploitation* in the optimization process). The direction and norm of $\mathbf{p}^{(t)}$, by contrast, which ultimately depends on the relationship between $\mathbf{p}^{(0)}$ and $\mathbf{p}^{(max)}$, determines how far from the principal component the exploration will deviate (therefore, it is directly related to the concept of degree of *exploration* in the optimization process). $cycle\_dir$ governs the direction in which the perturbations are changed, and is dependent on the length of the cycle as well as the current phase within the cycle. The values of $\mathbf{p}^{(0)}$, $\mathbf{p}^{max}$ and $\mathbf{x}$ are dependent on the cycle (or more precisely, on the discoveries made during the previous cycle), and are updated as follows:

$$\begin{aligned} p^{(0)}_{i,unnormed}[c] &= p_i^{(\arg\min_t h_i^t[c-1])}[c-1] \\ p^{max}_{i,unnormed}[c] &= p_i^{(0)}[c] + softmax(\sigma_{h_i}[c-1])(\sigma_{h_i}[c-1]+1) = \\ &= p_i^{(0)}[c] + \frac{\exp \sigma_{h_i}[c-1]}{\sum_I \exp \sigma_{h_I}[c-1]}\,[\sigma_{h_i}[c-1+1] \\ \mathbf{p}^{(0)}[c] &= ||\mathbf{p}^{(0)}_{unnormed}[c]|| \\ \mathbf{p}^{max}[c] &= ||\mathbf{p}^{max}_{unnormed}[c]|| \\ x_i[c] &= ||x_i[c-1] + \frac{p^{(0)}_{i\ unnormed}[c]}{||\mathbf{p}^{(0)}[c]||}|| \end{aligned} \tag{3}$$

Here, the value of a parameter within a cycle c is represented using square brackets, so that for example $h_i^t[c-1]$ refers to the value of the i-th hypervisor cell at time t of cycle $c-1$. σ_{h_i} denotes the standard deviation of value the i-th hypervisor cell. The update equations ensure that:

- the perturbations in the new cycle are centered, in each dimension, around

the perturbation that was associated with the lowest cost function value in the previous cycle (note that h_i refers to the i-th hypervisor cell)

- the maximum values of the perturbations are set to their starting value, plus a value that depends on the standard deviation of the corresponding hypervisor cell in the previous cycle, as well as its relation to the standard deviations of other hypervisor cells. In general, the larger the deviation in a given dimension in an absolute sense, the greater the distance will be between the initial and maximal perturbation in the following cycle (in which case the network will be more explorative in that dimension). Similarly, exploration in dimensions that are characterized by large relative deviations will also be higher in the following cycle.

- the principal component, $\mathbf{x}$ is set to the initial principal component plus the normalized value of the initial perturbation.

4 Analysis of the SDN Cell

In order to get a general glimpse into the operation of an SDN cell, we consider its performance on a simulation example introduced earlier in [3] (see also Figure 2):

$$z = \begin{cases} 500 & \text{if } x, y \notin (0, 10] \times (0, 10] \\ 70 & \text{if } x, y \in [1, 1.5] \times [2.75, 4.5] \\ -10 & \text{if } x, y \in [3.25, 3.5] \times [3.5, 4.25] \\ (x-5)^2 + \dots & \\ -2(y-2) + \dots & \text{otherwise} \\ x + e^{\frac{1}{(x+y)}} & \end{cases} \tag{4}$$

Figure 3 shows the points within the two-dimensional parameter space that are visited by the cell in each cycle for the first 15 cycles. Further, the plot for each cycle shows the principal component vector, initial perturbation vector and maximal perturbation vector computed for the following cycle. Based on the figure, the following conclusions can be drawn:

- Generally speaking, the initial perturbation vector (blue) for the next cycle will point towards the direction of the perturbation that was applied when the point with the smallest loss value was generated (largest point on the plot);

- The maximal perturbation vector (red) is obtained by adding to the initial perturbation vector (blue) a vector that characterizes, in each dimension, the variance of the feedback, both in absolute terms and in relative terms (i.e., as compared with other dimensions). Note that in the simulation, the output of the loss function is a scalar value and there is no credit assignment among

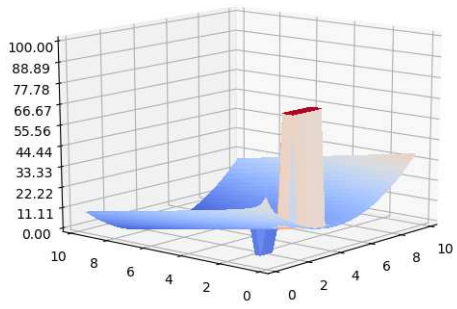


Figure 2
Plot of the loss function to be minimized in the simulation example.

the dimensions of the parameter space (the SDN model described in equation contains no such credit assignment), therefore both x and y components of the initial and maximal perturbation vectors are the same. This needs to be solved in future versions of SDN (including the hierarchical model proposed in this paper).

- As a result of the previous observation, the angle between the initial (blue) and maximal (red) perturbation vector largely depends on the orientation of the former (their difference is always a vector that points in the direction $[1, 1]$). This in turn influences the angle between $\mathbf{p}^{(0)} - \mathbf{x}$, and $\mathbf{p}^{max} - \mathbf{x}$, which is what determines the path of exploration.

5 Modified SDN Cell Model with Hypervisor Credit Assignment

Based on the above, a modified version of the SDN cell is proposed with hypervisor credit assignment. The approach introduces a feedback path from the output of the cell to the hypervisor cell, through the mediation of a time delay component called the **delta module** (Figure 4). The role of this module is to model the influence of changes in each of the output dimensions on the value of the loss function. Given the fact that the SDN model makes no assumptions on the loss function, it cannot be assumed that this influence can be modeled effectively over long periods of time. Therefore, changes through longer periods of time are discounted based on the following update equation:

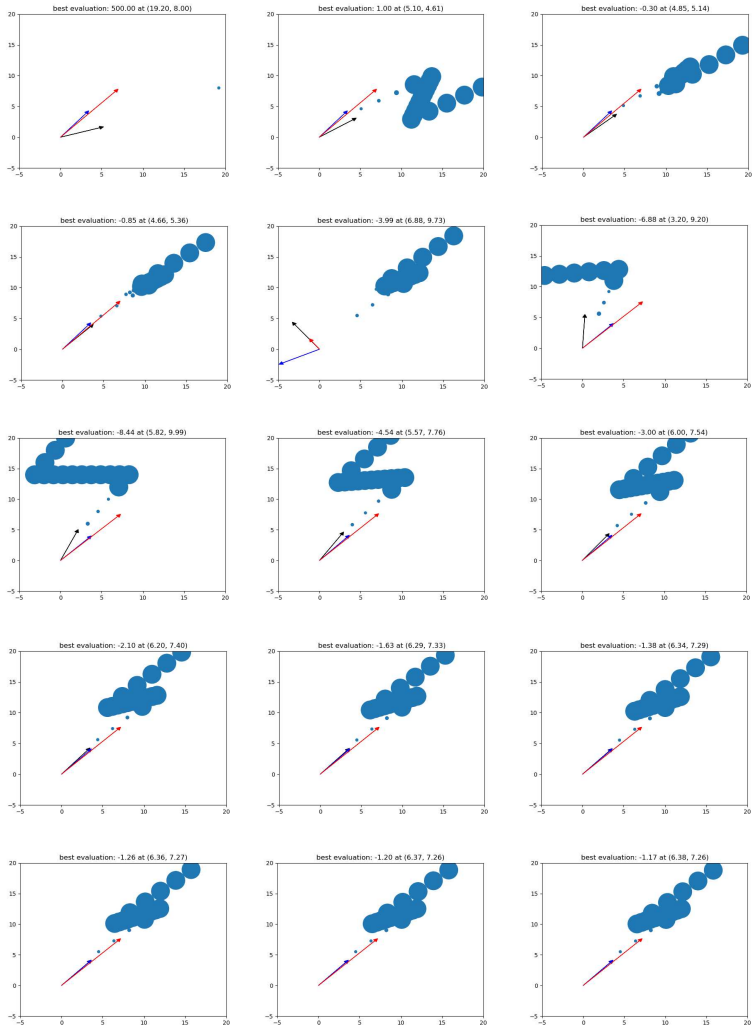


Figure 3

Plots of simulation in the first 15 cycles (left to right, top to bottom). The plot for each cycle shows the points within the parameter space that were visited in the cycle (with the sizes of the points being inversely proportional to the value of the loss function), as well as the unnormalized principal component vector (black), unnormalized initial perturbation vector (blue) and unnormalized maximal perturbation vector (red) computed for the next cycle (the length of all three vectors is scaled up for visibility).

$$h_i[t] = loss * \gamma_{h_i}[t] = \\ = loss * softmax\left(\sum_{i=1}^{\infty} i^{-1}(a_i[t] - a_i[t-i])^2\right) \tag{5}$$

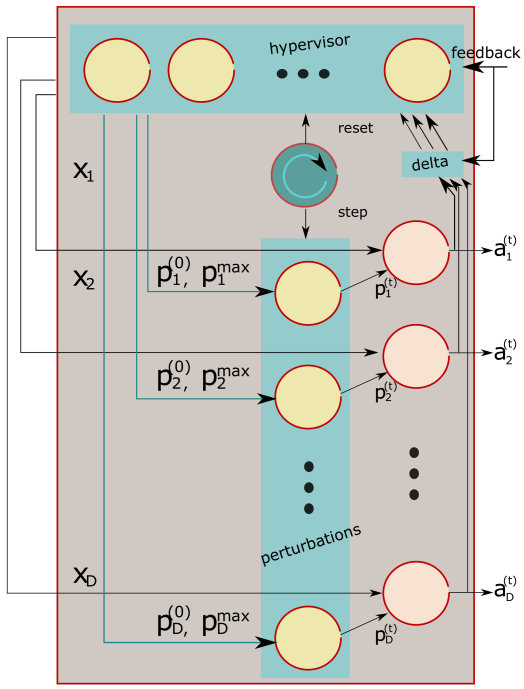


Figure 4
Neural network inspired formulation of an SDN cell with hypervisor credit assignment.

where γ is a scaling factor applied to the loss function value that is calculated through a softmax function and depends on all dimensions of the output activity.

Figure 5 shows the same simulation as earlier, this time carried out using the modified version of the SDN cell with hypervisor credit assignment. It is clear – as it should be – that it is no longer the case that the difference between the maximal and initial perturbation vectors is always a vector pointing in the direction $[1, 1]$ – hence, this model is better at achieving higher degrees of exploration when useful.

However, upon closer inspection it can be noticed that the evolution of the model seems to involve cycles and does not further adapt after a certain period of time. For example, the path of the outputs and the directions of the principal and perturbation vectors are very similar in the 15th cycle (fifth row, last plot in Figure 5) and in the 5th cycle (second row, middle plot in Figure 5). If the model is run longer, indeed it becomes clear that after a time, the same paths of exploration are revisited. Thus, the idea of adding further cells to the model – i.e., turning the model into a hierarchical network of SDN cells – seems to be worth exploring.

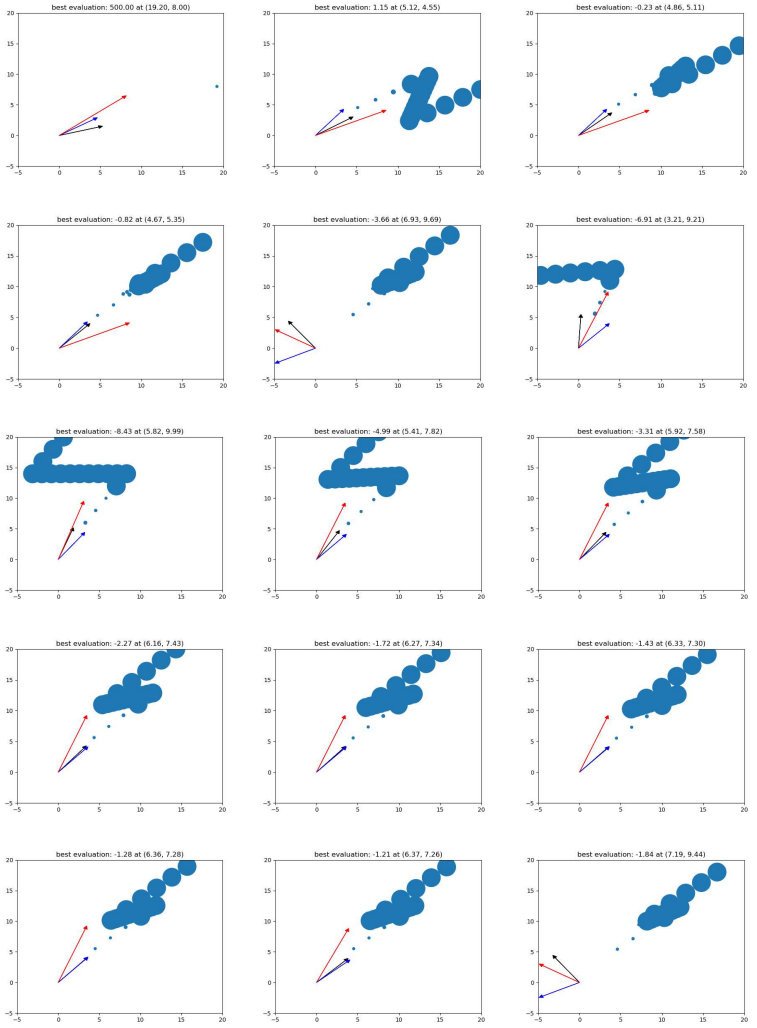


Figure 5

Plots of the second simulation including hypervisor credit assignment in the first 15 cycles (left to right, top to bottom). The plot for each cycle shows the points within the parameter space that were visited in the cycle (with the sizes of the points being inversely proportional to the value of the loss function), as well as the unnormalized principal component vector (black), unnormalized initial perturbation vector (blue) and unnormalized maximal perturbation vector (red) computed for the next cycle (the length of all three vectors is scaled up for visibility).

6 The Hierarchical SDN Model

Based on the above, a hierarchical SDN model can be proposed in which a top-level SDN cell computes the $step\_sz$ and principal component of the output layer

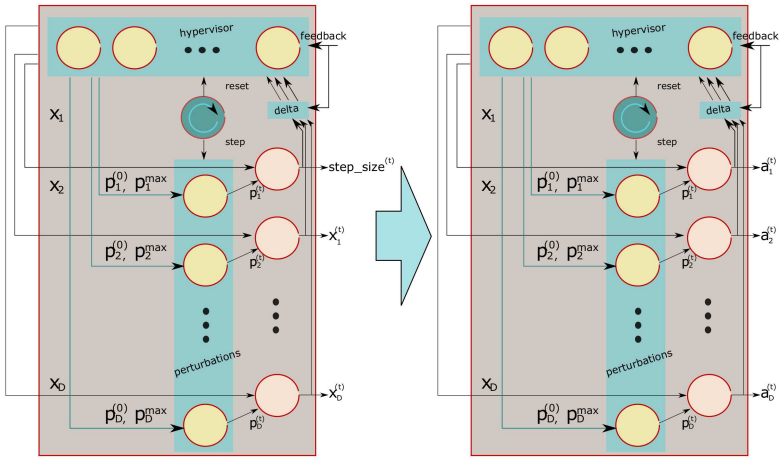


Figure 6
Hierarchical SDN comprised of two SDN cells.

(Figure 6). Following the *temporal slowness principle*, the lower-level SDN cell has a shorter period and produces a relatively large variety of outputs per input received from the top-level cell. Therefore, each output configuration of the top-level cell is associated with a single aggregated feedback from a multitude of outputs provided by the lower-level cell.

Accordingly, the third simulation presented in this paper was comprised of two SDN cells, both with a cycle of 20 steps and a period ratio of 20 lower level cycles per each cycle at the higher level. The smallest loss function value obtained over the 20 low-level cycles were fed back to the higher-level SDN cell and used as feedback for the configuration that yielded the corresponding $step\_sz$ and principal component vector ($\mathbf{x}$). Key results from the simulation are shown on Figure 7. The figure shows that the global minimum was found as early as in the 3rd high-level cycle. This is a qualitative improvement over all previous simulations, in which the search became stuck in cycles that didn't include the global minimum. A further observation that seems to underline the superiority of the hierarchical model in its modeling capabilities is the observation that no actual feedback was necessary from the lower-level layer – i.e. that it was sufficient to try 3 different higher-level output configurations to find the global minimum. This may of course also reflect the simplicity of the simulation problem.

Conclusions Evolutionary and neural network based parametric optimization approaches are different approaches, each with their own set of assumptions and limi-

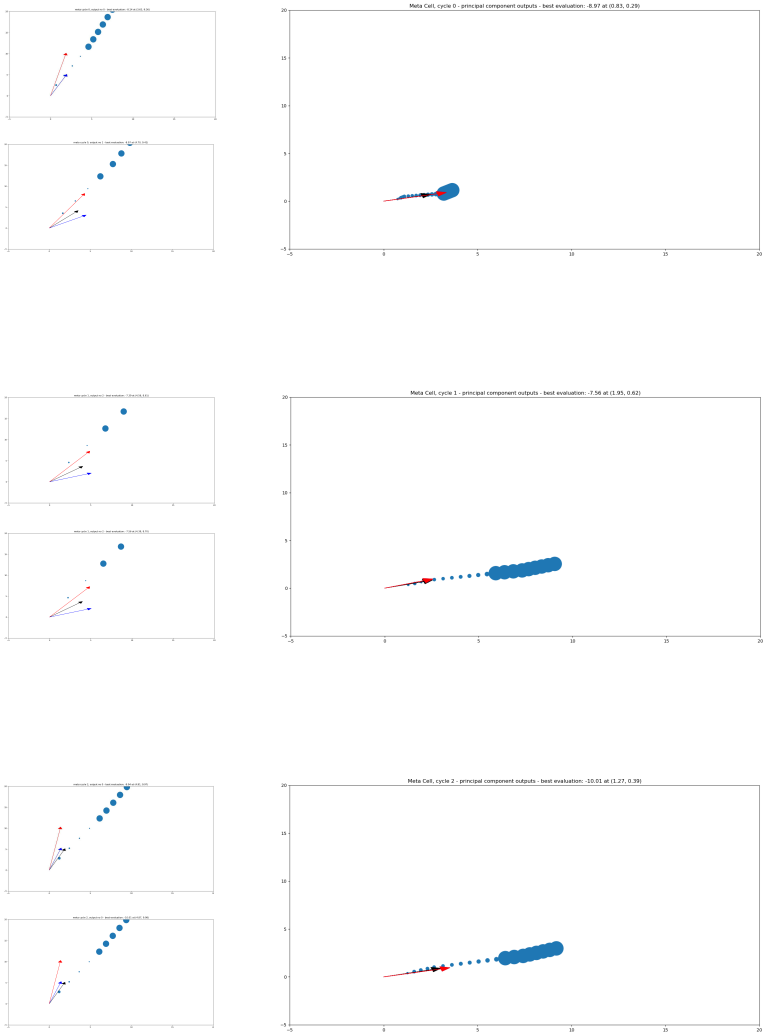


Figure 7
Plots of the third simulation – in this case using the hierarchical SDN model including cells with hypervisor credit assignment. The figure shows the top two results (left column) associated with the first three high-level (meta) cycles (right hand column). Because each meta cycle provides 3-dimensional outputs ($step\_sz$ and a two-dimensional principal component vector), only the elements of the principal component vector are shown in the right-hand column.

tations. Problem domains exist where neither approach is effective. First, it may be difficult to encode candidate solutions as evolutionary genotypes that behave well

(or provide useful results) under the operations of mutation and recombination. Second, the performance of candidate solutions sometimes cannot be evaluated using a globally defined and differentiable loss function. In such cases, the Spiral Discovery Network model and its extensions proposed in this paper can be used as an alternative approach: one that works even with feedback evaluations obtained less frequently, and without any explicit gradient information. The SDN approach can be characterized as follows: Instead of relying on evolutionary operations such as recombination and mutation, it operates directly within the search space in an auto-regressive fashion, and so the requirement of creating useful, hand-coded input encodings is alleviated; To compensate for the lack of a generally computable and differentiable loss function, it explores the search space along a parametric hyperspiral structure that implicitly generates differential feedback information, allowing the model to adapt its behavior in subsequent cycles of exploration.

References

[1] Peter Baranyi, Adam Csapo, and Gyula Sallai. *Cognitive Infocommunications (CogInfoCom)*. Springer International Publishing, 2015.

[2] A. Csapo and P. Baranyi. The Spiral Discovery Method: an Interpretable Tuning Model for CogInfoCom Channels. *Journal of Advanced Computational Intelligence and Intelligent Informatics*, 16(2):358–367, 2012.

[3] Adam B Csapo. The spiral discovery network as an automated general-purpose optimization tool. *Complexity*, 2018:1–8, 2018.

[4] Adam B Csapo. The spiral discovery network as an evolutionary model for gradient-free non-convex optimization. In *IEEE International Conference on Cognitive Infocommunications*, pages 1–6, 2018.

[5] Pedro Domingos. *The master algorithm: How the quest for the ultimate learning machine will remake our world*. Basic Books, 2015.

[6] John Duchi, Elad Hazan, and Yoram Singer. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research*, 12(Jul):2121–2159, 2011.

[7] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International Conference on Machine Learning*, pages 448–456, 2015.

[8] Diederik Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[9] Marvin Minsky. Steps toward artificial intelligence. *Proceedings of the IRE*, 49(1):8–30, 1961.

[10] Arvind Neelakantan, Luke Vilnis, Quoc V Le, Ilya Sutskever, Lukasz Kaiser, Karol Kurach, and James Martens. Adding gradient noise improves learning for very deep networks. *arXiv preprint arXiv:1511.06807*, 2015.

[11] Yurii Nesterov. *Introductory lectures on convex optimization: A basic course*, volume 87. Springer Science & Business Media, 2013.

[12] Ning Qian. On the momentum term in gradient descent learning algorithms. *Neural networks*, 12(1):145–151, 1999.

[13] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning internal representations by error propagation. Technical report, California Univ San Diego La Jolla Inst for Cognitive Science, 1985.

[14] Jürgen Schmidhuber. Deep learning in neural networks: An overview. *Neural networks*, 61:85–117, 2015.

[15] Ilya Sutskever, James Martens, George Dahl, and Geoffrey Hinton. On the importance of initialization and momentum in deep learning. In *International conference on machine learning*, pages 1139–1147, 2013.

[16] H. Takagi. Interactive evolutionary computation: Fusion of the capabilities of EC optimization and human evaluation. *Proceedings of the IEEE*, 89(9):1275–1296, 2001.

[17] Hideyuki Takagi and Hitoshi Iba. Interactive evolutionary computation. *New Generation Computing*, 23(2):113–114, 2005.

Connecting Bitcoin Blockchain with Digital Learning Chain Structure in Education

Krisztián Bálint[1,2], Dragan Cvetković[1], Márta Takács[3], Ildikó Holik[4], Alex Tóth[5]

[1]Singidunum University, Danijelova 32, 11000, Belgrade, Serbia

[2]Óbuda University, Doctoral School on Safety and Security Sciences, Népszínház u. 8, 1081 Budapest, Hungary

[3]Óbuda University, John von Neumann Faculty of Informatics, Bécsi út 96/b, 1034 Budapest, Hungary

[4]Óbuda University, Ágoston Trefort Centre for Engineering Education, Népszínház u. 8, 1081 Budapest, Hungary

[5]Municipal Museum of Subotica, Trg Sinagoge 3, 24000 Subotica, Serbia

E-mails: kristian.balint.12@singimail.rs, dcvetkovic@singidunum.ac.rs takacs.marta@nik.uni-obuda.hu, holik.ildiko@tmpk.uni-obuda.hu, muzejsubotica@mts.rs

Abstract: In the present paper, an extension of the Bitcoin (BTC) structure is introduced, with a proprietary and unique DLCC (Digital Learning Chain Structure) blockchain, through which an automated bursary payment system could be constituted, which has, until this day, never been implemented at universities. Thus, an alternative is offered to achieve a bursary payment system, operating in a fully automatic, periodically repeating manner, never requiring supervision. In the DLCC blockchain, specifically, in the Smart Contract, the terms of contract are kept, concerning the bursaries, terms agreed by both parties (students and universities alike). A great advantage of the DLCC system is that it is able to extract the required grades from the electronic grade book, thus supervising the students' progress throughout the semesters. Upon reaching the end of each semester, if the student qualifies, meeting the requirements, the system will transfer the required quantity of cryptocurrency from the university's account automatically to the student's account.

Keywords: Bitcoin; Smart Contract; Digital Learning Chain Structure

1 Introduction

Bitcoin is an open source digital currency, introduced in 2009 by an unknown user. The term is also used for the open source software managing the currency, as well as for the created distributed network. Bitcoin does not depend on central issuing agencies or authorities since it is supported on the distributed database stored at peer-to-peer network nodes. Bitcoins can be safely stored in a wallet, on a computer or mobile device, as well as in a cloud-based provider. Since the system cannot be controlled from the outside, the essence of the transaction remains hidden, which is why it is criticized for its potential illegal usage. At the same time, there are more and more applications using solutions based on Bitcoin blockchains.

Blockchain technologies make tracking and managing digital identities, both secure and efficient, resulting in seamless sign-on and reduced fraud. Be it banking, healthcare, national security, citizenship documentation or online retailing, identity authentication and authorization is a process intricately woven into commerce and culture worldwide. Blockchain technology can be applied to identity applications in the following areas: Digital Identities, Passports, E-Residency, Birth Certificates, Wedding Certificates, and Ids [1].

Blockchain technology can also be used in transportation. It provides support for tracking the goods with a blockchain-based solution. This way it is always possible to pinpoint the exact location of the goods anywhere in the world, en route between the manufacturer and user. [2].

Medical records are notoriously scattered and erroneous, with inconsistent data handling processes, i.e. hospitals and clinics, are often forced to work with incorrect or incomplete patient records. Healthcare projects such as MedRec are using the blockchain as a means of facilitating data sharing while providing authentication and maintaining confidentiality [3].

Over time, blockchain technology has found its way into various industry segments, too, including peer-to-peer taxi rides and other similar services. Arcade City is a blockchain-based ride sharing and car-for-hire application. Naturally, payment is organized via crypto currency [4].

Dubai has set sights on becoming the world's first blockchain-powered state. In 2016, representatives of 30 government departments formed a committee dedicated to investigating opportunities across health records, shipping, business registration and preventing the spread of conflict diamonds [5].

Traditional systems tend to be cumbersome, error-prone and rather slow. Intermediaries are often needed to mediate the process and resolve conflicts. As expected, this causes stress, and costs time, and money. In contrast, users find the blockchain cheaper, more transparent, and more effective. Unsurprisingly, an increasing number of financial services use this system to introduce innovations,

such as smart bonds and smart contracts. The former automatically pays bondholders their coupons once certain preprogrammed terms are met. The latter are digital contracts that self-execute and self-maintain, again when terms are met [6].

The New York-based Bitcoin exchange is working on Project Highline, a method of using the blockchain to settle and clear financial transactions in T+ 10 minutes rather than the customary T + 3 or T + 2 days [7].

Blockchain technology can be used in everyday life for gift cards and customer loyalty programs. The aim is to increase security, given that it stops the intermediaries by using unique BTC blockchain-based controls. By excluding the intermediaries, the gift cards make their way directly from the vendors to the customers without the customers ever having to give out personal information [8].

The DLCC (Digital Learning Chain Structure) system could also provide great help in education, as it could provide simplicity and transparency to a given system. The structure of Bitcoin has specific characteristics, whose possibilities have not been fully utilized in education so far, whereas, the system based on bursary payment could be realized with its help.

The aim of the research presented in this paper is to create a bursary payment system in higher education system, based on Bitcoin's structure, which could be utilized successfully by the institutions of higher education in the near future. The DLCC structure includes the following important elements, including:

- The generation of students' Smart Contracts based on BTC;
- The interconnection of electronic school registry (grade book) with the smart contracts;
- The examination of bursary terms by concentration of these systems;
- Online accessibility.

The development of the DLCC structure is complete, functioning and has been tested in practice.

Both the university, as well as the students must be open to introducing this system in practice. To this end the authors conducted an empirical research with the participation of 187 university students from Hungary and 102 students from Romania, so as to discover the students' opinions on Bitcoin, and whether they would be ready to accept a bursary system based on Bitcoin.

The paper first describes Bitcoin's basic elements, followed by the detailed description of the Digital Learning Chain Structure system in place. As mentioned briefly above, a case study is added to the research results, where the views of students from Hungary and Romania concerning Bitcoin are uncovered.

2 Bitcoin Basics

Bitcoin is one of many types of virtual currencies based on an algorithm that creates a direct peer-to-peer transaction system [9] using the necessary concepts for that.

The Bitcoin protocol implements an address propagation mechanism to help peers discover other peers in the P2P network. Each Bitcoin peer maintains a list of addresses of other peers in the network and each address is given a timestamp which determines its freshness. Peers can request addresses from this list from each other using GETADDR messages and unsolicitedly advertise addresses known to them using ADDR messages. Whenever a Bitcoin node receives an ADDR message, it decides individually for each address in the message whether or not to forward it to its neighbours. It first checks if:

- the total number of addresses in the corresponding ADDR message does not exceed 10, and
- the attached timestamp is not older than 10 minutes.

If either of these two checks fails, the address is not forwarded; otherwise the address is scheduled for forwarding to two of the node's neighbours in case the address is reachable and to one neighbour only if it is non-reachable. An address is considered reachable by a node if the node has a network interface associated with the same address family. Otherwise, the address is marked as unreachable [10].

The Bitcoin blockchain consists of a hashchain of blocks: every block contains an ordered set of transactions and a hash of the preceding block (starting from the initial, the so-called "genesis" block). The key part is the Proof-of-Work (PoW) aspect of the hashchain: a Bitcoin block contains nonces that a Bitcoin miner (i.e., a node attempting to add a block to the chain) must set in such a way that the hash of the entire block is smaller than a known target, which is typically a very small number. In the early days of Bitcoin, the performance scalability of its probabilistic PoW-based blockchain was not a major issue. Even today, Bitcoin works with a consensus latency of about an hour (for the recommended 6-block transaction confirmation), and with up to seven transactions per second peak throughput (with smallest 200-250 byte transactions). On top of this, the Bitcoin network uses a lot of power, which, in 2014, was roughly estimated to be in the ballpark of 0.1-10 GW2.1.4 [11].

Users transfer coins (BTCs) to each other by issuing a transaction. A transaction is formed by digitally signing a hash of the transaction through which a BTC was acquired. Given that in Bitcoin there is one-to-one correspondence between signature public keys and addresses, a transaction taking place between two addresses aS and aR has the following form: $\tau(aS \rightarrow aR) = \{source, B, aR, SIGskaS (source, B, aR)\}$. Here, SIGskaS is the signature using the private key skaS that corresponds to the public key associated with the aS, B is the amount of

BTCs transferred, and source is a reference to the most recent transaction that aS acquired the B BTCs from. After their creation, Bitcoin transactions are released in the Bitcoin network. Once the validity of τ is confirmed, aR can subsequently use this transaction as a reference to spend the acquired BTCs. Consequently, Bitcoin transactions form a public record and any user can verify the authenticity of a BTC by checking the chain of signatures of the transactions in which the BTC was involved [12].

3 Solution for Communication between Bitcoin and DLCC Blockchains

Although it is true that Bitcoin has an open source system based on mathematical algorithms, on the other hand, it is almost an impossible and unaccomplishable task to make changes to the BTC blockchain. This is due to the fact that the transactions are continuously built on each other from the beginning of the system. Therefore, if a change in the blockchain had to be made, it would be necessary to retrace it to the first ever transaction and that is a serious task. Hence, the changes would be most expedient if made to another blockchain, which, in turn, would be connected to the existing and stable Bitcoin blockchain. The creation of a DLCC blockchain would provide a new solution, which could be harnessed in the educational system. As a first step, this new blockchain would allow the signing of Smart Contracts. So far, the Smart Contracts have been exclusive elements of the Ethereum system. Created in 2015, they started to be widely used in 2016. The Smart Contract system has seen steady adoption, supporting tens of thousands of contracts, holding millions of dollars' worth of virtual coins [13]. Therefore, the base of DLCC would consist of Smart Contracts. Based on the experiences, the following advantages and drawbacks of Smart Contracts can be identified:

- Fair exchange between mutually distrustful parties with rich contract rules expressible in a programmable logic; this feature prevents parties from cheating by aborting an exchange protocol, yet removes the need for physical meetings and (potentially cheating) third-party intermediaries;
- Minimized interaction between parties also for a rich set of contracts expressible in a programmable logic, reducing opportunities for unwanted monitoring and tracking;
- Enriched transactions with external state by allowing input authenticated data feeds (attestations) provided by brokers on physical and other events outside the Smart Contract system, e.g., stock tickers, weather reports, etc.

Unfortunately, despite all their benefits, these properties also have a dark side, potentially facilitating crime because:

- Fair exchange enables transactions between mutually distrustful criminal parties, eliminating the need for today's fragile reputation systems and/or potentially duplicitous or law-enforcement-infiltrated third-party intermediaries;
- Minimized interaction renders illegal activities harder for law enforcement to monitor [14].

Considering the recent experiences shared in the trade periodicals, which discuss the Smart Contracts, it would be possible to eliminate numerous drawbacks of DLCC, starting from the initial phase. Considering the nature of faculty systems, it is to be assumed that they are not invented and created for conducting criminal activities, but rather for raising the standards of educational institutions', as well as the students' level of satisfaction. For the accomplishment of the mentioned project, the School Boards must strive to wholeheartedly provide all competency, and maximum correctness.

3.1 Soft-Fork and Hard-Fork in Bitcoin Terminology

Connecting the Bitcoins blockchain with the DLCC blockchain is a challenging task. The accomplishment of communication between the two blockchains requires the involvement of Bitcoins protocol. There are two possibilities for this: the Soft-Fork and the Hard-Fork.

In Bitcoin, the terms Soft-Fork and Hard-Fork describe compatibility breaking changes in the Bitcoin protocol: should the community be irreconcilably divided on such an issue, the old version and the new version of Bitcoin could emerge as distinct projects thereafter. While both versions of the Bitcoin protocol are in use, the differences in acceptance may cause a lasting blockchain-fork, i.e. the two distinctly longest chains which are both considered valid by part of the network.

- Soft-Forks are forward compatible,
- Hard-Forks are not forward compatible [15].

A Hard-Fork occurs when blocks that would have previously been considered invalid, are now valid. Any Bitcoin user, miner, exchanger, etc. with the intention to stay in consensus with the network must upgrade their software during a Hard-Fork; otherwise, some new block that the network accepts will appear as invalid to it.

A Soft-Fork occurs when blocks that would have previously been considered valid, are now invalid. Upgrading software during a consensus Soft-Fork is always optional for any Bitcoin user, miner or exchanger, with the following caveats:

- If the Soft-Fork introduces a new feature that the user wishes to implement as either the sender or recipient, they must upgrade in order to use it.
- At least 51% of miners must upgrade to adopt the Soft-Fork, otherwise, it will always appear as the shortest chain and become orphaned by the network.
- Refusal to accept the Soft-Fork can reduce one's security. Given that Soft-Forked transactions are normally considered invalid, Bitcoin developers use various tricks to make these transactions appear valid, while also reducing the client's capacity to process exactly why they are valid [16]. The first figure presents the operation of the Soft-Fork blocks (Figure 1).

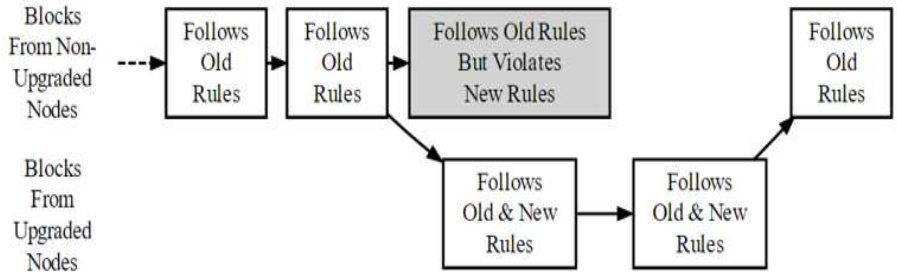


Figure 1

A Soft-Fork: Blocks violating new rules are made stale by the upgraded mining majority

For the implementation of the protocol between the BTC blockchain and DLCC blockchain, in this case, the best solution would be the Soft-Fork. The Figure 2 shows how to connect the DLCC blockchain with the Bitcoin blockchain.

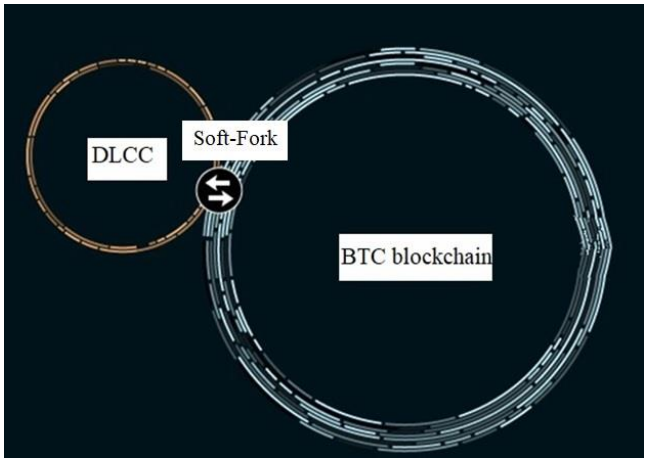


Figure 2
Link between DLCC's and Bitcoin's blockchains

The process of creating the DLCC blockchain can be seen in the Table 1.

Table1
Blockchain creating process

Mychain-util generate DLCC the default settings would be used: */default ~ mychain/DLCC/params.dat* *params.dat include:* *Contract addresses [receiver (student) IP address, sender (university) IP address],* *BTC wallet addresses [receiver (student) IP address, sender (university) IP address],* *Terms of contract.* Next the DLCC blockchain would be initialized, and the genesis block would be mined *mychain DLCC* The server will be started in those few seconds after the genesis block has been found, then the node address needs to be connected: *DLCC@192.168.0.1:8008*

After these steps, the connection can be attempted from a second server:

mychain DLCC@192.168.0.1:8008

After the message confirming the chain has been initialized, the permission is not given for connection to the wallet. The address would be copied and pasted: 192.168.0.2

finally, permission for connection would be granted:

mychain DLCC grant 192.168.0.2 connect

At this point, an attempt can be made to reconnect to the second server, by a multi-chained DLCC – daemon. After the setup of the blockchain, the interactive mode would be used, where the setup would commence, so permissions addresses, peers, parameters of blockchain, native parameters, transaction meta-data, and streams would be given and mining would commence [17].

4 Operating Principles of DLCC

The application of networked computer systems has brought a great change in higher education in the form of electronic teaching materials, communication using web pages, and course administration by purposeful software systems [18].

The next step, utilizing these purposeful software systems, the upper management of the universities ought to perform management tasks, not yet used until now, in an automated way, with the help of DLCC. This is a novel possibility of performing these tasks, unavailable until several years ago. In order to create a successfully operating DLCC project, the adequate competence must be ensured not only from the university's side, but also from the students' side, as the digital literacy represents a mixture of consciousness, trending and capabilities, which enables the sufficient and safe application of digital means and institutions, with the aim of identifying, reaching, using, integrating, rating, and synthesizing of the digital sources, as well as the production of new knowledge and media publications, also with the goal of communication and mutual reflection regarding this process amongst ourselves [19].

The first step in the implementation of DLCC is signing the contract between the faculty and the student. This is to happen immediately after the students' enrollment to their first year of study. It would be highly useful in the case of enrollment for bursary. By signing the given contract, a possibility would arise for DLCC to automatically follow the students' results during their study years, with the help of marks in the electronic grade book. Thus, if a student meets the required terms, the bursary would be automatically be paid at the end of respective

semesters. This solution would take off an enormous burden of faculties' administration. According to the contract, Krisztián is the university student, while Professor Dragan is the university representative, with whom Krisztián signs the bursary contract. The structure of DLCC Smart Contract consists of the elements presented in the Figure 3.

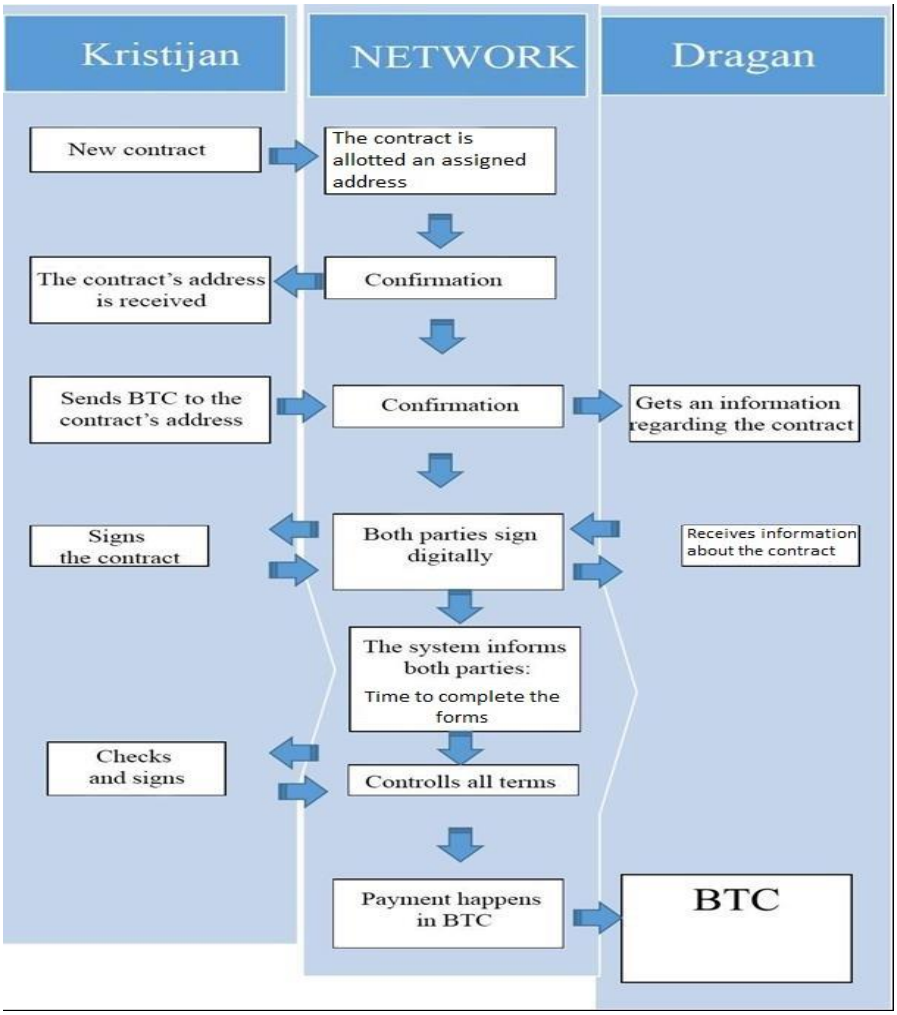


Figure 3
DLCC Bitcoin contract

Looking at this from another perspective, the mentioned innovation would be advantageous, because by the utilization of such systems, it would be possible to raise the students' financial proficiency in everyday life. This way, the students could be establishing a direct connection with modern financial systems even

during their years of education. Such a solution is all the more thought-provoking given the fact that all people have finances, which need personal attention. Some people are prone to quick decision making, while others are prefer to obtain substantial quantity of information before each transaction, whereas there are also people who like to rely on their intuition. Therefore, aside from skills in economy, financial proficiency is considerably determined by societal and social relations, monetary attitudes and traits of personality. International research has proven the existence of significant differences among youth, considering the acquisition, interpretation and implementation of financial knowledge [20]. Information aggregation is one of the key issues in development of intelligent systems [21]. The Smart Contract automatically loads information needed for its operation from many sources, requiring no more than a safe access to the Internet. Necessary information includes, for instance, the daily Bitcoin exchange rate, the date and time, or the records and marks from the electronic grade book. As DLCC is based on Bitcoin's structure, thus, its secrecy is outstanding, as BTC uses military grade encryption, while the school based computer systems do not, and the Smart Contract is formulated on these. Security threats are closely related to the overall level of software security in computer systems [22]. Therefore, the school-based computer systems need to be upgraded, concerning the adequate level of protection.

The structure of DLCC would consists of the elements described in the following sections.

4.1 The Definition of Terms of Contract

A great advantage of Smart Contracts is that there is no possibility of later modification or manipulation. For this reason, it is important to ensure that the contracts are constructed with due diligence. In these contracts, concerning the applications for bursary, the students would accept the obligation, and state that they will fulfill the necessary terms, while the faculty would warrant regular and timely payment. In addition, the regularity of payments would be defined with exact dates, as the system would execute these payments automatically. For example, it would be ensured that payments would be executed on the first day of every month, or on the last day of every semester. Such payments have a great advantage, as they are not affected by holidays, weekends, because Bitcoin's network functions ceaselessly every day of the year.

4.2 Electronic Grade Book

It is important to secure, even on the institution level, a unified registry of curriculum, and a connecting informational system for leaders, which is necessary for the functioning of educational institutions, so they handle the data emerging from their functioning in a consistent way [23]. The electronic registry fulfills

these terms completely, as the electronic rating (awarding of marks) is completed in this system, and similarly, the later recalling of marks by DLCC, according to the terms of contract.

4.3 Gathering of Information

Firstly, DLCC collects the necessary information from the electronic diaries, in a way that it continuously extracts the marks of a student, and at the end of each semester, the total marks of the subjects. Secondly, the system follows the corresponding dates, with the aim of keeping up with the dates of mentioned marks. DLCC can also read the actual date from the electronic grade book, if it is available, and if unavailable, the system can perform this task, with the help of access to the Internet.

Apart from collecting information, the constant following of Bitcoin's exchange rate is also very important, as the price of Bitcoin is rather dynamic, as this kind of currency has gone through many years of turbulent evolution in past years [24]. The high fluctuation of the exchange rate is characteristic it [25], however, the exchange rate of Bitcoin seems to undergo stabilization [26]. For this reason, it is vital for DLCC to use the Internet so that it continuously traces the fluctuation of exchange rate, as it can greatly influence the amount of payments.

4.4 Examination of Terms in Contracts

DLCC categorizes and evaluates the collected information. It examines the fulfillment of terms in contracts. If the terms are met, the next step will take effect: the payment. If the student fails to meet the requirements from the contract, then DLCC does not allow the completion of payment.

4.5 Payment

Lastly, the DLCC system pays the student; it transfers the necessary money to the corresponding electronic wallet. To this end, the student needs to possess a digital wallet. The users keep their wallets on their own machines, but in order to minimizing the risk of theft, they can demand the service of online wallets. Every wallet is based on a pair of keys – a public key and a confidential key, these two fulfill different tasks. The public key creates the address: it is basically a string of letters and numbers ranging from 27 to 34 characters. The secret key is used to validate the transactions. The addresses do not contain data about the user, but by signing of the public keys, the transaction can be retrieved. Although the transactions in Bitcoin can be retrieved based on the addresses, the owners of individual addresses remain unknown [27].

The construction of the DLCC structure is presented in the fourth figure in the series. As can be seen, the first step is the definition of the contract conditions,

then the system can extract the data from the electronic log, compile them, and finally analyze them as to whether or not the necessary requirements for payment were met. As a final step, the bursary is transferred to the student. It is vital that the university has an up-to-date firewall, in this way it is possible to prevent any loss of grades in the electronic log.

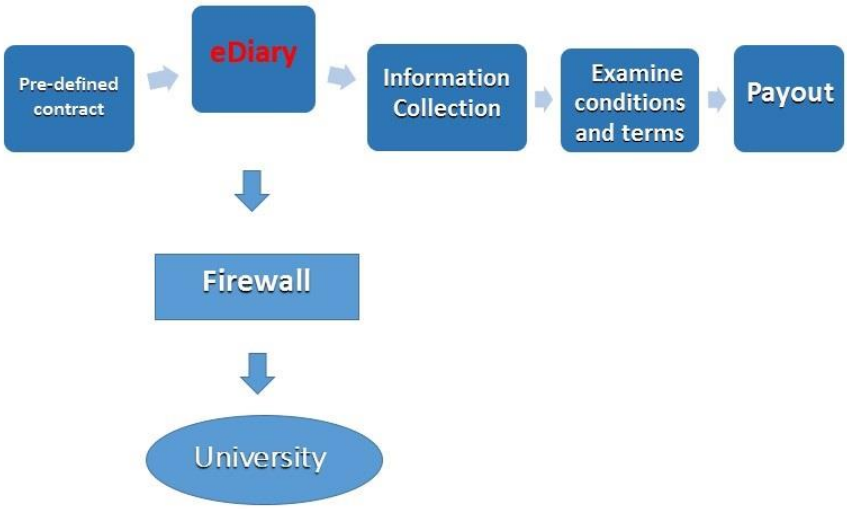


Figure 4
Main elements of the DLCC structure's operation

5 Case Study about the Possible Introduction of the System for Students

Before the commencement of the empiric investigation, it was an imperative to examine the economic situation of digital currency regarding countries where the research subjects (students) live, as the standpoints of some countries are were significantly divided regarding the topic of cryptocurrency, thereby influencing the test subjects' answers.

5.1 Status of Hungary in the World of Cryptocurrency

In Hungary, the following legislative decrees define the notion of electronic money:

The notions of electronic money and electronic means of payment have been defined by Law No. CXII. Since 1996, until its amendment on October 1, 2009, stating the following:

"5.2 Electronic money: an amount (worth) of money, stored in electronic currency, issued in turn of receiving hard cash or a sum from a balance, which is accepted by other parties than the issuer, in order of electronic payment."

"5.3 Electronic means of payment: means of payment represents a substitute for hard cash – as such, particularly a card for storing value, or computer memory – which is used to store electronic money, and with which the client is able to directly perform operations of payment " [28].

To date (2017), the National Bank of Hungary (MNB) has not formally recognized Bitcoin as a currency. According to the National Bank of Hungary, some of the means suitable for payment (specially Bitcoin) are by far more risky, than the electronic payment solutions already well-known to users (e.g. the credit cards, electronic money, etc.), as they do not have a formal issuer, they do not fall under supervision of any state-owned authorities or national banks, and there are no rules concerning the accountability, warranty, or damage management in existence. The National Bank of Hungary calls upon consumers' attention to be very aware, to act with utmost caution, before the use of such means of payment, for example Bitcoin [29].

5.2 Status of Romania in the World of Cryptocurrency

Up to 2017, the National Bank of Romania (BNR) has failed to pass any the normative regulation regarding Bitcoin, in other words, it does not prosecute, nor does it support the use of cryptocurrency. Romania's economy is surprisingly open towards the cryptocurrency in comparison to the neighboring countries, as one can obtain cryptocurrency in exchange for domestic currency in numerous places in Romania. There are 3 Bitcoin ATM machines working in Bucharest, Romania, since 2014. Within a one-way ATM transaction, it is only possible to exchange Romanian Lei for Bitcoins.

Moreover, thanks to the mutual cooperation of the operator of terminals, named Zebra Pay and Bitcoin Romania there is a possibility of more than 800 terminals and 160 cities in Romania to buy Bitcoins in exchange for domestic money. The transactions through the exchange agency named Coin trader carry a 4% transfer fee. With this new, simplified process of transaction, though, the transaction requires much less time [30]. However, regrettably, BTCX change, one of the greatest exchange offices of Romania concerning cryptocurrency, suspended its services for an indefinite period in early December, 2014. There is a background story about a misunderstanding between the lead programmer and the manager of the website. According to the programmer's statement, he had not received his salary; hence he closed down the servers, making them effectively inaccessible. These news caused significant surprise within the BTC community, as during almost 8 months of the website's operation, there were BTC transactions worth of 2 million dollars completed [31].

5.3 Results of Empirical Research

The empirical study was conducted with the aim to find out the attitude of Romanian and Hungarian students towards Bitcoin and the possible bursary system based on Bitcoin blockchain technology.

Table 2
Number of students participating in the present research, per states

States	Faculties	Number of students
Hungary	Óbuda University – Kandó Kálmán Faculty of Electrical Engineering – Electrical Engineering specialization	76
	Budapest University of Technology and Economics – Social sciences – Manager of technology specialization	69
	Óbuda University - Ágoston Trefort Centre for Engineering Education - Professor engineer specialization	42
Romania	Babes-Bolyai University – Faculty of Economics and Management – Financial and banking specialization	43
	Babes-Bolyai University – Faculty of Economics and Management – Management specialization	15
	Babes-Bolyai University – Faculty of Economics and Management – Tourism and economy specialization	28
	Partium Catholic Faculty – Social Sciences and Economics – Manager of tourism specialization	16
Total number of students		**289**

The hypothesis was: is it assumed that the participating 187 Hungarian and 102 Romanian students would accept their bursaries to be paid in Bitcoin?

The overall number of participants was initially 289. As a first step, it was examined if the students had heard about Bitcoin at all. A total of 41% of the students surveyed (that was 118 students) had not heard about Bitcoin at all, therefore, their further answers were ignored. This was necessary because these subjects could presumably not provide any relevant answers in a topic they had never heard of before. This led to the fact that in the later stages of research, the remaining 171 enquires were taken to be 100% of the students. The following questions were directed towards mapping the students' opinion about Bitcoin, as a prerequisite of the successful implementation of DLCC.

For the question whether Bitcoin would represent the alternative of traditional money, the students gave the following answers: 85% of the subjects answered negatively, while 15% gave a positive answer. This is fully understandable, as the dominance of paper money has long been present in everyday life, for this purpose, it is natural that the people have been forced to put their faith into it.

Considering the question if Bitcoin could represent the alternative of "traditional" electronic means of payment (E-banking, postal order, PayPal, etc.) in everyday

life, the answers were more divided, as 56% of the subjects has given a positive, while 44% of them gave a negative answer. The significance level of the chi-square test regarding this question was p=0,417, so, it can be stated that there is no significant correlation present in this case.

The data indicates that the students are certainly divided when considering Bitcoin's future. This is supported by the question examining the participants' views on how the cryptocurrency would prevail in the next 5 years. The obtained answers are depicted on the Likert scale below.

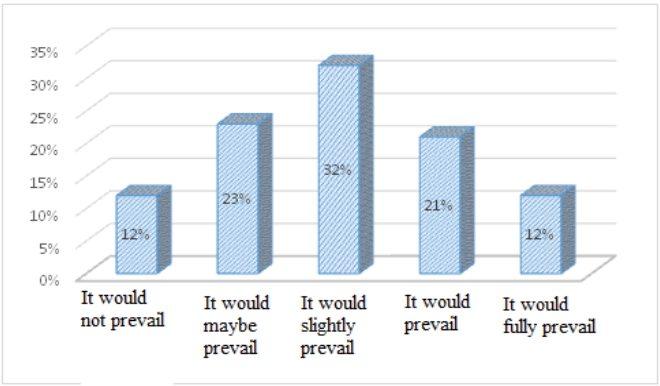


Figure5
What is your opinion about Bitcoin's future?
How will it prevail as a digital currency in the next five years? (n=187)

It can therefore be established that the opinions of students from various countries do not differ significantly regarding the question referring to how Bitcoin would prevail in the next five years (analysis of variance: p=0,619).

The participating university students were asked to give their views on the following question "How would a certain fact influence Bitcoin's future, namely that Bitcoin does not belong under supervision of any state's jurisdiction, or central bank, instead, it is based on a mathematical structure? " The answers can be summarized as follows:

- Unfortunately, knowing Bitcoin's current status, I cannot answer this question (30%).
- The systems used in banks are not fully dependable themselves (27%).
- Regarding Bitcoin's future, in the long term, the P2P structure might have a negative influence (14%).
- Looking at Bitcoin's future, in the long term, the P2P structure might have a positive influence (29%).

Based on the answers given, it can be clearly seen that the future of Bitcoin remains obscure to the students, too. However, what is more surprising is that in

their view, the systems used in banking are not fully dependable. The significance level of chi-square test is p=0,614, therefore, it can be stated that there is no significant correlation.

The last and also most important question was: "If you had the possibility of receiving a bursary, would you accept it in Bitcoins?" The percentage-ratio of positive and negative answers is shown in the graph below.

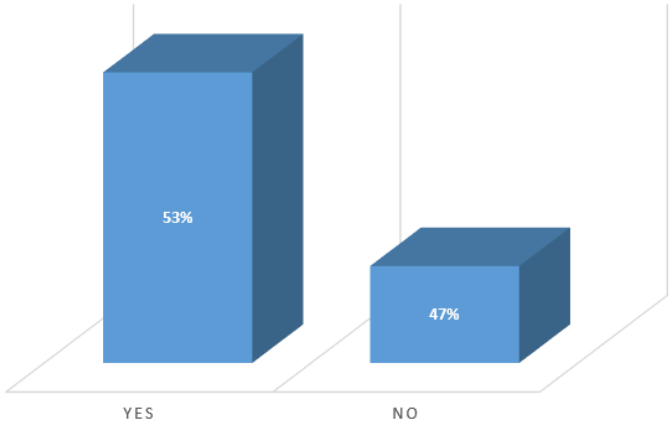


Figure 6
If you had the possibility of receiving a bursary, would you accept it in Bitcoins? (n=187)

It can be established that the hypothesis n, according to which the students would accept a bursary paid in Bitcoins, has been prove, even if they felt both unsure and skeptic about the future of cryptocurrency.

Conclusions

The paper presents a system for introducing a Bitcoin-based bursary system. The technical realization, obtained results and a detailed description of the operational conditions are laid out in this work.

To sum up, in order for the DLCC system to work, there are a number of conditions that must be met in order to work. It is obvious that changing the Bitcoin blockchain is not an easy task, so the paper suggests the use of Soft-Fork for connecting two blockchains. In the DLCC system, within the smart contract's default settings (params.dat) the first step is to define the contract conditions, then setting up the address of the university, then the BTC wallet address of the student. The system extracts the study results from the electronic log and if those meet the requirements included in the contract, the bursary will be paid. Since the system is fully automatic after the setup, it is important to keep the firewall and virus protection up to date, so as to make sure that no data from the electronic log is compromised. Given that the DLCC system is based on the Bitcoin system, this enables a military-level protection that makes it almost impossible to manipulate data.

Further, it is crucial that by the implementation of DLCC system, the administrative tasks would become simpler and more transparent. It is widely known that the institutions of higher education are usually struggling with meager financial bases; nevertheless, the realization of DLCC blockchain would not require enormous spending. One of the possible deficiencies of this system is the lack of trust, as it is utterly new and has never been utilized by the institutions of higher education until this day. If the top management of the universities gave the DLCC solution a vote of confidence, the realization of this Smart Contract bursary payment system could easily be accomplished. In the current situation, what is needed most, is a considerable amount of dedication, rather than serious financial investment.

References

[1] Ameer Rosic: 5 Blockchain Applications That Are Shaping Your Future, https://goo.gl/mY46eM (Download time: 2018-08-09) 2018

[2] Lányi, Márton.: Block technology for logistics. Bánki Reports, (Blokklánc technológia a logisztika szolgálatában), Bánki 2 Közlemények, 1.1 (2018): 5-10

[3] Matteo Ginpietro Zago: 50+ Examples of How Blockchains are Taking Over the World, https://goo.gl/hyYr7y (Download time: 2018-08-09) 2018

[4] Jayanard Sagar: Arcade City Taps Blockchain Technology to Create New-Age Uber, https://goo.gl/AZUfZa, 2016 (Download time: 2018-08-09) 2018

[5] Bernard Mar: 35 Amazing Real World Examples of How Blockchain Is Changing Our World, https://goo.gl/Nv4yWx (Download time: 2018-08-09) 2018

[6] Blockgeeks: 17 Blockchain Applications That Are Transforming Society, https://goo.gl/3qe1Fu (Download time: 2018-08-09) 2018

[7] Adam Hayes: Bitcoin 2.9 Applications, https://goo.gl/rauQa9 (Download time: 2018-08-09) 2018

[8] Upwork: 8 Blockchain Applications That Could Help Your Small Business, https://goo.gl/SVHNcq (Download time: 2018-08-09) 2018

[9] Wiseman, Scott A.: Property or Currency: The Tax Dilemma behind Bitcoin, Utah L. Rev. (2016): 417

[10] Biryukov, Alex, Dmitry Khovratovich, and Ivan Pustogarov: Deanonymisation of clients in Bitcoin P2P network, Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security. ACM, 2014

[11] Vukolić, Marko: The quest for scalable blockchain fabric: Proof-of-work vs. BFT replication, International Workshop on Open Problems in Network Security. Springer, Cham, 2015

[12] Androulaki, Elli, et al: Evaluating user privacy in bitcoin, International Conference on Financial Cryptography and Data Security. Springer, Berlin, Heidelberg, 2013

[13] Miller, Mark S., and Marc Stiegler: The digital path: smart contracts and the Third World, 2003

[14] Juels, Ari, Ahmed Kosba, and Elaine Shi: The ring of gyges: Investigating the future of criminal smart contracts. Proceedings of the 2016 ACM SIGSAC, Conference on Computer and Communications Security. ACM, 2016

[15] StackExchange: Soft-Fork and Hard-Fork in Bitcoin terminology, https://goo.gl/TTsb3E (Download time: 2016-12-12)

[16] Blockchain Blog: A Brief History of Bitcoin Forks, https://goo.gl/YZJ3mV (Download time: 2016-12-12) 2016

[17] MultiChain: Getting Started with MultiChain, https://goo.gl/VO5tvC (Download time: 2016-12-12) 2016

[18] Horváth, László, and Imre J. Rudas: Course Modeling for Student Profile Based Flexible Higher Education on the Internet, J. UCS 12.9 (2006): 1254-1266

[19] Nyikes, Zoltán: Security awareness in the light of digital competence (A biztonságtudatosság a digitális kompetencia tükrében) 2016

[20] Andrea, Hornyák: Segmentation of young people on the basis of financial behavior (A fiatal korosztály szegmentálása a pénzügyi viselkedés alapján) Economists' Forum, Vol. 16, No. 112, 2013

[21] Rudas, Imre J., and János Fodor: Information aggregation in intelligent systems using generalized operators, International Journal of Computers Communications & Control 1.1 (2006): 47-57

[22] Vokorokos, Liberios, Anton Baláž, and Branislav Madoš: Application security through sandbox virtualization, Acta Polytechnica Hungarica 12.1 (2015): 83-101

[23] Beke, Béla: Schedule of study records, (Tanulmányi nyilvántartás rendszerterve) https://goo.gl/AAL3da (Download time: 2016-12-12) (2009)

[24] Kristoufek, Ladislav: What are the main drivers of the Bitcoin price, Evidence from wavelet coherence analysis (2015): e0123923

[25] Rogojanu, Angela, and Liana Badea: The issue of competing currencies, 2014

[26] Bonneau, Joseph, et al: SoK: Research perspectives and challenges for Bitcoin and cryptocurrencies, Security and Privacy (SP), 2015 IEEE Symposium on. IEEE, 2015

[27] Dion, Derek A: I'll Gladly Trade You Two Bits on Tuesday for a Byte Today: Bitcoin, Regulating Fraud in the E-Conomy of Hacker-Cash. U. Ill. JL Tech. & Pol'y (2013): 165

[28] Gál, Veronika Alexandra, and Katalin Gáspár Bencéné Vér: E-money-local money (E-pénz–helyi pénz) Acta Scientiarum Socialium 15.38 (2013)

[29] Press release: National Bank of Hungary presumes virtual means of payment to be risky Bitcoin for example, https://goo.gl/0sfQJk (Download time: 2016-12-12) 2016

[30] Magyar Bitcoin Portál: It is possible to obtain Bitcoin in Romania at 874 terminals, (Romániában már 874 terminálon lehet bitcoint vásárolni), https://goo.gl/YXnyQi (Download time: 2016-12-12) 2016

[31] CCN: Bitcoin Exchange Closes after Lead Programmer Holds Servers Hostage, https://goo.gl/mHGw17 (Download time: 2016-12-12) 2016

High Level Kinematic and Low Level Nonlinear Dynamic Control of Unmanned Ground Vehicles

Béla Lantos, Zsófia Bodó

Budapest University of Technology and Economics, Hungary
H-1117 Budapest, Magyar Tudósok krt. 2., Hungary
E-mail: *lantos@iit.bme.hu*, *zsbodo@iit.bme.hu*

Abstract: High level kinematic model-based control of vehicles is an often used technique in the presence of a driver. Existing robust low level linear (speed, steering, brake, suspension etc.) control components are available in cars which can be influenced using the outputs of the kinematic control as reference signals. If problems arise then the driver can modify the internal control based on the visual information of the path and the observed car motion. In case of unmanned ground vehicles (UGVs) this modification is no more evident. In the paper an approach is presented to estimate the errors in real UGV situations where the road-tire contacts generate special sliding effects in behavior of the UGV. These effects are considered as disturbances and are involved in both the kinematic and dynamic models. The novelties of the paper are the consideration of the sliding effects in the kinematic control and the application of sophisticated nonlinear methods for low level dynamic control. It is demonstrated by simulation that high level kinematic control based approaches can cause lateral errors in the order of 1m. In the experiments three types of low level dynamic controls were considered: i) a simplified one using the steering angle of kinematic control, ii) nonlinear input-output linearization (DGA method), and iii) flatness control. They can supply the sliding angle information for the kinematic control.

Keywords: Kinematic Control, Sliding Effects, UGV, Kinematic and Dynamic Coupling, Path Following, Input-Output Linearization, Flatness Control

1 Introduction

Vehicle control based on the kinematic model is a popular approach delivering speed and steering angle commands for the existing robust low level control subsystems. If problems arise and a driver is present then the necessary corrections can be performed manually using the available visual information and the observed difference between the path and the car's motion. However, in case of unmanned ground vehicles (UGVs) this modification is no more possible.

In the paper an approach is presented to estimate the errors in real UGV situations where the road-tire contacts generate special sliding effects in the dynamic behavior of the UGV. These effects are usually not considered in the kinematic modeling where the side motion of the vehicle is neglected and nonholonomic constraints are assumed for the front and rear wheels, see e.g. De Luca and coworkers [1].

A remarkable exception is the approach of Arogeti and Berman [2] whose method can also manage the sliding effects involved in the kinematic model in the form of disturbances. Their method is based on the results of Scherer and Weiland [3] for decreasing the peak-to-peak L_∞ (or generalized H_2) disturbance effects in single variable (SISO) systems. Arogeti and Berman involved the sliding effects in the kinematic model and presented a modified path following kinematic control method. Since the kinematic and dynamic models are coupled through the sliding effects therefore a realistic testing cannot be performed without an appropriate dynamic control method. The paper [2] demonstrates using simulation that with the modified kinematic control the slip angles remain in realistic and acceptable domains. Unfortunately, it cannot be pointed out from [2] what was the dynamic control method during the test. Similarly, no data was shown about the path following errors.

Our earlier paper [4] considered the similar problem but the main goal was to show what is the order of the lateral error if the steering angle of the modified kinematic control is saved in the low level dynamic control, and if it is large, how can it be decreased by dynamic control. Since kinematic and dynamic models are coupled through the slip angles hence realistic dynamic control of the velocity and the acceleration is needed for correct analysis of the lateral error (the slip angles depend on the velocities and the steering angle). For this purpose a dynamic control was also developed which is based on nonlinear input-output linearization (dynamic inversion). The main novelty of our present paper is the extension of the dynamic control methods with the nonlinear flatness control and the proof of the flatness property both for front and rear wheel driven cars.

Other popular methods exist using PID-type control of linekeeping [5], potential field technique [6] and nonlinear time-optimal control [7].

The paper is organized as follows. Section 2 presents the modified kinematic control if the slip angles are taken into consideration and the disturbance effects have to be reduced. Section 3 deals with the suggested nonlinear dynamic control methods for testing. Section 4 shows the numerical results of the simulation for the parallel running modified kinematic and dynamical control and the analysis of the lateral error. Section 5 concludes the paper.

2 MODIFIED KINEMATIC CONTROL

For the kinematic and dynamical investigations in the paper the well known two wheel bicycle model will be used, see Fig. 1. The notations are as usual, i.e. front (F) and rear (R) are wheels, longitudinal (l), and transversal (t) stand for forces, M is the moment, CoG is the center of gravity, v is velocity, β stands for the side slip

angle, α denote the slip angles of the wheels, ψ is the orientation (heading), and δ_w is the steering angle in the figure. The other parameters are geometrical ones and $L := l_R + l_F$. From the frames x_0 and y_0 is the inertia system, x_{CoG} and y_{CoG} is the body system and x_w and y_w is the front wheel system. In the paper front wheel steering and rear wheel accelerating are assumed.

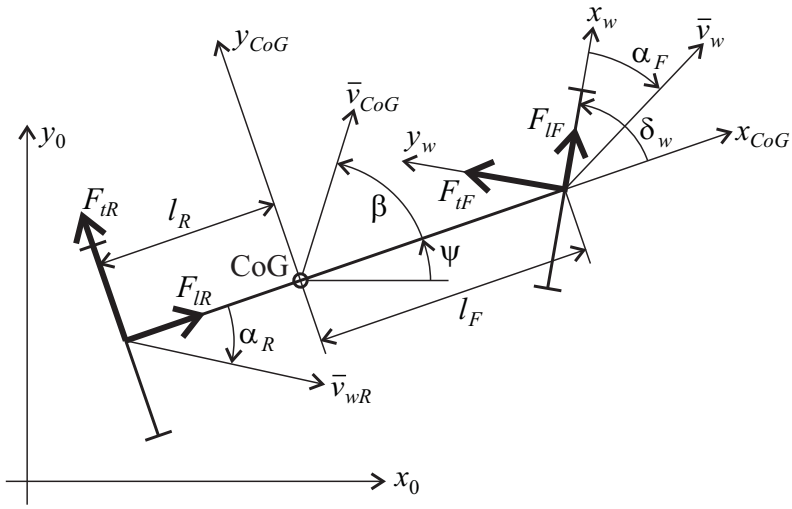


Figure 1
The two wheel (bicycle) structure

For path design and kinematic modeling the coordinate system will be fixed to the middle point of the rear axle instead of the CoG. Kinematic models satisfying the nonholonomic constraints can be brought to chain form and stabilized by state feedback [1].

The convergence of error decaying strongly depends on the speed variable $\bar{v} = \dot{x} = \cos(\psi)v$ that depends also on the orientation ψ. Furthermore, the singularities of the chain transformation should be avoided during path design.

The modified form will be discussed in two steps. First the error definition is modified and after it the slip angles will be taken into consideration.

2.1 Modified error definition for the chain form

In order to eliminate the dependence of the control u on the orientation, the basic paper of Arogeti and Berman [2] defines the tracking error by

$$\begin{aligned} e_1 &= f(x) - y \\ e_2 &= f'(x)\cos(\psi) - \sin(\psi) \\ e_3 &= f''(x)\cos^2(\psi) - \frac{\tan(\delta_w)}{L}\left(f'(x)\sin(\psi) + \cos(\psi)\right) \end{aligned} \tag{1}$$

where e_1 represents the position error in lateral direction, e_2 is the orientation error and e_3 is the steering error, and $y_d = f(x)$ is the desired path.

Denote v the absolute value (magnitude) of the velocity in the middle point of the rear axle (i.e. $\bar{v} := v$) which makes a great difference to the original method because v does not depend on the orientation ψ.

The transformation to the chain form is completed by the control signal

$$\begin{aligned} w = &\Big[\Big(f'''(x)\cos^3(\psi) - 3\frac{f''(x)\cos(\psi)\sin(\psi)\tan(\delta_w)}{L} \\ &- \frac{f'(x)\cos(\psi)\tan^2(\delta_w)}{L^2} + \frac{\sin(\psi)\tan^2(\delta_w)}{L^2}\Big)v - u\Big] \\ &\times \frac{L\cos^2(\delta_w)}{f'(x)\sin(\psi) + \cos(\psi)} \end{aligned} \tag{2}$$

where u is the stabilizing state feedback.

The new chain form is

$$\dot{e}_1 = e_2 v, \quad \dot{e}_2 = e_3 v, \quad \dot{e}_3 = u \tag{3}$$

The physical interpretation of e_1 is the vehicle lateral error and e_2 is the orientation (heading) error. Along the path e_1 and e_2 are zero hence the reference value of the orientation is $\tan(\psi_r(x)) = f'(x)$, i.e.

$$\psi_r(x) = \arctan(f'(x)) \tag{4}$$

Considering e_3 for $e_2 = 0$ it yields $f'(x)\sin(\psi_r) + \cos(\psi_r) = \tan(\psi_r)\sin(\psi_r) + \cos(\psi_r) = 1/\cos(\psi_r)$ and one obtains for $e_3 = 0$:

$$\tan(\delta_{wr}) = Lf''(x)\cos^3(\psi_r) \tag{5}$$

2.2 Kinematic control in the presence of sliding effects

The earlier discussion assumed rolling without side motion (slipping) of the wheels. Considering the bicycle model, the velocity vector v_R, the slip angles α_R and α_F and denoting the projection of the velocity vector in x-direction of the car by $v = |v_R|\cos(\alpha_R) \Leftrightarrow |v_R| = v/\cos(\alpha_R)$, then the kinematic equations in the presence of sliding effects can be written as follows:

$$\begin{aligned} \dot{x} &= \frac{\cos(\psi + \alpha_R)}{\cos(\alpha_R)} v = (\cos(\psi) - \tan(\alpha_R)\sin(\psi))\, v \\ \dot{y} &= \frac{\sin(\psi + \alpha_R)}{\cos(\alpha_R)} v = (\sin(\psi) - \tan(\alpha_R)\cos(\psi))\, v \\ \dot{\psi} &= \frac{\tan(\delta_w - \alpha_F) + \tan(\alpha_R)}{L} v \\ \dot{\delta}_w &= w \end{aligned} \tag{6}$$

The design objective is to follow the prescribed reference path $y_d = f(x)$. Using (1), (2) and (6) the tracking error can be written in the form consisting of two components:

$$\begin{bmatrix} \dot{e}_1 \\ \dot{e}_2 \\ \dot{e}_3 \end{bmatrix} = \begin{bmatrix} e_2 v \\ e_3 v \\ u \end{bmatrix} + \begin{bmatrix} g_1(\psi, f', \alpha_R) \\ g_2(\psi, \delta_w, f', f'', \alpha_R, \alpha_F) \\ g_3(\psi, \delta_w, f', f'', f''', \alpha_R, \alpha_F) \end{bmatrix} v \tag{7}$$

The functions g_1, g_2 and g_3 are nonlinear functions defined in [2]. The vehicle heading ψ_r and steering angle δ_{wr} along the path are given in (4) and (5). This second nonlinear term in (7) can be linearized in a small neighborhood of the desired path and the zero slip angles resulting in

$$\begin{aligned} \dot{e} = & \underbrace{\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}}_{A} v e + \underbrace{\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}}_{B_2} u \\ & + \underbrace{\begin{bmatrix} \bar{g}_{11}(\psi_r) & 0 \\ \bar{g}_{21}(\psi_r, \delta_{wr}) & \bar{g}_{22}(\psi_r, \delta_{wr}) \\ \bar{g}_{31}(\psi_r, \delta_{wr}, f''') & \bar{g}_{32}(\psi_r, \delta_{wr}) \end{bmatrix}}_{B_1(t)=B_1(\psi_r, \delta_{wr}, f''')} v \underbrace{\begin{bmatrix} \alpha_R \\ \alpha_F \end{bmatrix}}_{d(t)} \end{aligned} \tag{8}$$

2.3 Robust kinematic control

For an LTI systems of class $\dot{e} = Ae + Bd$, $z = Ce$, $e \in R^n$, $d \in R^m$ and $z \in R^p$ Scherer and Weiland [3] developed a method for the design of a control u satisfying peak–to–peak performance, i.e. $||z||_\infty \leq \gamma ||d||_\infty$ for given $\gamma > 0$ based on LMI technique.

The functions g_i and $\bar{g}_{ij}$ are listed in Arogeti and Berman [2] without derivation. Because of the central role of these functions their validity was also checked by the authors of this paper. In the sequel some detected errors of the above paper are also corrected, especially the correct order of the terms to find upper bounds for $B_1(t)B_1^T(t) < \bar{B}_1 \bar{B}_1^T$, $\forall t$.

The model (8) consists of two parts. The first part is the new chain form $\dot{e} = Aev + B_2 u$ according to (3), while the second $B_1(\psi_r, \delta_{wr}, f''')vd(t)$ can be treated as an unknown model disturbance. The matrix $B_1(\cdot)$ is a function of the reference path thus all its elements are bounded, i.e. they are in L_∞, thus the robust design approach should be based on the nonstandard L_∞ valued performance optimization. Notice that model (8) belongs to the LTI system class in [2] with $n = 3$ and $m = 2$.

First we considered the 3×3 type matrix $B_1(t)B_1(t)^T$ along the path and determined a constant matrix $\bar{B}_1$ satisfying $B_1(t)B_1^T(t) < \bar{B}_1 \bar{B}_1^T$ for $\forall t$.

Notice that model (8) belongs to the LTI system class in [2] with $n = 3$, $m = 2$ and $p = 4$. Consider for $e(0) = 0$ the linear time-varying system

$$\begin{aligned} \dot{e}(t) &= Av(t)e(t) + B_2 u(t) + B_1 v(t) d(t) \\ z(t) &= Ce(t) + Du(t) \end{aligned} \tag{9}$$

with the input $u \in R$, $d \in R^2$ and the controlled output $z \in R^p$ and the bounds

$$\begin{aligned} 0 < \eta_1 < v(t) < \eta_2 < \infty \\ B_1(t)B_1^T(t) < \bar{B}_1\bar{B}_1^T \end{aligned} \tag{10}$$

Using the state feedback $u(t) = Ke(t)v(t)$ the closed loop system will be

$$\begin{aligned} \dot{e}(t) &= (A + B_2K)v(t)e(t) + B_2u(t) + B_1v(t)d(t) \\ z(t) &= (C + DKv(t))e(t) = \bar{C}e(t) \end{aligned} \tag{11}$$

Based on the results of [3] and using the bounds in (10) it was shown in [2] that given the system (11) and a scalar $\gamma > 0$, assume there exist $0 < \lambda \in R$, $0 < Q \in R^{n\times n}$ and $Y \in R^{p\times n}$ such that the two LMIs are satisfied, i.e.

$$\begin{aligned} &\begin{bmatrix} (QA^T + AQ + Y^TB_2^T + B_2Y)\eta_1 + \lambda nQ & \bar{B}_1\eta_2 \\ \bar{B}_1^T\eta_2 & -\gamma I_m \end{bmatrix} < 0 \\ &\begin{bmatrix} \lambda Q & QC^T + Y^TD^T \\ CQ + DY & \gamma I_p \end{bmatrix} > 0 \end{aligned} \tag{12}$$

Then, for control gains given by $K = YQ^{-1}$, the closed loop system norm satisfies $||L_{cl}||_\infty < \gamma$, and the system is internally asymptotically stable. Notice yet that the two LMIs are coupled in Q and Y that determine the state feedback K.

3 ADVANCED NONLINEAR DYNAMIC CONTROL

The dynamic model of the vehicle will be considered in the frame fixed to CoG while the origin of the frame used for kinematic control is at the middle point of the rear axle. Fortunately, the two frames are parallel. For simplicity denote v_G the absolute value of the velocity at CoG. The angles β, α_F, α_R are usually called the vehicle body side slip angle, the tire slide slip angle front and the tire slide slip angle rear, respectively.

The tire slip angles are defined by

$$\tan(\alpha_R) = \frac{l_R\dot{\psi} - v_G\sin(\beta)}{v_G\cos(\beta)} \tag{13}$$

$$\tan(\delta_W - \alpha_F) = \frac{l_F\dot{\psi} + v_G\sin(\beta)}{v_G\cos(\beta)} \tag{14}$$

The forces acting at the origin of the coordinate system of the front wheel are the longitudinal force F_{lF} and the transversal force F_{tF}. It should be underlined that in the state equations of the *dynamic modeling* the transversal forces are described by the Pacejka's equations [8] in order to obtain reliable results for the lateral errors. On the other hand, during the *dynamic control design*, as usual in the vehicle control literature, the lateral forces are approximated by the cornering stiffnesses, in order to omit nonlinear dynamic optimization in real time.

3.1 Input affine dynamic model for control design

For dynamic control design it is assumed that the transversal components are $F_{tF} = c_F\alpha_F$ and $F_{tR} = c_R\alpha_R$, respectively, where the cornering stiffnesses c_F and c_R are constants. Assuming small $\delta_W - \alpha_F$, α_R and β and using the approximations $\tan(\delta_W - \alpha_F) \approx \delta_W - \alpha_F$, $\tan(\alpha_R) \approx \alpha_R$, $\sin(\beta) \approx \beta$ and $\cos(\beta) \approx 1$ it follows

$$\alpha_F = \delta_W - \beta - \frac{l_F\dot{\psi}}{v_G}, \quad \alpha_R = -\beta + \frac{l_R\dot{\psi}}{v_G}$$

Applying the usual notations $\cos(\beta) = C_\beta$, $\sin(\beta) = S_\beta$, and the differentiation rule in moving frames, then the kinematic model and based on it the dynamic model of the car can be derived.

$$\begin{aligned}
\bar{v}_{COG} &= \begin{pmatrix} C_\beta & S_\beta & 0 \end{pmatrix}^T v_G \\
\bar{a}_{COG} &= \dot{\bar{v}}_{COG} + \omega \times \bar{v}_{COG} \\
&= \begin{bmatrix} C_\beta & -v_G S_\beta \\ S_\beta & v_G C_\beta \\ 0 & 0 \end{bmatrix} \begin{pmatrix} \dot{v}_G \\ \dot{\beta} \end{pmatrix} + \begin{pmatrix} -S_\beta \\ C_\beta \\ 0 \end{pmatrix} \dot{\psi} v_G
\end{aligned}$$

Taking into account the direction of the forces, dividing by the mass m_v of the car, and considering only the nontrivial components of $\bar{a}_{COG}$, the acceleration is obtained:

$$\bar{a}_{COG} = \frac{1}{m_v}\begin{pmatrix} F_x \\ F_y \end{pmatrix} = \frac{1}{m_v}\begin{pmatrix} F_{lF}C_{\delta_w} - F_{tF}S_{\delta_w} + F_{lR} \\ F_{lF}S_{\delta_w} + F_{tF}C_{\delta_w} + F_{tR} \end{pmatrix}$$

3.2 Nominal dynamic control saving kinematic steering angle

Because forward axle steering and rear axle driving are assumed in this paper therefor the (e.g. industrial) low level dynamic control can be modeled in such a simple form that the rear longitudinal force is determined from the state equation and the steering angle of the kinematic control is used as steering angle of the dynamic control. This approach will be called *nominal control*. The driving force in nominal control can easily be computed:

$$F_{lR} = m_v(\dot{v}_{cx} - v_{cy}\dot{\psi}) + F_{tF}S_{\delta_w} \tag{15}$$

The differentiation of v_{cx} can be approximated by a fictitious control loop, see also [5], but with other choice of the parameters in the fictitious controller.

3.3 Differential Geometry Based Control Algorithm

It is useful to introduce the rear ($S_h = F_{tR}$) and front ($S_v = F_{tF}$) side forces where

$$S_h = c_R(-\beta + l_R\dot{\psi}/v_G), \tag{16}$$

$$S_v = c_F(\delta_w - \beta - l_F\dot{\psi}/v_G). \tag{17}$$

It is clear that steering angle δ_w can be determined for control implementation by

$$\delta_w = \frac{1}{c_F} S_v + \beta + l_F \dot{\psi} / v_G \tag{18}$$

Assuming small angles the input affine model arises as

$$\dot{x} = \begin{bmatrix} -x_3 + S_h/(m_v x_4) \\ x_3 \\ -S_h l_R / I_z \\ 0 \\ x_4 C_{12} \\ x_4 S_{12} \end{bmatrix} + \begin{bmatrix} 1/(m_v x_4) & -x_1/(m_v x_4) \\ 0 & 0 \\ l_F/I_z & 0 \\ 0 & 1/m_v \\ 0 & 0 \\ 0 & 0 \end{bmatrix} u, \tag{19}$$

$$\dot{x} = A(x) + B(x)u, \quad y = (x_5, x_6)^T = C(x),$$
$$x = (\beta, \psi, \dot{\psi}, v_G, X, Y)^T, \; u = (S_v, F_{lR})^T, \; y = (X, Y)^T.$$

Here we used the notation $C_{12} = \cos(x_1 + x_2)$ etc. and X and Y are the coordinates of the CoG in the inertia frame.

It can be shown [9] that the above approximated dynamic model has vector relative degrees $r_1 = r_2 = 2$. Thus (the observable subsystem) has the form

$$\begin{pmatrix} \ddot{y}_1 \\ \ddot{y}_2 \end{pmatrix} = q(x) + S(x)u \tag{20}$$

$$S(x) = \frac{1}{m_v} \begin{bmatrix} -S_{12} & C_{12} + x_1 S_{12} \\ C_{12} & S_{12} - x_1 C_{12} \end{bmatrix}, \tag{21}$$

$$q(x) = \frac{1}{m_v} \begin{pmatrix} -S_{12} \\ C_{12} \end{pmatrix} S_h. \tag{22}$$

Hence the system can be input-output linearized by an internal nonlinear feedback and the resulting system consists of two double integrators:

$$u := S^{-1}[v - q(x)], \tag{23}$$
$$\ddot{y}_i = v_i, \;\; i = 1, 2 \tag{24}$$

The stability of the zero dynamics was proven in [9].

First we assume a prescribed stable and sufficiently quick error dynamics

$$(\ddot{y}_{di} - \ddot{y}_i) + \alpha_{1i}(\dot{y}_{di} - \dot{y}_i) + \lambda_i(y_{di} - y_i) = 0 \tag{25}$$

Let us use the notation $w_i := y_{di} + (\alpha_{1i}\dot{y}_{di} + \ddot{y}_{di})/\lambda_i$, then

$$\ddot{y}_i = v_i = \lambda_i w_i - \alpha_{1i}\dot{y}_i - \lambda_i y_i \tag{26}$$

Observe that $\lambda_i w_i$ depends only on the reference signal and its derivatives (feed forward from the reference signal) and $-\alpha_{1i}\dot{y}_i - \lambda_i y_i$ is the state feedback stabilizing

the system where α_{1i} and λ_i are positive for stable error dynamics. Let $\lambda_1 = \lambda_2 := \lambda$, $\alpha_{11} = \alpha_{12} = 2\sqrt{\lambda}$ (or similar ones) where $\lambda > 0$, then two decoupled linear systems are arising whose characteristic equation and differential equation are, respectively,

$$s^2 + 2\sqrt{\lambda}s + \lambda = 0 \Rightarrow s_{1,2} = -\sqrt{\lambda}, \tag{27}$$

$$\ddot{y}_i + \alpha_{1i}\dot{y}_i + \lambda y_i = \lambda_i w_i. \tag{28}$$

The steps of the *DGA Control Algorithm (DGA)*:

1. $w_i := y_{di} + \frac{1}{\lambda_i}(\alpha_{1i}\dot{y}_{di} + \ddot{y}_{di}),\ \ i = 1,2$
2. $\bar{y}_1 := \lambda_1 w_1 - \alpha_{11}(x_4 C_{12}) - \lambda_1 x_5$
3. $\bar{y}_2 := \lambda_2 w_2 - \alpha_{12}(x_4 S_{12}) - \lambda_2 x_6$
4. $S_h := c_R\,[-x_1 + (l_R x_3 / x_4)]$
5. $u_1 := -S_h + m_v\,[(x_1 C_{12} - S_{12})\bar{y}_1 + (x_1 S_{12} + C_{12})\bar{y}_2]$
6. $u_2 := m_v(C_{12}\bar{y}_1 + S_{12}\bar{y}_2)$
7. $\delta_w := (u_1 / c_F) + x_1 + (l_F x_3 / x_4)$
8. $F_{lR} := u_2$

Notice, because not all state variables can be measured, a state estimator has to be implemented in order to supply the necessary state information for the controller. This problem was discussed in Chapter 5, pp. 189-195 of [9].

3.4 Flatness control

A nonlinear system $\dot{x} = f(x,u)$, $x \in R^n$, $u \in R^m$ is said to be differentially flat if there exists a vector $y = (y_1, \ldots, y_m)^T \in R^m$ called the flat output and vector valued functions and integers such that $y = h(x, u, \dot{u}, \ldots, u^{(r)})$, $x = A(y, \dot{y}, \ldots, y^{(r_x)})$, and $u = B(y, \dot{y}, \ldots, y^{(r_u)})$, see [10] and [9]. Notice that y and u have equal dimension.

The two wheels (bicycle) vehicle dynamic model can be approximated by using the flat outputs $y_1 = V_x$ and $y_2 = l_F m V_y - I_z \dot{\psi}$, and the controls $u_1 = T_m - T_b$ and $u_2 = \delta_w$, respectively.

3.4.1 Flatness proof for front wheel or rear wheel driven cars

A sketch of the flatness proof can be found for front steering and front driving in the recent paper [11]. Since rear wheel driving is used in the kinematical part of our paper hence a generalization for both driving cases will be given.

In order to show the similarities and/or differences we will use a similar notation as the cited paper. Most of the notations are evident: for simplicity we omit the index w in δ_w; V_x and V_y are the velocity components of CoG; instead of l (longitudinal) and t (transversal) indexes of forces the indexes x and y will be used. It will be

assumed that the braking forces satisfy $T_{br} = rT_{bf}$, $r \in [0,1]$, hence the total braking force is $T_b = (1+r)T_{bf}$ and thus

$$T_{bf} = \frac{1}{1+r}T_b, \quad T_{br} = \frac{r}{1+r}T_b \tag{29}$$

Denote the tire effective radius R, the wheel inertia I_ω and the wheel angular velocities ω_f and ω_r, respectively. The wheel angular velocities are assumed to be measured by odometers.

For *front wheel driven car* yields:

$$\begin{aligned} F_{xf} &= \frac{1}{R}\left(-I_\omega \dot{\omega}_f + T_m - T_{bf}\right) \\ F_{xr} &= \frac{1}{R}\left(-I_\omega \dot{\omega}_r - T_{br}\right) \end{aligned} \tag{30}$$

For *rear wheel driven car* yields:

$$\begin{aligned} F_{xf} &= \frac{1}{R}\left(-I_\omega \dot{\omega}_f - T_{bf}\right) \\ F_{xr} &= \frac{1}{R}\left(-I_\omega \dot{\omega}_r + T_m - T_{br}\right) \end{aligned} \tag{31}$$

The basic dynamic motion equations are as follows:

$$\begin{aligned} m(\dot{V}_x - \dot{\psi}V_y) &= F_{xf}\cos(\delta) - F_{yf}\sin(\delta) + F_{xr} \\ m(\dot{V}_y + \dot{\psi}V_x) &= F_{xf}\sin(\delta) + F_{yf}\cos(\delta) + Fyr \\ I_z\ddot{\psi} &= l_F\left(F_{yf}\cos(\delta) + F_{xf}\sin(\delta)\right) - l_R F_{yr} \end{aligned} \tag{32}$$

Assuming as usual small angles and lateral forces approximated by using cornering stiffness the above equations can be simplified.

For front wheel driven car:

$$\begin{aligned} \dot{V}_x &= \dot{\psi}V_y - \frac{I_\omega}{mR}(\dot{\omega}_r + \dot{\omega}_f) + \frac{1}{mR}(T_m - T_b) + \frac{c_F}{m}\frac{V_y + l_F\dot{\psi}}{V_x}\delta - \frac{c_F}{m}\delta^2 \\ \dot{V}_y &= -\dot{\psi}V_x - \frac{c_F}{m}\frac{V_y + l_F\dot{\psi}}{V_x} - \frac{c_R}{m}\frac{V_y - l_R\dot{\psi}}{V_x} + \frac{1}{mR}(T_m - T_{bf})\delta + \frac{c_F R - I_\omega\dot{\omega}_f}{mR}\delta \\ \ddot{\psi} &= \frac{1}{I_z}\left[-l_F c_F\frac{V_y + l_F\dot{\psi}}{V_x} + l_R c_R\frac{V_y - l_R\dot{\psi}}{V_x} + \frac{l_F}{R}(T_m - T_{bf})\delta + \frac{l_F}{R}(c_F R - I_\omega\dot{\omega}_f)\delta\right] \end{aligned} \tag{33}$$

For rear wheel driven car:

$$\begin{aligned} \dot{V}_x &= \dot{\psi}V_y - \frac{I_\omega}{mR}(\dot{\omega}_r + \dot{\omega}_f) + \frac{1}{mR}(T_m - T_b) + \frac{c_F}{m}\frac{V_y + l_F\dot{\psi}}{V_x}\delta - \frac{c_F}{m}\delta^2 \\ \dot{V}_y &= -\dot{\psi}V_x - \frac{c_F}{m}\frac{V_y + l_F\dot{\psi}}{V_x} - \frac{c_R}{m}\frac{V_y - l_R\dot{\psi}}{V_x} - \frac{1}{mR}T_{bf}\delta + \frac{c_F R - I_\omega\dot{\omega}_f}{mR}\delta \\ \ddot{\psi} &= \frac{1}{I_z}\left[-l_F c_F\frac{V_y + l_F\dot{\psi}}{V_x} + l_R c_R\frac{V_y - l_R\dot{\psi}}{V_x} - \frac{l_F}{R}T_{bf}\delta + \frac{l_F}{R}(c_F R - I_\omega\dot{\omega}_f)\delta\right] \end{aligned} \tag{34}$$

As can be seen, the differences in the two driving modes are in $\dot{V}_y$ and $\ddot{\psi}$. Moreover, if $u_1 = T_m - T_b$ is already known then for $u_1 > 0 \Rightarrow T_m = u_1, T_b = 0$ while for $u_1 \leq 0 \Rightarrow T_m = 0, T_b = -u_1$ can be chosen and T_{bf}, T_{br} can also be computed using the selected value of r.

Hence the two driving modes can be standardized using the notation $\gamma(u_1) = T_m - T_b$ for rear wheel driven car and $\gamma(u_1) = -T_b$ for front wheel driven car, respectively.

With the notations

$$f(x) = \begin{bmatrix} \dot{\psi} V_y - \frac{I_\omega}{mR}(\dot{\omega}_r + \dot{\omega}_f) \\ -\dot{\psi} V_x - \frac{c_F}{m}\frac{V_y + l_F\dot{\psi}}{V_x} - \frac{c_R}{m}\frac{V_y - l_R\dot{\psi}}{V_x} \\ \frac{1}{I_z}\left(-l_F c_F \frac{V_y + l_F\dot{\psi}}{V_x} + l_R c_R \frac{V_y - l_R\dot{\psi}}{V_x}\right) \end{bmatrix}$$

$$g(x) = \begin{bmatrix} \frac{1}{mR} & \frac{c_F}{m}\frac{V_y + l_F\dot{\psi}}{V_x} \\ 0 & \frac{c_F R - I_\omega \dot{\omega}_f}{mR} \\ 0 & \frac{l_F}{I_z R}(c_F R - I_\omega \dot{\omega}_f) \end{bmatrix}, \quad g_1 = \begin{bmatrix} 0 \\ \frac{1}{mR} \\ \frac{l_R}{I_z R} \end{bmatrix}, \quad g_2 = \begin{bmatrix} -\frac{c_F}{m} \\ 0 \\ 0 \end{bmatrix} \tag{35}$$

the state equation can be written as follows:

$$\dot{x} = f(x) + g(x)u + g_1\gamma(u_1)u_2 + g_2 u_2^2 \tag{36}$$

It remains to prove the flatness of the system for the outputs $y_1 = V_x$ and $y_2 = l_F m V_y - I_z\dot{\psi}$. Using the earlier results and adding zero in special form yields

$$\begin{aligned} \dot{y}_2 =& l_F\left[-m\dot{\psi}V_x - c_F\frac{V_y + l_F\dot{\psi}}{V_x} - c_R\frac{V_y - l_R\dot{\psi}}{V_x} + \frac{1}{R}\gamma(u_1)\delta + \frac{c_F R - I_\omega\dot{\omega}_f}{R}\delta\right] \\ &+ \left[l_F c_F\frac{V_y + l_F\dot{\psi}}{V_x} - l_R c_R\frac{V_y - l_R\dot{\psi}}{V_x} - \frac{l_F}{R}\gamma(u_1)\delta - \frac{l_F}{R}(c_F R - I_\omega\dot{\omega}_f)\delta\right] \\ &+ (l_F c_R - l_F c_R)\frac{V_y - l_R\dot{\psi}}{V_x} \end{aligned}$$

Now we can cancel the appropriate positive and negative terms in pair and obtain

$$\dot{y}_2 = -l_F m\dot{\psi}V_x - (l_F + l_R)c_R\frac{V_y - l_R\dot{\psi}}{V_x} \tag{37}$$

Multiplying with $y_1 = V_x$ it follows

$$y_1\dot{y}_2 = \left[-l_F m y_1^2 + (l_F + l_R)c_R l_R\right]\dot{\psi} - (l_F + l_R)c_R\frac{y_2 + I_z\dot{\psi}}{l_F m}$$

from which $\dot{\psi}$ and V_y can be determined:

$$\dot{\psi} = -\frac{l_F m y_1\dot{y}_2 + (l_F + l_R)c_R y_2}{(l_F + l_R)c_R(I_z - l_F l_R m) + (l_F m y_1)^2} \tag{38}$$

$$V_y = \frac{y_2 + I_z\dot{\psi}}{l_F m} = \frac{y_2}{l_F m} - \frac{I_z}{l_F m}\frac{l_F m y_1\dot{y}_2 + (l_F + l_R)c_R y_2}{(l_F + l_R)c_R(I_z - l_F l_R m) + (l_F m y_1)^2} \tag{39}$$

Hence $x = (V_x, V_y, \dot{\psi})^T = A(y_1, y_2, \dot{y}_2)$ and $r_x = 1$.

Unfortunately, to the flatness property of u we need also $\ddot{y}_2$:

$$\ddot{y}_2 = -l_F m \ddot{\psi} V_x - l_F m \dot{\psi} \dot{V}_x - \frac{(l_F + l_R) c_R (\dot{V}_y - l_R \ddot{\psi})}{V_x} + \frac{(l_F + l_R) c_R (V_y - l_R \dot{\psi}) \dot{V}_x}{V_x^2} \tag{40}$$

The following form will be derived:

$$\begin{bmatrix} \dot{y}_1 \\ \ddot{y}_2 \end{bmatrix} = \Delta(y_1, y_2, \dot{y}_2) \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} + \Phi(y_1, y_2, \dot{y}_2) \tag{41}$$

Using (35) and (36) for the derivatives of the state variables and the structure of (40), then the following choice can be made:

$$\begin{aligned}
\Delta_{11} =& g_{11} = \frac{1}{mR} \\
\Delta_{12} =& g_{12} = \frac{c_F}{m} \frac{V_y + l_F \dot{\psi}}{y_1} \\
\Delta_{21} =& \frac{c_R (l_F + l_R)(V_y - l_R \dot{\psi}) - l_F m \dot{\psi} y_1^2}{m R y_1^2} \\
\Delta_{22} =& \frac{l_R c_R (l_F + l_R) - l_F m y_1^2}{y_1} \frac{l_F c_F R - l_F I_\omega \dot{\omega}_f}{I_z R} \\
&+ \frac{c_R (l_F + l_R)(V_y - l_R \dot{\psi}) - l_F m \dot{\psi} y_1^2}{y_1^2} \frac{c_F (V_y + l_F \dot{\psi}}{m y_1} \\
&- \frac{c_R (l_F + l_R)}{y_1} \frac{R c_F - I_\omega \dot{\omega}_f}{mR}
\end{aligned} \tag{42}$$

$$\begin{aligned}
\Phi_1 =& f_1(x) + g_{21} \delta^2 = \dot{\psi} V_y - \frac{I_\omega}{mR} (\dot{\omega}_r + \dot{\omega}_f) - \frac{c_F}{m} u_2^2 \\
\Phi_2 =& - l_F m y_1 [f_3(x) + g_{13} \gamma(u_1) u_2] \\
&- \frac{(l_F + l_R) c_R}{y_1} [f_2(x) + g_{12} \gamma(u_1) u_2] \\
&+ \frac{(l_F + l_R) c_R (V_y - l_R \dot{\psi}) - l_F m \dot{\psi} y_1^2}{y_1^2} \left[f_1(x) + g_{21} u_2^2 \right] \\
&+ \frac{(l_F + l_R) c_R l_R}{y_1} [f_3(x) + g_{13} \gamma(u_1) u_2]
\end{aligned} \tag{43}$$

The determinant of the matrix Δ satisfies

$$\begin{aligned}
\det(\Delta) &= \Delta_{11} \Delta_{22} - \Delta_{21} \Delta_{12} \\
&= \frac{(I_\omega \dot{\omega}_f - c_F R) \left[l_F^2 y_1^2 m^2 - c_R (l_F + l_R) l_R l_F m + c_R I_z (l_F + l_R) \right]}{I_z R^2 y_1 m^2} \neq 0
\end{aligned} \tag{44}$$

Then taking into consideration that for typical cars Rc_F/I_ω is around 10^4 and if $I_z > l_F m$ then $c_R(l_R+l_F)(I_z-l_F m)+l_F^2 m^2 y_1^2 \neq 0$, hence neglecting the non-dominant terms $\gamma(u_1)u_2$ and u_2^2 in Φ the control u can be determined from the flatness variables and their derivatives:

$$u = B(y_1, y_2, \dot{y}_1, \dot{y}_2, \ddot{y}_2) = \Delta^{-1}\left(\begin{bmatrix} \dot{y}_1 \\ \ddot{y}_2 \end{bmatrix} - \Phi\right), \quad r_u = 2 \tag{45}$$

Remark: Based on this initial value the control can be further improved by considering the nonlinear terms in Φ and finding the fix point in iterations.

3.4.2 Flatness based control algorithm

For the different flatness variables linear reference systems can be prescribed:

$$\begin{bmatrix} \dot{y}_1 \\ \ddot{y}_2 \end{bmatrix} = \begin{bmatrix} \dot{y}_1^{ref} + k_{1p}e_{y_1} + k_{1I}\int e_{y_1}dt \\ \ddot{y}_2^{ref} + k_{2p}e_{y_2} + k_{2I}\int e_{y_2}dt + k_{2D}\dot{e}_{y_2} \end{bmatrix} \tag{46}$$

where $e_{y_1} = y_1^{ref} - y_1 = V_x^{ref} - V_x$ and $e_{y_2} = e_{y_2}^{ref} - y_2$ and $y_2^{ref} = l_F m V_y^{ref} - Iz\dot{\psi}^{ref}$. The reference signals can be obtained from the high level path design, or in our case from the kinematic control. The angular velocities of the axels can be measured by odometers. All the signals are typically superposed with noises hence reliable filtering and differentiation are needed.

For this purpose Savitzky-Golay filters, fictitious control loops or algebraic estimation can be suggested [12]. From the later two typical methods are presented. Denote $y(t)$ the noisy function to be filtered or differentiated, T is the sampling time. The integration can be performed by the trapezoidal rule.

Filtering using integration:

$$\hat{y}(t) = \frac{2!}{T^2}\int_{t-T}^{t}[-3(t-\tau)+2T]\,y(\tau)d\tau \tag{47}$$

Numerical differentiation using integration:

$$\hat{\dot{y}}(t) = -\frac{3!}{T^3}\int_{t-T}^{t}[2(t-\tau)-T]\,y(\tau)d\tau \tag{48}$$

The steps of the *Flatness Control Algorithm (FCA)*:

1. Reading the signals V_x^{ref}, V_y^{ref}, $\dot{\psi}^{ref}$ from the kinematic control level (or path design) and the signals y_1, y_2, $\dot{y}_2$ from the dynamic system (or its model).
2. Computation of the signals $\hat{y}_1^{ref} = V_x^{ref}$, $\hat{\dot{y}}_1^{ref} = \hat{\dot{V}}_x^{ref}$, $\hat{y}_2^{ref} = l_F m\hat{V}_y^{ref} - I_z\hat{\dot{\psi}}^{ref}$, $\hat{\dot{y}}_2^{ref} = l_F m\hat{\dot{V}}_y^{ref} - I_z\hat{\ddot{\psi}}^{ref}$, $\hat{\ddot{y}}_2^{ref} = l_F m\hat{\ddot{V}}_y^{ref} - I_z\hat{\dddot{\psi}}^{ref}$.

3. Computation of the tracking errors $\hat{e}_{y_1} = \hat{y}_1^{ref} - y_1$, $\hat{e}_{y_2} = \hat{y}_2^{ref} - y_2$, $\hat{\dot{e}}_{y_2} = \hat{\dot{y}}_2^{ref} - \dot{y}_2$.
4. Determine the control signals by

$$u = \begin{bmatrix} T_\omega \\ \delta \end{bmatrix} = \tag{49}$$
$$= \Delta^{-1}(y_1, y_2, \dot{y}_2) \left(\begin{bmatrix} \dot{y}_1^{ref} + k_{1p} e_{y_1} + k_{1I} \int e_{y_1} dt \\ \ddot{y}_2^{ref} + k_{2p} e_{y_2} + k_{2I} \int e_{y_2} dt + k_{2D} \dot{e}_{y_2} \end{bmatrix} - \Phi(y_1, y_2, \dot{y}_2) \right)$$

5. Provide the control signals to the dynamic system.

4 NUMERICAL RESULTS

In the sequel the dynamic control part in the hierarchy will be limited to the nominal control and the DGA methods. Flatness control will be investigated in a separate future paper based on the here developed algorithm.

In order to have comparable results with the approach of Arogeti and Berman [2], we have chosen similar path and vehicle parameters in our experiments, namely $l_R = 1.35$, $l_F = 1.35$, $m_v = 1600$ and $I_{zz} = 2200$ (belonging to COG), all in SI units.

The road-tire relation was described by Pacejka's model:

$$F_{ti} = 2D_M \sin\{C_M \arctan[B_M \alpha_i - E_M \times (B_M \alpha_i - \arctan(B_M \alpha_i)))]\}, \quad i \in \{R, F\}$$

(α_i should be substituted in degree, not in radian). The parameters are $B_M = 0.239$, $C_M = 1.19$, $D_M = 3600\,[N]$ and $E_M = -0.678$. After linear approximation the cornering stiffnesses are $c_R = c_F = 1.1896 \cdot 10^5$ in N/rad.

The reference path is given by

$$y_d = 2\sin(0.25x) + 0.25x + 1, \quad x \in [0, 60]\text{m}$$

which is a straight line with additional slalom.

The velocity limits are $\eta_1 = 8.5$ m/s and $\eta_2 = 9.5$ m/s.

4.1 Checking the available data

For the LMIs the matrices C and D weighting the performance and the magnitude of the control and the disturbance weighting parameter λ were chosen in the referred paper by

$$C = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \\ 0 & 0 & 0 \end{bmatrix}, \quad D = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1/v(t) \end{bmatrix}$$

where $D(4)$ was corrected to make $\bar{C}$ constant which is necessary to the cited theory.

We have corrected the formula for the disturbance bounds and obtained that the Frobenius norm should be used. With its use the supremum of the Frobenius norm of $B_1(t)B_1^T(t)$ along the path is 1.8174 and its square root is 1.3446, so that its upper bound can be chosen as

$$\bar{B}_1 = 1.3446 \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

which is in good correspondence ($1.35I_3$) of the referred paper.

Using the above paper's result $1.35I_3$, the solutions of the two LMIs have also been determined:

$$Q = \begin{bmatrix} 12.9587 & -3.4709 & -0.4358 \\ -3.4709 & 2.1682 & -1.1092 \\ -0.4358 & -1.1092 & 1.4821 \end{bmatrix}$$
$$Y = \begin{bmatrix} 1.0548 & -0.9856 & -0.9252 \end{bmatrix}$$
$$K = \begin{bmatrix} -1.9357 & -6.7468 & -6.2429 \end{bmatrix}$$

The state feedback $K = \begin{bmatrix} -1.9 & -6.7 & -6.2 \end{bmatrix}$ is in good correspondence with the paper's solution which will be used later on.

4.2 The nonlinear dynamic model and the DGA control

The kinematic control and the dynamic control are running in different frames, and the dynamic modeling too. The velocity computation between them is as usual. Denote v_K the velocity vector at the origin of the kinematic frame (origin on the rear axle) and v_C the velocity of the origin of the dynamical frame (at the CoG), respectively. Using two dimensional vectors, the transformations between them are as follows:

$$\begin{aligned} v_C &= v_K + \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} l_R \dot{\psi} \\ a_C &= \dot{v}_C + \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \dot{\psi} v_C \\ v_F &= v_K + \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} L\dot{\psi} = v_C + \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} l_F \dot{\psi} \\ a_K &= \dot{v}_K + \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \dot{\psi} v_K \end{aligned} \tag{50}$$

From the velocities of the axles the slip angles can be determined and used to find the lateral forces by Pacejka's Magic Formula. For rear driving ($F_{xR} = F_{lR}$, $F_{yR} = F_{lR}$)

and front steering (which is assumed, $F_{xF} = F_{lF} = 0$, $F_{yF} = F_{tF}$) the dynamic motion equations for plane motion are given as

$$m_v a_C = \begin{bmatrix} C_{\delta_w} & -S_{\delta_w} \\ S_{\delta_w} & C_{\delta_w} \end{bmatrix} \begin{bmatrix} 0 \\ F_{yF} \end{bmatrix} + \begin{bmatrix} F_{xR} \\ F_{yR} \end{bmatrix}$$
$$I_{zz}\ddot{\psi} = l_F F_{yF} C_{\delta_w} - l_R F_{yR} \tag{51}$$

The differential equations (51) of the dynamic model and those consisting of (6) for the kinematic model are parallel running. From the differential equations the state equations can easily be formed.

For DGA control the higher order derivatives of the desired path by the time are needed. For this purpose the function `movingslope` is used which is available in MATLAB environment. This method of John D'Errico (woodchips@rochester.rr.com) uses filter to determine the slope of a curve stored as an equally spaced sequence of points. With 3 point window the method is similar to the derivation. It was used three times assuming constant speed 9 m/s along the prescribed path. Notice, that two path design is necessary, one for the K–frame origin and another for the C–frame origin.

The continuous time models and the DGA controller were discretized by Euler method with sampling time $T_S = 0.01$ sec and used in the realization.

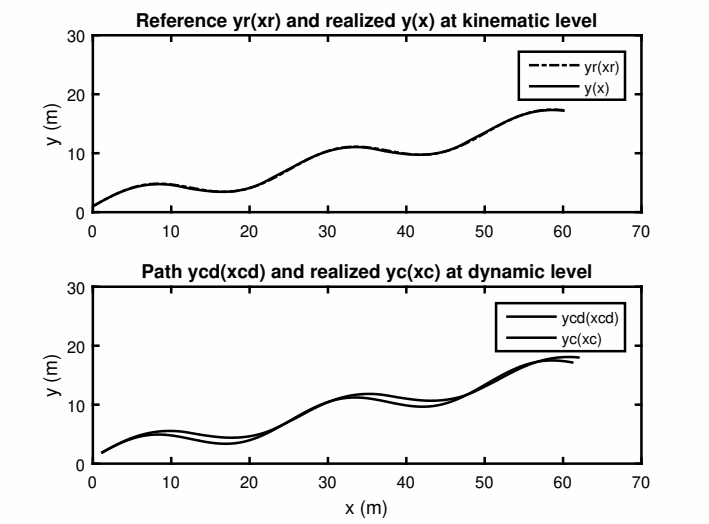


Figure 2
Kinematic and nominal control results

4.3 Simulation results

Simulation experiments were performed with high level modified kinematic control and low level nominal and DGA controls. Kinematic and dynamical controls are

running parallel, the latter supplies the kinematic control with the slipping angles.

The path design is performed in K-frame and the path is transformed to the C-frame using (50). The constant state feedback matrix $K = YQ^{-1}$ is computed offline from the solution of (12). The nominal dynamic control is based on (15). The DGA control was deeply described in the steps of the *DGA Control Algorithm*.

Denote $X_c = (v_B, x_c, y_c, \psi, \dot{\psi}, \delta_W)^T$ the state and $U_c = (F_{lR}, \dot{\delta}_w = w)^T$ the control of the C-frame. Let $X_k = (x, y, \psi, \delta_w)^T$ be the state and $U_k = (v_r, u, w = \dot{\delta}_w)^T$ the control of the K-frame, i.e. the middle point of the of the rear axle. The kinematic control is based on X_k and the slip angles $\alpha_R = \alpha_1$ and $\alpha_F = \alpha_2$ are determined from the dynamic control using Pacejka's magic formula. Non-measurable state variables can be determined in real time by the fusion of a common sensory system and state estimation for all the control methods. Since they are common for all the controllers, hence in the simulation they are emulated by the integration of the dynamics of X_c. Two position vectors can be computed for the CoG, p_{cc} from X_c and another p_{ck} from X_k.

The cycle of the simulation for one sampling instant T_s repeats the following steps:

1. Reading X_c and p_{ck}. Compute $v_R, v_F, \alpha_R, \alpha_F, \beta$ from X_c.
2. Reading X_k and p_{ck}. Compute the path $y_d(x) = f(x)$ and its derivatives $f'(x)$, $f''(x)$, $f'''(x)$.
3. Determine the kinematic error $e = (e_1, e_2, e_3)^T$.
4. Compute the kinematic control, i.e. choose v_r and use the state feedback matrix to compute the kinematic control $u = K * e * v_r$ and $w = \dot{\delta}_w$.
5. Integrate the kinematic state equation by computing the derivatives at the right side of the DE and using Euler method for the new X_k and p_{ck}.
6. Determine the transversal forces $F_R = F_1$ and $F_F = F_2$ using Pacejka's formula.
7. Perform numerical differentiations for the necessary variables and compute the new internal states needed to them.
8. Compute the control outputs F_{lR} and $w = \dot{\delta}_w$ for nominal control, and S_h, u_1, $u_2 = F_{lR}$ and $w = \dot{\delta}_w$ for the DGA algorithm, respectively.
9. Model the saturation of the control forces between $\pm$8000N.
10. Integrate the dynamic state equation by computing the derivatives at the right side of the DE and using Euler method for computing the new X_c and p_{cc}.
11. Storing the new states X_c, p_{cc} and the new control signals U_c for dynamic control, and the new states X_k, p_{ck} and the new control signals U_k for kinematic control.

Fig. 2 and Fig. 4 show that the kinematic level works well in the presence of acceptable rear and front slip angles, see also Fig. 3 and Fig. 5. The steering angles are in acceptable domains. From the lower part of Fig. 2 can be seen the main result

that if the steering angle is saved in the dynamic control (e.g. modeling an industrial controller with quick transients) than lateral errors of order 1m can be observed which may be critical in case of UGVs if no visual information for correction is available. This can be seen deeper in Fig. 6 for both dynamical control forms. For DGA it can also be seen that, although the nonlinear input–output linearization (DGA) can well stabilize the system, it cannot essentially decrease the lateral error.

5 CONCLUSION

In this paper the problem of the hierarchical control of UGVs was considered. High level kinematic control in the presence of sliding effects was analyzed using the modified kinematic control method of Arogeti and Berman. The non-published derivations of some important details were checked. In order to improve the precision of the path tracking low level nonlinear dynamic control methods were suggested. Novelties of our paper are:

i) Development of three low level techniques: nominal, DGA and flatness based dynamic control methods to supply the high level modified kinematic control with realistic front and rear sliding angles.

ii) The high level modified kinematic control method can well tolerate the sliding angles in the realistic domain of less than 15 degree and the path errors remain small at kinematic level.

iii) Using the nominal control it was experimentally proven the fact that if the steering angle of the kinematic control is used as the reference signal for the low level (e.g. industrial) steering control then this approach may cause problems for UGVs because the lateral error is in the order of 1m. This may be critical in the lack of visual information since no driver is present to make corrections.

iv) Simple methods, like nonlinear input–output linearization in the form of the DGA dynamic control with own reference signals (i.e. the path), can stabilize the vehicle but cannot considerably decrease the lateral error.

Further researches are necessary to develop new dynamic control methods that are able to decrease the path errors and are simple enough for real-time implementation. The flatness based control is one of the methods in this direction. A future paper will consider this investigation based on the here presented approach.

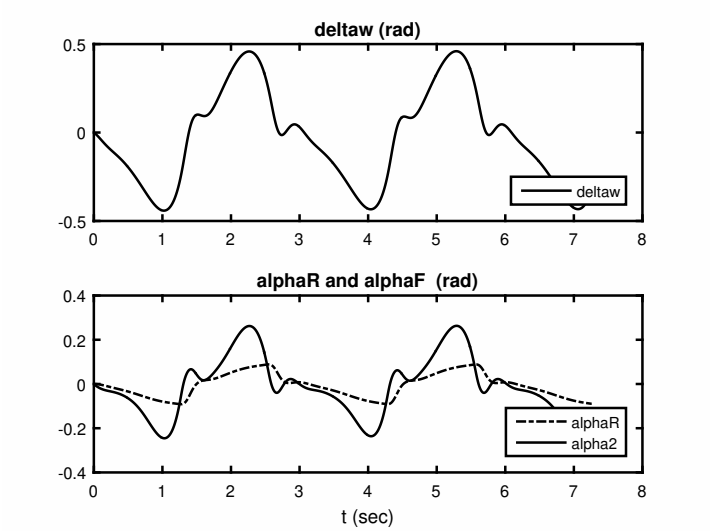


Figure 3
Steering and slipping angles with nominal control

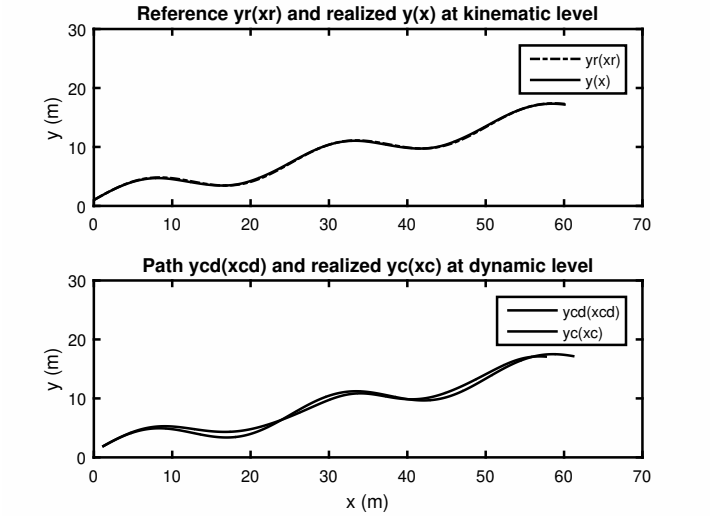


Figure 4
Kinematic and DGA control results

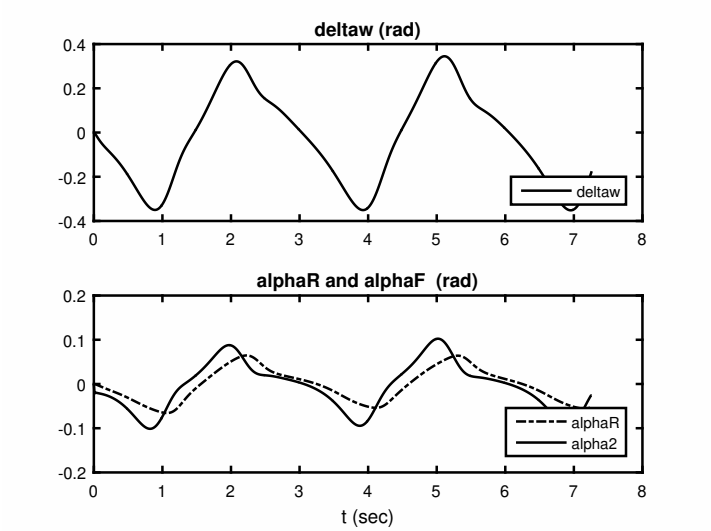


Figure 5
Steering and slipping angles with DGA control

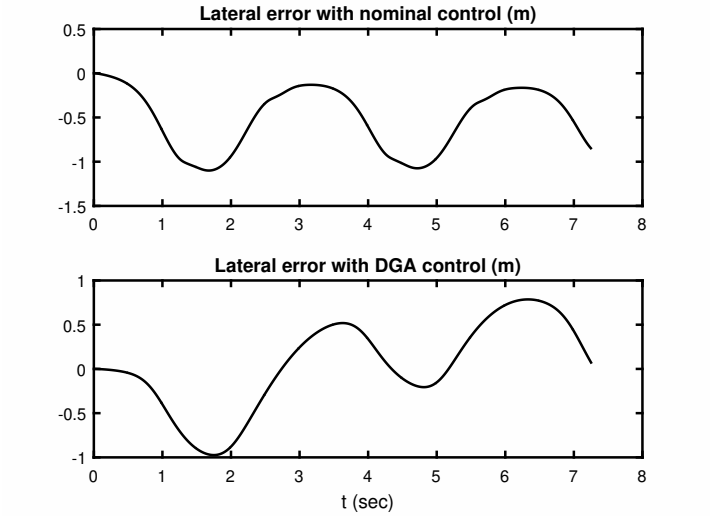


Figure 6
Lateral errors with nominal and DGA control

Acknowledgements

The research of B. Lantos was supported by the MTA-BME Control Engineering Research Group.

References

[1] A. D. Luca, G. Oriolo, and C. Samson, *Feedback control of a nonholonomic car-like robot*. J.P. Laumond, Ed. ser. Lecture Notes in Control and Information Sciences, Springer Verlag, New York, 1998.

[2] A. Arogeti and N. Berman, "Path following of autonomous vehicles in the presence of sliding effects." *IEEE Transaction on Vehicular Technology*, vol. 61, no. 4, pp. 1481–1492, 2012.

[3] C. Scherer and S. Weiland, *Lecture Notes DISC Course on Linear Matrix Inequalities in Control, ver. 2.0, pp. 50–57*, 1999.

[4] Z. Bodó and B. Lantos, "Error caused by kinematic control in dynamic behavior of unmanned ground vehicles," in *IEEE Symposium on Applied Computational Intelligence and Informatics SACI2018*, Timisoara, Romania, May 2018, pp. 1–6.

[5] R. Rajamani, *Vehicle dynamics and control*. Springer, New York, 2006.

[6] L. Talvala, K. Kritayakirana, and J. Gerdes, "Pushing the limits: From lane-keeping to autonomous racing." *Annu. Rev. Control*, vol. 35, no. 1, pp. 137–148, 2011.

[7] G. Max and B. Lantos, "Time optimal control of four-in–wheel-motors driven electric cars." *Periodica Polytechnica Electrical Engineering and Computer Science*, vol. 58, no. 4, pp. 149–159, 2014.

[8] H. Pacejka, *Tire and vehicle dynamics*. SAE International, 2005.

[9] B. Lantos and L. Marton, *Nonlinear control of vehicles and robots*. Springer, London, 2011.

[10] M. Fliess, J. Levine, P. Martin, and P. Rouchon, "Lie–Bäcklund approach to equivalence and flatness of nonlinear systems," *IEEE Transaction on Automatic Control*, vol. 44, no. 5, pp. 922–937, May 1999.

[11] L. Menhour, B. d'Andrea Novel, M. Fliess, and H. Mounier, "Coupled nonlinear vehicle control: Flatness-based setting with algebraic estimation techniques." *Control Engineering Practice*, vol. 22, 2014.

[12] E. Diekema and T. Koornwinder, "Differentiation by integration using orthogonal polynomials, a survey," *Journal of Approximation Theory*, vol. 164, 2012.

Techno-Economic Analysis of Several Energy Storage Options for Off-Grid Renewable Energy Systems

Juan Lata-García[1,2], Francisco Jurado[2], Luis M. Fernández-Ramírez[3], Pablo Parra[1],Víctor Larco[1]

[1]Department of Electrical Engineering GIPI, Universidad Politécnica Salesiana Guayaquil, Ecuador, jlatag@ups.edu.ec, pparra@ups.edu.ec, vlarco@ups.edu.ec

[2] Research Group in Research and Electrical Technology (PAIDI-TEP-152), Department of Electrical Engineering, EPS Linares, University of Jaén, 23700 Linares, Jaén, Spain, fjurado@ujaen.es

[3] Research Group in Electrical Technologies for Sustainable and Renewable Energy (PAIDI-TEP-023), Department of Electrical Engineering, EPS Algeciras, University of Cádiz, 11202 Algeciras, Cádiz, Spain, luis.fernandez@uca.es

Abstract: The increase in the production of energy through renewable energies and the reduced predictability of meteorological variables leads to consider reliable energy storage systems to ensure the constant supply of energy and the stabilization of the network. This paper studies four commercial energy storage systems for energy supply to an island by using photovoltaic panels and hydrokinetic turbine as primary energy sources. The technical characteristics of the energy storage systems are compared with respect to capacity shortage, autonomy storage and expected life storage, and the economic analysis is carried out. The system with li-ion batteries has 15 year expect life storage, the fuel cell based system has the highest storage autonomy 253 hr, and the pump hydro storage system has the lowest levelized cost of energy 0.178 $/kWh with 7 year expect life storage.

Keywords: energy storage; hybrid system; off-grid applications; renewable energy

1 Introduction

The development of renewable energy with marginal costs such as solar or wind energy is achieving increasing funding in liberalized energy markets. Latin America and the Caribbean have 173.25 GW of installed capacity in renewable energy, mainly from three primary sources, hydroelectric (159 GW), wind (12.3 GW) and solar (1.95 GW) developed in recent years [1], [2]. The sizing of a hybrid system and the adequate energy management from intermittent sources is a

complex task due to the poor predictability of the climate. Several researchers have proposed methodologies, mathematical algorithms and software applications to achieve optimal sizing [3]–[5].

Energy storage systems fulfill two main functions. The first function is to store the excess energy when the load does not require it and the second is to smooth the variations in the network as a result of the intermittence of the primary sources or the starting of some equipment. Said [6] presented an overview of different electrical energy storage technologies to reduce these fluctuations.

Currently, there are several types of energy storage commercially available, used in different projects. Among them are the hydraulic pumping, batteries, compressed air, super capacitor and flywheel [7]. Evans [8] reviewed energy storage technologies, comparing parameters such as efficiency, energy capacity, energy density, execution time, capital investment costs, response time, lifespan in years and cycles, and self-discharge. This work concluded that the choice of the storage system depends on the individual requirements, such as the capacity of the system and security in the delivery of energy.

Energy storage systems represent the highest cost in the implementation of hybrid systems. In [9], the authors presented the technical economic analysis of a hybrid system composed of solar panels, hydrokinetic turbine, fuel cell and battery acid lead. The hydrogen subsystem represented 45.5% of the total cost of the hybrid system, whereas the battery bank was 24.9%. Several successful models to solve optimization problems have been reviewed. A stochastic approximation algorithm (SA) was used in [10], which was based on a simultaneous perturbation gradient approximation instead of the standard Kiefer-Wolfowitz approximation. The numerical results obtained showed that the SA algorithm can be more efficient in problems of large dimensions.

An algorithm called Quantum Structure Analyzer 1.0 was presented in [11], whose operating principle was based on the Self-Organization Network by Kohonen. The algorithm was used to examine the grouping of nanostructures that are formed in a process of self-organization, technical parameters such as size, shape and spatial distribution in order to determine the characteristics of the devices. In addition, a grouping algorithm combined with a diffuse inference algorithm was used.

In data mining, the grouping of information is an important technique to create groups of similar objects within a grouping and different object between the different groupings. In [12], the authors presented an improved hybrid algorithm (IABCFCM) based on Fuzzy c-means (FCM), which used data grouping and an Artificial Bee Colony (ABC) algorithm. ABC is based on swarms inspired by the intelligent behavior of bees. The improved IABCFCM algorithm helped the fuzzy c-means cluster to escape from local optima and provide better experimental results in known data sets.

The problem of optimal sizing of storage systems was investigated in [13]. The authors considered two different storage systems (batteries and hydraulic storage pumped). The solution was achieved through optimization, using a non-linear programming approach in GAMS and considering several operational constraints such as efficiency, depth of discharge, response speed in order to obtain the minimum total cost of the system. It was experimentally tested and a comparative study was carried out. Jacob [14] developed a methodology for sizing a storage system using pinch analysis. The analysis was performed through a simulation of time series of the system, where the generation has to always exceed the load for an isolated hybrid system based on renewable energies. The methodology defined the design space as feasible combinations of storage in the short, medium and long term.

Young [15] studied the sizing of a battery energy storage system (BESS), and used financial, technical and hybrid indicators. The size of the subsystem can be determined by using methods and techniques, such as probabilistic methods, analytical methods, directed, search-based methods and hybrid methods. This work concluded that the proper method for sizing the battery bank is determined by the type of application and the system size.

Commercial programs can be used for the optimal sizing of hybrid systems. HOMER is a tool widely used by researchers. Through simulations, the number of components required to satisfy the demand is determined to achieve the minimum costs of the system and operation, and to comply with the technical limitations and emissions [16].

So far the research developed has focused on the optimized size of the storage systems for the balance between operation, support and cost, since it is one of the most expensive elements of the system. This work pretends to serve as a useful guidance for selecting energy storage systems for off-grid hybrid renewable energy systems. Furthermore, it provides a techno-economic analysis of four types of storage systems that have been chosen based on the following criteria: ability to provide energy for a long time; widely used technology; 3) fully renewable technology; and 4) new energy backup technique. Economic technical optimization is achieved through the HOMER software, which uses patented "derivative free" algorithm to identify the lowest energy cost for the hybrid system.

This paper is structured as follows. After the introduction, the different energy storage systems are reviewed in Section 2. Section 3 shows the mathematical model of the components of the system. Section 4 describes the different configurations simulated. The different configurations under study are compared in technical and economic terms in Section 5. Finally, the conclusions of the paper are drawn in Section 6.

2 Energy Storage Systems

In this section, the different types of energy storage systems are reviewed. The delivery of stored energy depends on the needs of the load. The different storage technologies are categorized into short, medium, and long-term storage. In Fig. 1, the duration of the download of the different storage systems in time scale is shown.

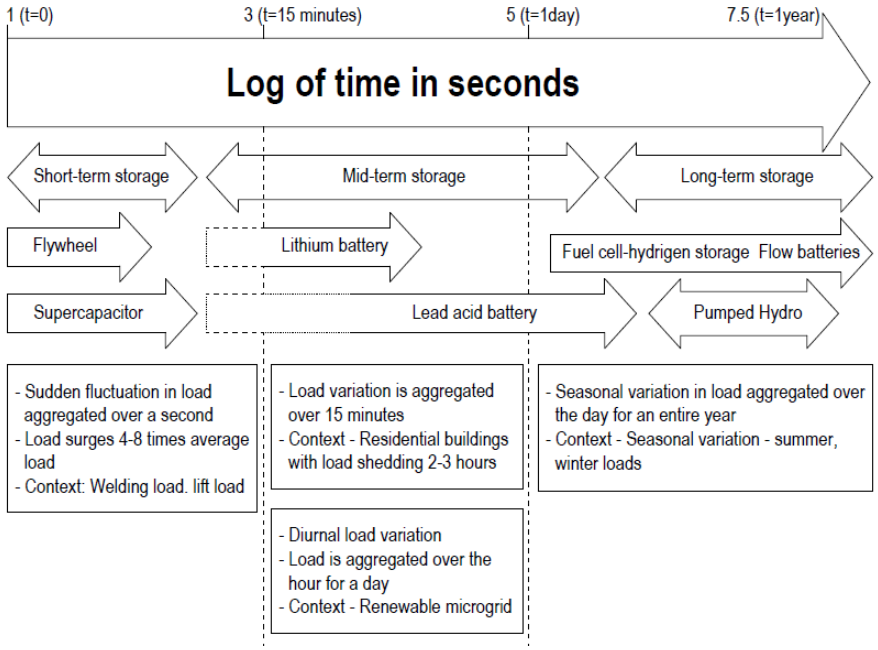


Figure 1
Energy storage technologies [14]

2.1 Flywheel Energy Storage

Flywheel energy storage represents an ecological option since it is made of non-hazardous base metals and carbon fibers. It uses a flywheel that rotates at very high speed. Inertia wheels store energy in its kinetic form [15]. Flywheels provide an amount of energy in a short time. The operational life is several million complete cycles of discharge, unlike batteries can reach up to 20 years of useable life [17]. The amount of energy stored is given by Equation (1).

$$E = \frac{1}{2} I \cdot \omega^2 \tag{1}$$

Where I is the inertia moment of the steering wheel and ω is the angular velocity.

2.2 Supercapacitor

Supercapacitor is composed of two electrodes in parallel with a dielectric material between them. It differs from the conventional capacitor because of its high capacity and discharge current [18]. Supercapacitors can be charged and discharged quickly. The long cycle of operation without degradation makes it attractive for use. They are mainly used to start motors, actuators and electric vehicles [19].

2.3 Lithium Battery

Lithium battery transforms the chemical energy of its metallic lithium composites into electrical energy through oxidation and reduction reactions. The advantages offered are: 100% efficiency, the highest achieved of all types of batteries; load densities are between 90-190 Wh/kg; and low environmental impact due to their recyclable components.

The main barrier in the extended use of this type of battery is the cost due to the special packaging and the internal overload protection circuit [17]. The electrical storage capacity of a Li-ion battery is limited by the amount of lithium that can be trapped at the battery anode. The lifetime is 3,000 cycles at 80% depth of discharge, the cost is 800-1100 $/kWh [20].

2.4 Lead Acid Battery

Lead-acid batteries are widely used in renewable energy systems and other applications. They are rechargeable batteries and are composed of an anode of metallic lead, a cathode of lead dioxide and an electrolyte of sulfuric acid solution. Among the advantages that they present are: the cost under 500-150 $/kWh, simplicity in the manufacture, fast electrochemical reaction, and efficiency of 72%. The main disadvantage is the use of heavy metal components, dangerous for the environment. The lifetime is limited. Several types of lead-acid batteries have been developed, such as lead antimony batteries, SLI (starting, lighting and ignition) batteries, valve-regulated lead acid batteries (VRLA), flooded lead acid batteries, lead and calcium batteries, AGM absorbed glass mats batteries, gel cells and deep cycle batteries. The last two are commonly used in renewable energy systems [21].

2.5 Electrolyzer and Fuel Cell

Hydrogen energy storage system is a clean technology. It is an acceptable option for remote communities that are not connected to a conventional network. There is currently a wide variety of commercially available fuel cells (FC). Proton exchange membrane(PEM) type is used for small-scale hybrid systems [22]. The

excess energy produced by the generators feeds the electrolyzer, which is a device that generates hydrogen and oxygen from water. The produced hydrogen is stored in a tank or taken directly to the FC.

2.6 Pumped Hydro Storage

Pumped hydro storage (PHS) is a mature and commercially available technology. It stores and generates electricity through two water tanks at different heights, and energy produced in excess feeds a pump that sends water from a lower reservoir to an upper reservoir [23].

When the demand is greater than the generation, the water stored in the upper tank passes through a turbine that produces electrical energy. In some cases, pumping systems are based on reversible electric and hydraulic machines. The energy used to pump water from the lower reservoir is obtained by Equation (2).

$$E_{pumping} = \frac{\rho \,.\, g \,.\, h \,.\, V}{\eta_p} \tag{2}$$

where V is the volume of water, h is the height of the upper reservoir, and η_p is the pumping efficiency.

The energy supplied by the generator to the network can be obtained by Equation (3).

$$E_{generator} = \rho \,.\, g \,.\, h \,.\, V \,.\, \eta_g \tag{3}$$

where η_g is the generator efficiency.

3 Mathematical Model of System under Study

The hybrid systems under study in this work use the following components: photovoltaic (PV) panels, hydrokinetic (HKT) turbine, batteries, hydrogen system (FC, FC, electrolyzer and tank), pumped hydro storage and electronic power converters. This section describes the mathematical models used to represent the behavior of these components.

3.1 PV Panel

Santay Island receives around 12 hours of sun a day with a high global radiation so that PV energy is a promising source to be used for electrification purposes.

For the modeling of the PV panels, polycrystalline type models CS6K-285M-FG are commercially available, which present the following characteristics: a power of 285 W; an efficiency in standard test conditions (STC) of 17.41%; the open circuit voltage (V_{oc}) is 38.6 V_{dc}; the short circuit current (I_{sc}) is 9.51 A; the

dimensions (mm) are 1,650-992-40 mm; the capital cost is US$ 300/285 Wp; the replacement cost is US$ 300/285 Wp; the operating and maintenance cost $0 for not having moving parts; and the lifetime is 25 years.

The estimation of the maximum power of the PV generator is an important parameter to consider in the system. The maximum output power of the PV module can be calculated under operating conditions as follows:

$$P_{PV} = f_{PV} * Y_{PV} * \frac{I_T}{I_S} \quad (4)$$

where f_{PV} is the reduction factor, which takes into account the losses due to high temperatures, dirt, wiring, that reduce the performance of the panel; Y_{PV} is the rated capacity of the PV array (kW); I_T is the total incident radiation on the panel surface (kWh/m²); and I_s is 1000 W/m².

3.2 HKT Turbine

The energy from the flow of water that surrounds the island can be exploited by a HKT turbine to be converted into electrical energy. In this work, a 5 kW smart monofloat turbine from the manufacturer Smart hydro power was used [24], which present the following characteristics: the dimension are 2640-1120-1120 mm; the rotational speed is 90 to 230 rpm; the weight is 380 kg; the number of rotor blades is 3; the cost of capital is $ 11179; and the cost of operation and maintenance is 100 $/year. The energy generated by the HKT turbine (E_{HKT}) can be calculated as:

$$E_{HKT} = \frac{1}{2} * \rho_W * A * v^3 * C_{p,H} * \eta_{HKT} * t \quad (5)$$

where $\rho_W$is the water density in kg/m³; η_{HKT} is the combined HKT-generator efficiency; $C_{p,H}$is the HKT performance coefficient; v is the water flow velocity in m/s; A is the HKT area in m²; and t is the time in s.

3.3 Batteries

Battery consists of electrochemical cells that can convert chemical energy into electrical energy. The energy produced in excess can be stored in the batteries, which is used when there is a peak consumption that cannot be supplied by the renewable generation.

The state-of-charge (SOC) of the battery for the charging and discharging modes can be calculated by [25]:

$$SOC(t) = SOC(t-1) + \frac{E_{bat}(t) * \eta_{cbat}}{P_{bat}} * 100 \quad (6)$$

$$SOC(t) = SOC(t-1) + \frac{E_{bat}(t) * \eta_{dbat}}{P_{bat}} * 100 \quad (7)$$

where $SOC(t)$ is the state of charge at time *t*; P_{bat} is the nominal capacity of the battery; E_{bat} is the power exchange during the time step Δt; and $\eta_{cbat,}\eta_{dbat}$ are the battery charge and discharge efficiency, respectively.

3.4 FC

In this work, a PEM FC is considered for the hydrogen energy storage system. FC uses hydrogen to produce DC electrical energy. PEM FC is one of the best options for distributed generation, since it has advantages such as its low operating temperature, high-power density, longevity, specific power [26].

Different works indicate an initial capital cost of the FC system from $2,000/kW to $6.000/kW. The installation, operational costs and replacement are estimated depending on the power of the system at $3,000, $0.080/h and $2,500 respectively [9].

The electric power output from the FC, P_{fc}, can be calculated as follow [27]:

$$P_{fc} = U_{Stack} \times I = U_{SC} \times N_{CELLS} \times I \tag{8}$$

The hydrogen consumption of the FC in one hour can be calculated from the FC power as follows:

$$q_{H_2}^{con} = \frac{P_{fc}}{E_{low,H2} * \eta_{therm} * U_f * \eta_{fc}} \tag{9}$$

where P_{fc} is the output power supplied by the FC; $U_{SC}$is the average single cell voltage; I is the current; $E_{low,H2}$ is the lower heating value of hydrogen ($E_{low,H2}$ =33.35 kWh/kg); η_{therm} is the thermodynamic efficiency (0.98 at 298 K); U_f is the fuel utilization efficiency and η_{fc} is the FC efficiency.

3.5 Electrolyzer

An electrolyzer converts DC electrical energy into chemical energy stored in hydrogen. An electric current drives the decomposition of water, which is fed to the anode, into hydrogen and oxygen. The occurring reactions in a PEM electrolyzer are:

$$\text{anode} \quad H_2O \rightarrow 2H^+ + 1/2O_2 + 2e^-$$

$$\text{cathode} \quad 2H^+ + 2e^- \rightarrow H_2$$

The advantages of PEM electrolyzer are: high voltage efficiency, high current densities, good partial load range, high gas purity, rapid system response. PEM electrolyzers have a conversion efficiency of 40 to 60% similar to alkaline electrolyzers, due to the membrane and noble metals the PEM are more expensive, although promising studies consider an 85% efficiency [28].

The electrical power consumed by the electrolyzer can be calculated as follow:

$$P_E = \frac{\dot{m}_{H_2} \times HHV_{H_2}}{\eta_E} \quad (10)$$

where, P_E is the power consumption of the element; $\eta_E$is the electrolyzer efficiency;$\dot{m}_{H_2}$ is the produced hydrogen mass flow rate (kg/s); andHHV_{H_2} is the gross calorific value (MJ/kg).

3.5 Hydrogen Storage Tank

The hydrogen generated by the electrolyzer needs to be stored in a tank to supply the FC. The price of a tank with 1 kg of hydrogen capacity is estimated at $1,500. The replacement costs are estimated at $1,500 per kg similar to studies performed. The useful life of the hydrogen tank is estimated at 25 years [9].The tank autonomy can be calculated using the following equation [29]:

$$A_{htank} = \frac{Y_{htank} \times LHV_{H_2}(24h/d)}{L_{prim,ave}(3.6MJ/kWh)} \quad (11)$$

where Y_{htank} is the capacity of the hydrogen tank (kg);LHV_{H_2} is the energy content of hydrogen; and$L_{prim,ave}$ is the average primary load (kWh/kg).

3.5 Pumped Hydro

Pumped hydro system allows storing water in a tank at a certain height when there is excess energy. It converts the potential energy into electricity by a turbine when the energy generation is required.

In this work a generic 245 kWh pumped hydro system was used. It presents a nominal capacity of. 1,059.000 Ah.

The energy storage capacity can be calculated as follow [29]:

$$E[J] = 9.81 P_{water} \times V_{res} \times h_{head} \times \eta \quad (12)$$

where E is the energy stored (J); P_{water} is the density of water (1000 kg/m3);V_{res} is the volume of the reservoir (m^3); h_{head} is the head height in meters; and η is the efficiency.

3.6 Electronic Power Converters

All configurations in this work need DC/AC converters for the connection to the AC network. The connection between the DC bus and the loads is made by a DC/AC converter. The converters must respond to several challenges, in terms of voltage ratio, energy efficiency, output current ripple, cost and reliability in case of power switch fault-tolerance. For this work, a range of inverter (0-20 kW) with efficiency of 95% and a useful life of 15 years is considered. The installation cost

is $300/kW and the replacement cost is equal to the initial capital. The cost of O&M is zero for inverter [9].

4 Studied Configurations

The hybrid systems under study in this work are designed to satisfy the demand of the inhabitants of Santay Island. The island is located in the Gulf of Guayaquil, 800 m from the city of Guayaquil, with an area of 21.79 km². This island has a population of 56 families living in 46 houses. The consumption load is characterized by the habits of the residents of the coastal area of Ecuador [30] with a minimum consumption (3.1 kW), average consumption (4.18 kW) and maximum consumption (5.6 kW). The renewable energy resources (radiation solar and river speed) are studied below.

4.1 Renewable Energy Resources

The abundant solar radiation (2°12.5'S, 79°52.2'W) is shown in Figure 2. The data were obtained through the HOMER software. The annual average is 4,630 kWh/m² /day [29].

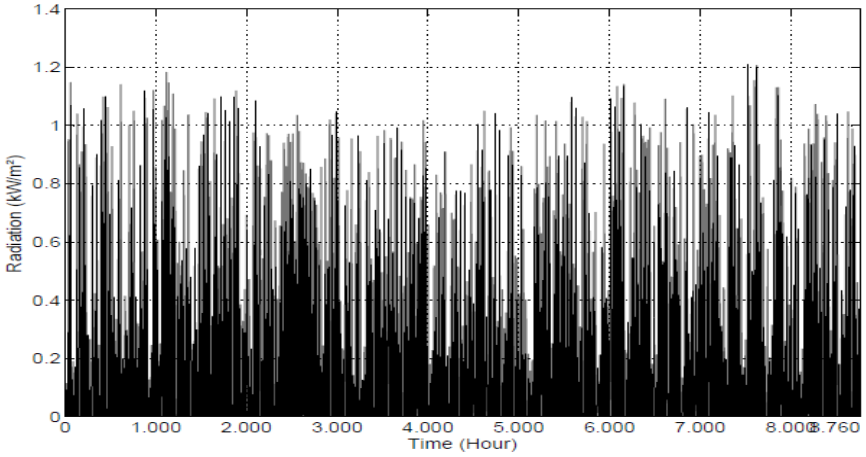


Figure 2
Radiation solar annual

The kinetic energy of the river flow is exploited. The annual river velocity data is shown in Figure 3. These data are provided by the Oceanographic Institute of the Ecuadorian Navy (INOCAR) [31]. The highest recorded speed is 2.26 m/s, the slow speed is 0.31m/s and the average is 1.39 m/s. To carry out the simulation, the monthly average values are taken.

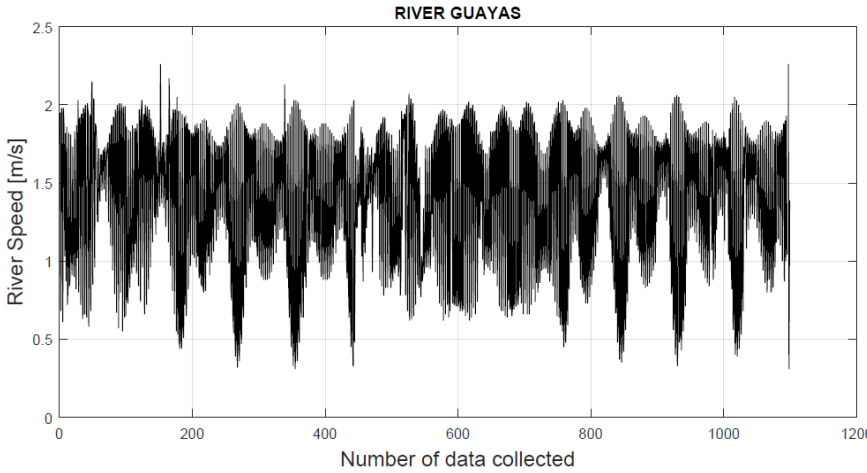


Figure 3
River speed

4.2 Systems under Study

Four hybrid systems have been studied. In all of them, PV panels and HKT are used as primary energy sources, the ESS with little capacity to provide energy for long time are not considered in the study. The studied configurations have been chosen based on widely used energy storage: 1) lead acid and 2) lithium batteries commonly used; 3) hydrogen system (electrolyzer, hydrogen tank and FC) is a totally renewable technology that is being used despite its high costs; 4) pumped hydro is a promising technology with a high application capacity.

The optimal configuration of each hybrid system under study is obtained by HOMER. The program carries out hundreds or even thousands of configurations to achieve the optimum sizing of the hybrid system. Each system is evaluated by four parameters: 1) lower initial capital; 2) total net present cost, which is the present value of all system costs over the useful life; 3) cost of energy, which is the average cost per kWh produced by the system; and 4) maximum capacity shortage or maximum allowable value of the capacity shortage fraction.

4.2.1 PV, HKT, Lithium Battery-based Hybrid System

Figure 4 shows the hybrid system under study. The resulting optimal sizing is composed of a PV array (42 kW), 3 HKT turbines (15 kW), a 1 kWh Li-Ion battery bank (50 units) and a DC/AC power converter (10 kW).

Table 1 shows the technical characteristics of the system. The simulation provided the following results: the initial capital of the system is $48,400, TNPC is $96,804, LCOE is 0.178 $/kWh, the capacity shortage is 0.29% and the electrical

load that the system is unable to serve (Unmed Electric Load) is 72.2 kWh/yr (0.20%) of total load.

The model used is a generic 1 kWh li-ion battery, which has a nominal voltage of 3.7 V, a nominal capacity of 1.02 kWh, a maximum capacity of 276 Ah, a capacity ratio of 1, a rate constant of 1 (1/h), 8% in other round –trip losses, and the effective series resistance is 0.00036 ohms.

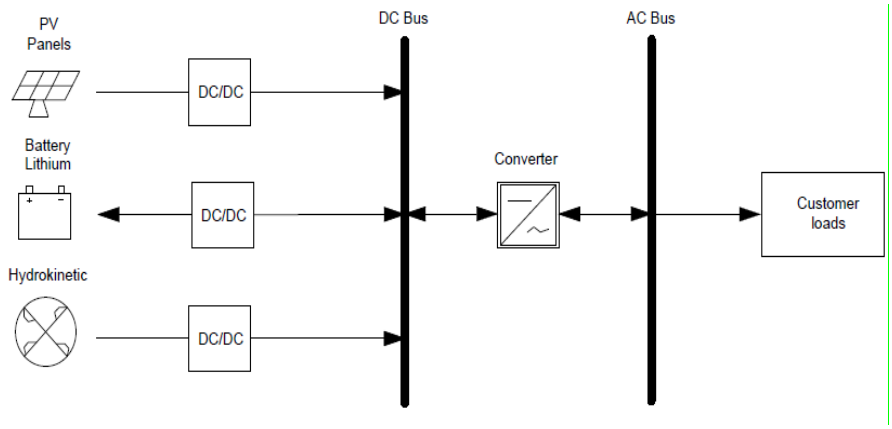


Figure 4
Configuration of the PV, HKT, Lithium Battery-based hybrid system

Table 1
Technical characteristics of the PV, HKT, Lithium Battery-based hybrid system

Components	PV	HKT	Converter
Rated Capacity (kW)	42	15	10
Mean Output (kW)	6.53	3.38	3.95
Maximum Output (kW)	34.5	5.98	8.54
Capacity Factor (%)	15.6	22.5	39.5
Total Production (kWh/yr)	57,236	29,547	-----
PV Penetration (%)	165	85.4	-----
Hours of Operation	4,189	8,760	8760

The per year operating cost is $2,435, the levelized costs for the PV and HKT are $0.0149 and $0.065 per $/kWh, respectively. The energy stored by the batteries is 6,592 kWh/yr, the energy consumed out is 6,063 kWh/yr, the storage depletion (the difference between the battery state of charge at the start of the year and at the end of the year) is 6.28 kWh/yr and the losses are 565 kWh/yr, the lifetime throughput 94,587 kWh and expected life 15 yr.

The components that have a higher cost of capital are batteries and PV panels, with $30,000 and $13,440 respectively. The component with the highest replacement cost is the battery bank with $16,847. The cost per operation and

maintenance of the battery is $7,876. The total cost of the system is $96,803. It is composed of the following items: capital $58,440, replacement $32,260, O&M $12,601, salvage $6,498.

4.2.2 PV, HKT, Lead Acid Battery-based Hybrid System

Figure 5 shows the proposed hybrid system, which is formed by a PVarray (45 kW), 3 units of HKT turbines (15 kW), a battery bank of 55 units (1 kWh) and a DC/AC converter (10 kW). The energy storage system is composed by a generic 1 kWh lead-acid battery, which has a nominal voltage of 12 V, a maximum charge current of 16.7 A, a capacity ratio of 0.403, a rate constant of 0.827 16 1/h and a maximum capacity of 83.4 Ah.

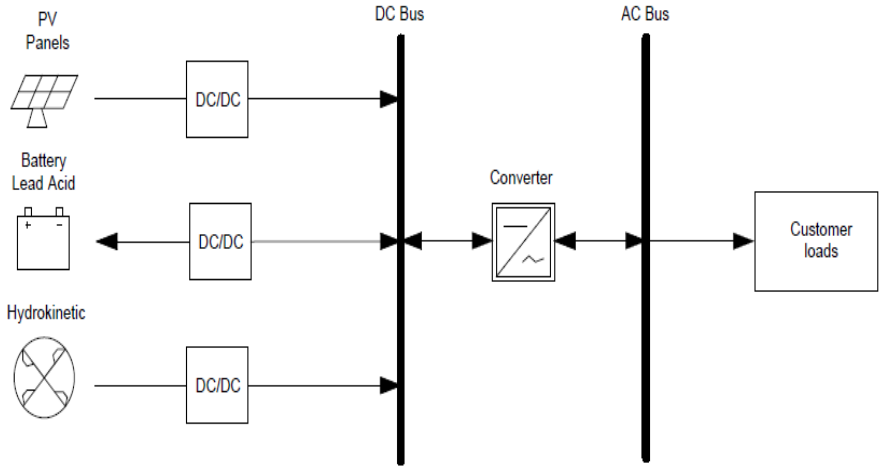


Figure 5
Configuration of the PV, HKT, Lithium Battery-based hybrid system

The initial capital cost of the system under study is $51,900, the TNPC is $106,854, the LCOE is 0.196 $/kWh, the capacity shortage is 0.232% and the unmed electric load is 57.7 kWh/yr (0.16%) of total load. The per year operating cost is $3,489, the levelized costs for the PV and HKT are $0.0149 and $0.0605 per $/kWh, respectively.

The excess energy stored by the batteries is 7,530 kWh/yr, the output energy is 6,029 kWh/yr, the storage depletion is 5.97 kWh/yr and the losses is 1,507 kWh/yr, the lifetime throughput 60,000 kWh and the expected life 8.90 yr.

The components that have a higher capital cost are the batteries and PV panels with $22,500 and $14,400 respectively. The component with the highest replacement cost is the battery bank with $27,321, the same happens with the cost per operation and maintenance, $11,814. The total cost of the system is $106,853 (capital $51,900, replacement $42,734, O&M $16,540, salvage $4,320).

The optimization results given by HOMER according to the specifications are shown in Table 2.

Table 2
Technical characteristics of the PV, HKT, lead acid battery-based hybrid system

Components	PV	HKT	Converter
Rated Capacity (kW)	45	15	10
Mean Output (kW)	7	3.38	3.95
Maximum Output (kW)	37	5.98	8.54
Capacity Factor (%)	15.6	22.5	39.5
Total Production (kWh/yr)	61,324	29,547	-----
PV Penetration (%)	177	85.4	-----
Hours of Operation	4,189	8,760	8760

4.2.3 PV, HKT, Hydrogen-based Hybrid System

The architecture of the proposed system is presented in Figure 6, where the system is composed of a PV array (40 kW), HKT turbines (15 W), a FC (6 kW), an electrolyzer (10 kW), a hydrogen tank (30 kg) and a DC/AC power converter (10 kW). The optimization results given by HOMER according to the specifications are shown in Table 3.

The simulation provided the following results: the initial capital of the system is $110,800, the TNPC is $189,640, the LCOE is 0.348 $/kWh, the operating cost is $5,005, the capacity shortage is 0.27% and the unmed electric load is 62.6 kWh/yr (0.18%) of total load.

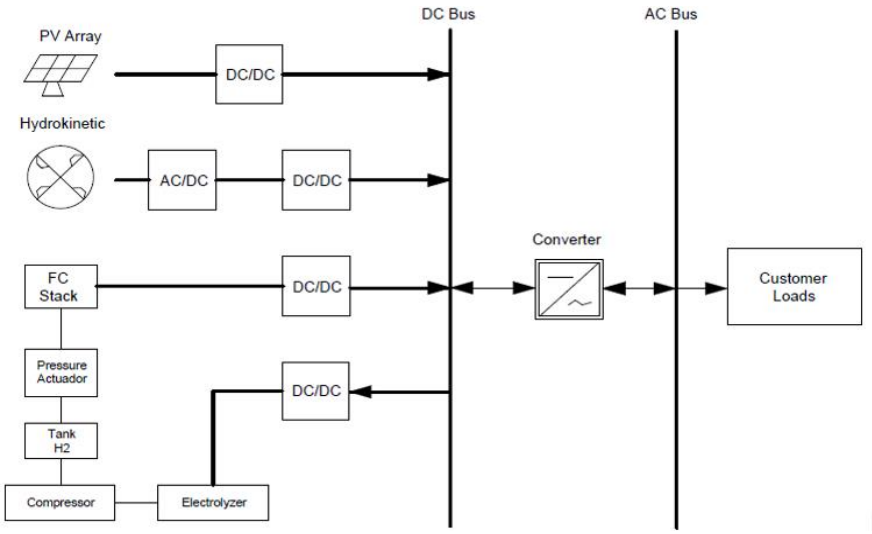


Figure 6
Configuration of the PV, HKT, hydrogen-based hybrid system

The levelized costs for the PV and HKT are $0.0149 and $0.0605 per $/kWh, respectively. The hydrogen subsystem presents the following results: electrical production 6058 kWh/yr, hours of operation 3,805 hrs/yr, numbers of starts 623 starts/yr, fixed generation cost 1.05 $/hr, fuel consumption 363 kg and operational life 105 yr.

In the third system under study, the capital costs of hydrogen tank, electrolyzer and FC are the highest $ 45,000, $20,000 and $18,000 respectively. The FC has the highest replacement cost and O&M with $20,030 and $35,962, respectively.

The HKT has the second highest cost of O&M with $4,725, while the cost of O&M for the hydrogen tank and PV is zero due to having no moving parts or consumables. The total cost of the system is $189,639, which consists of the following items: capital $110,800, replacement $43,867, O&M $43,838, salvage $8,866.

Table 3
Technical characteristics of the PV, HKT, hydrogen-based hybrid system

Components	PV	HKT	FC	Electrolyzer	Converter
Rated Capacity (kW)	40	15	6	10	10
Mean Output (kW)	6.22	3.38	1.59	2.07	3.95
Maximum Output (kW)	32.9	5.98	6	0.215 (kg/hr)	8.54
Capacity Factor (%)	15.6	22.5	11.1	19.5	38.9
Production (kWh/yr)	54,510	29,572	6,058	369 (kg/yr)	-----
Penetration (%)	157	85.4	-----	-----	-----
Operation (hr/yr)	4,189	8,760	3,805	3,308	8760

4.2.4 PV, HKT, Pumped Hydro Storage-based Hybrid System

The pumped hydro has a reservoir that can store a capacity of 1000 m^3 of water, which can be discharged over 12 hours. The system under study is shown in Figure 7. The resulting components are a PV array (30 kW), HKT turbines (15 kW), a pump hydro (245 kWh) and a DC/AC power converter (10 kW). The pumped hydro has a nominal voltage of 240 V, a maximum charge current of 91.6 A, a maximum discharge current of 91.6 A and the generator efficiency is 90%.

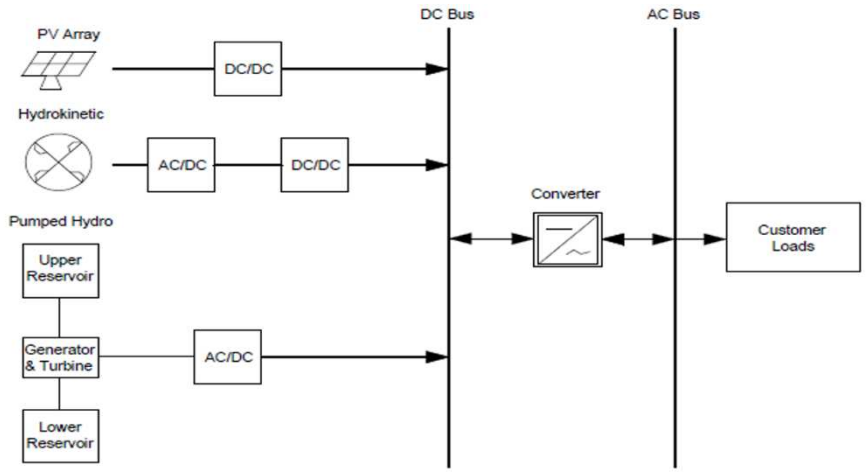


Figure 7

Configuration of the PV, HKT, pumped hydro storage-based hybrid system

Table 4 shows the parameters of system under study. The initial capital cost of the system is $42,600, the TNPC is $86,996, the LCOE is $0.159/kWh, the capacity shortage is 0.04% and the unmed electric load is 10.4 kWh/yr (0.02%) of total load. The per year operating cost is $2,818, the levelized costs for the PV and HKT are $0.0149 and $0.0605 per $/kWh, respectively.

The excess energy stored by pumped hydro is 12,540 kWh/yr, the output energy is 10,178 kWh/yr, the storage depletion is 23.5 kWh/yr, the losses is 2,385 kWh/yr, the lifetime throughput 79,164 kWh and the expected life is 7 yr.

The components that have a higher capital cost are the pumped hydro and the PV panels with $22,000 and $9,600, respectively. The component with the highest replacement cost is the HKT turbine with $9,152, while the largest cost per operation and maintenance is the pumped hydro with $31,504. The total cost of the system is $86,995 (capital $42,600, replacement $1,733, O&M $34,655, salvage $1,993).

Table 4

Technical characteristics of the PV, HKT, pumped hydro storage-based hybrid system

Components	PV	HKT	Converter
Rated Capacity (kW)	30	15	10
Mean Output (kW)	4.67	3.38	3.95
Maximum Output (kW)	24.7	5.98	8.54
Capacity Factor (%)	15.6	22.5	39.5
Total Production (kWh/yr)	40,883	29,572	-----
PV Penetration (%)	118	85.4	-----
Hours of Operation	4,189	8,760	8760

5 Comparative Study

5.1 Economic Parameters

The economic feasibility of the proposed system is performed taking into account five parameters: 1) initial capital cost, which is the total installed cost of the components at the beginning of the project; 2) TNPC; 3) capacity shortage; 4) levelized cost of energy (LCOE); and 5) annualized maintenance, operating and replacement cost. With these parameters, HOMER determines the optimal system. Each of the parameters and the equations for its calculation are presented in this work.

5.2 TNPC

The TNPC of a system during its useful life is defined as the present value of all costs incurred throughout life, less the present value of all the income obtained during its operation. It incorporates operating and maintenance (O&M) costs, fuel costs, emission penalties and component replacement costs, minus the cost of rescuing each component. TNPC is the main economic indicator of HOMER [29].

The TNPC can be calculated as follows:

$$C_{TPNC} = \frac{C_{ann,tot}}{CRF\left(i,R_{proj}\right)} \tag{13}$$

$$CRF_{(i,N)} = \frac{i(1+i)^N}{(1+i)^N-1} \tag{14}$$

where $C_{ann,tot}$ is the total annual cost (\$/yr); CRF is the capital recovery factor; I is the interest rate (%); and R_{proj} is the project lifetime (N).

5.3 Capacity Shortage

The maximum annual capacity shortage is the maximum allowable value of the capacity shortage fraction, which is the total capacity shortage divided by the total electric load. HOMER considers infeasible (or unacceptable) any system with a higher value of the capacity shortage fraction. Allowing some capacity shortage can change the results dramatically in some cases. This might happen if a very high peak occurs for a very short time [32].

5.4 Levelized Cost of Energy (LCOE)

The levelized cost of energy (LCOE) is defined as the average cost per kWh of useful electrical energy produced by the system. This provides a way to compare the total energy cost for each specific electrification scenario that is being considered. The result is achieved dividing the annualized cost of electricity

produced by the total electric load served. The LCOE is calculated by equation (15).

$$LCOE = \frac{LCC}{\sum_{j=1}^{N} \cdot \left(\frac{E_{GEN}(j)}{(1+d)^j}\right)} \quad (15)$$

where *LCC* is the life cycle cost; and E_{GEN} represents the cost of the life cycle of the energy produced by the proposed system in the given year *j* [29].

5.5 Annualized Maintenance, Operating and Replacement Cost

The annual cost of operating and maintenance increases with the operating time of the system. The cost can be determined by equation (16), while the annualized cost is determined by equation (17).

$$\boldsymbol{C_{OP}(j) = C_O(j) + C_m(j) + C_r(j)} \quad (16)$$

$$\boldsymbol{AV_{op} = \left(\frac{d(1+d)^j}{(1+d)^j - 1}\right) \cdot \left(\sum_{j=1}^{N} \frac{C_{op}(j)}{(1+d)^n}\right)} \quad (17)$$

whereC_O isthe operational cost of any component of the energy system in the year *j*; C_m is the maintenance cost of any component of the energy system incurred in the year *j*; and C_r is the replacement cost for any system component in the year *j*.

5.6 Results and Discussion

Table 5 shows the results of all the hybrid systems under study obtained from HOMER. The comparative study is performed by analyzing the differences in economic parameters and technical parameters(capacity shortage, renewable production, autonomy storage, annual throughput storage and expected life storage) of each of the systems under study[29].

The PV/HKT/pump hydro configuration achieves the best results in the LCOE with 0.159 $/kWh, the TNPC with $86,996, the operating cost with $2,818, the initial capital with $42,600. The capacity shortage (%) is only 0.04% with 14.3 kWh/yr, which ensures the constant supply of energy, and the autonomy storage is 64.3 hours. This configuration presents the smallest PV system (30 kW), and the production of renewable energy is 60,597 kWh.

Table 5
Economic results of simulation

Systems	PV/HKT/BLITIO	PV/HKT/B LEAD ACID	PV/HKT/FC	PV/HKT/PUMP HYDRO
LCOE ($)	0.17	0.19	0.34	0.15
TNPC ($)	96,804	106,854	189,640	86,996

Operating cost ($)	2,435	3,489	5,005	2,818
Initial capital ($)	58,440	51,900	110,800	42,600
Capital Cost PV ($)	13,440	14,400	12,800	9,600
Capacity Shortage (%)	0.29	0.23	0.27	0.04
Autonomy Storage (h)	9	11.4	253	64.3
Annual Throughput Storage (kWh)	6,285	6,741	-----	11,309
Expected life Storage (yr)	15	8.9	10.5	7

The annual discharge of the energy storage systems under study is illustrated in Figure 9.

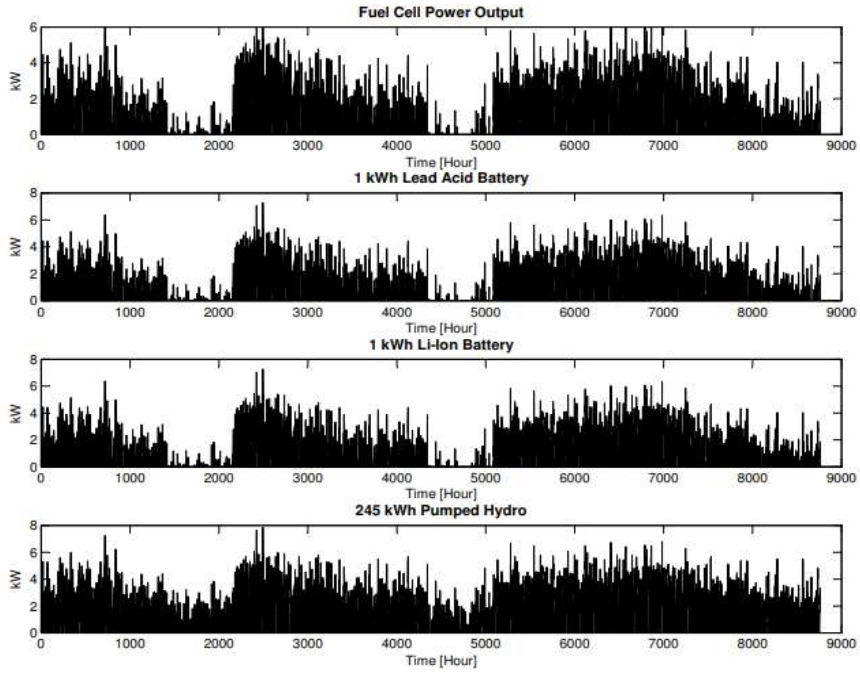


Figure 9

Annual discharge of the energy storage systems

The maximum discharge values of the FC, lead acid battery, lithium battery and pumped hydro are 6 kW, 7.26 kW, 7.26 kW and 7.84 kW, respectively. The total power provided by each of the energy storage system for a year is 6,056 kW (FC), 6,029 kW (lead acid), 6,033 kW (li-ion), and 10,178 kW (pumped hydro).

Figure 10 shows the state of charge of the three energy storage system. The battery of lead acid has a discharge of 40% which promotes the increase of the system useful life. The maximum discharge is 22.5% for the battery Li-Ion, while

the pumped hydro system delivers all the energy stored in the tank at the end of January due to the lower renewable generation.

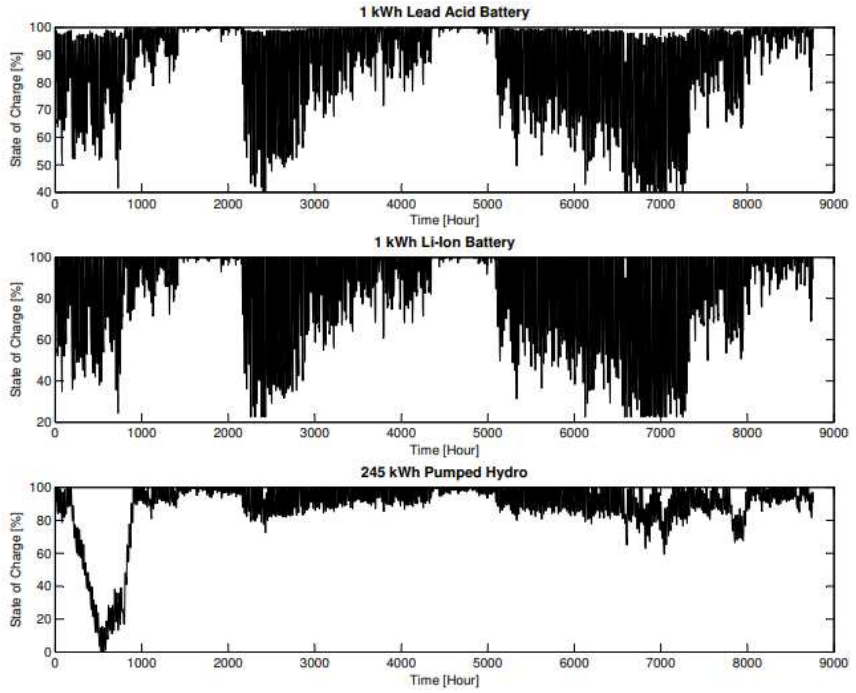


Figure 10

State of charge of the energy storage systems

Conclusions

This paper has presented a study of different energy storage technologies applied to a hybrid system composed of PV generator and HKT turbine. The results showed that all the systems analyzed satisfied over 99.8% of the energy required by the demand.

The system based of li-ion battery had the second lowest LCOE (0.178 $/kWh), and the TNPC was $96,804. The advantage of this configuration is a battery life with 15 years, whereas the disadvantage was the low autonomy (9 hr) of the stored energy.

The system based on lead acid battery had the third highest cost in LCOE (0.196 $/kWh). The NPC was $106,854, the autonomy of the system was 11.4 hr, but the expected life of the storage was reduced to 8.9 yr.

The system with hydrogen-based energy storage presented the highest TNPC ($189,639). However, it is expected that its implementation costs will be significantly reduced in the coming years and it is a totally ecological technology.

The system based on pumped hydro obtained the best results in the unmet electric load (10.4 kWh/yr), TNPC ($86,995) and LCOE (0.159 $/kWh). However, it presented the lowest expected life of the storage (7 yr).

References

[1] World energy, “Latin America & The Caribbean.” [Online]. Available: https://www.worldenergy.org/data/resources/region/latin-america-the-caribbean/. [Accessed: 02-Feb-2018]

[2] “Renewable Capacity Statistics 2017,” */publications/2017/Mar/Renewable-Capacity-Statistics-2017*, 2017

[3] J. Lata-Garcia, C. Reyes-Lopez, F. Jurado, L. M. Fernandez-Ramirez, and H. Sanchez, “Sizing optimization of a small hydro/photovoltaic hybrid system for electricity generation in Santay Island, Ecuador by two methods,” in *2017 CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies (CHILECON)*, 2017, pp. 1-6

[4] S. Ahmadi and S. Abdi, “Application of the Hybrid Big Bang–Big Crunch algorithm for optimal sizing of a stand-alone hybrid PV/wind/battery system,” *Sol. Energy*, Vol. 134, 2016, pp. 366-374

[5] A. Cano, F. Jurado, H. Sánchez, L. M. Fernández, and M. Castañeda, “Optimal sizing of stand-alone hybrid systems based on PV/WT/FC by using several methodologies,” *J. Energy Inst.*, Vol. 87, No. 4, 2014, pp. 330-340

[6] S. O. Amrouche, D. Rekioua, and T. Rekioua, “Overview of energy storage in renewable energy systems,” in *2015 3rd International Renewable and Sustainable Energy Conference (IRSEC)*, 2015, pp. 1-6

[7] A. K. Rohit and S. Rangnekar, “An overview of energy storage and its importance in Indian renewable energy sector: Part II – energy storage applications, benefits and market potential,” *J. Energy Storage*, Vol. 13,, Oct. 2017, pp. 447-456

[8] A. Evans, V. Strezov, and T. J. Evans, “Assessment of utility energy storage options for increased renewable energy penetration,” *Renew. Sustain. Energy Rev.*, Vol. 16, No. 6, Aug. 2012, pp. 4141-4147

[9] J. Lata-García, F. Jurado, L. M. Fernández-Ramírez, and H. Sánchez-Sainz, “Optimal hydrokinetic turbine location and techno-economic analysis of a hybrid system based on photovoltaic/hydrokinetic/hydrogen/battery,” *Energy*, Vol. 159, Sep. 2018, pp. 611-620

[10] J. C. Spall, “Multivariate stochastic approximation using a simultaneous perturbation gradient approximation,” *IEEE Trans. Automat. Contr.*, Vol. 37, No. 3, Mar. 1992, pp. 332-341

[11] A. Ürmös, Z. Farkas, M. Farkas, T. Sándor, L. T. Kóczy, and Á. Nemcsics,

"Application of Self-Organizing Maps for Technological Support of Droplet Epitaxy"

[12] A. Impr. Bourg-offset), D. Kumar, and S. K. Jarial, *A Hybrid Clustering Method Based on Improved Artificial Bee Colony and Fuzzy C-Means Algorithm*, Vol. 15, No. 2. Fontaine Picard, 1997

[13] B. Das and A. Kumar, "A NLP approach to optimally size an energy storage system for proper utilization of renewable energy sources," *Procedia Comput. Sci.*, Vol. 125, Jan. 2018, pp. 483-491

[14] A. S. Jacob, R. Banerjee, and P. C. Ghosh, "Sizing of hybrid energy storage system for a PV based microgrid through design space approach," *Appl. Energy*, Vol. 212, Feb. 2018, pp. 640-653

[15] Y. Yang, S. Bremner, C. Menictas, and M. Kay, "Battery energy storage system size determination in renewable energy systems: A review," *Renew. Sustain. Energy Rev.*, Vol. 91, Aug. 2018, pp. 109-125

[16] S. Bahramara, M. P. Moghaddam, and M. R. Haghifam, "Optimal planning of hybrid renewable energy systems using HOMER: A review," *Renew. Sustain. Energy Rev.*, Vol. 62, 2016, pp. 609-620

[17] T. M. I. Mahlia, T. J. Saktisahdan, A. Jannifar, M. H. Hasan, and H. S. C. Matseelar, "A review of available methods and development on energy storage; technology update," *Renew. Sustain. Energy Rev.*, Vol. 33, May 2014, pp. 532-545

[18] Amin, R. T. Bambang, A. S. Rohman, C. J. Dronkers, R. Ortega, and A. Sasongko, "Energy Management of Fuel Cell/Battery/Supercapacitor Hybrid Power Sources Using Model Predictive Control," *IEEE Trans. Ind. Informatics*, Vol. 10, No. 4, Nov. 2014, pp. 1992-2002

[19] Policy department Economic and Scientific, "Outlook of energy storage technologies"

[20] University Basque country, "Energy Storage Technologies for Electric Applications." [Online]. Available: http://www.sc.ehu.es/sbweb/energias-renovables/temas/almacenamiento_1/almacenamiento_1.html [Accessed: 10-Feb-2018]

[21] A. Raj K, S. Bag, A. Roy, U. Pal, and S. Mitra, "Battery Technologies for Energy Storage," in *Encyclopedia of Sustainable Technologies*, Elsevier, 2017, pp. 469-486

[22] C. Wang, M. H. Nehrir, and S. R. Shaw, "Dynamic Models and Model Validation for PEM Fuel Cells Using Electrical Circuits," *IEEE Trans. Energy Convers.*, Vol. 20, No. 2, Jun. 2005, pp. 442-451

[23] S. Kapila, A. O. Oni, and A. Kumar, "The development of techno-economic models for large-scale energy storage systems," *Energy*, Vol. 140, Dec. 2017, pp. 656-672

[24] Smart Hydro Power, "Smart monofloat turbine." Postdam, p. 1, 2015

[25] M. Mehrabankhomartash, M. Rayati, A. Sheikhi, and A. M. Ranjbar, "Practical battery size optimization of a PV system by considering individual customer damage function," *Renew. Sustain. Energy Rev.*, Vol. 67, 2017, pp. 36-50

[26] "PEM fuel cell degradation effects on the performance of a stand-alone solar energy system," *Int. J. Hydrogen Energy*, Vol. 42, No. 18, May 2017, pp. 13217-13225

[27] C. Ghenai and M. Bettayeb, "Modelling and performance analysis of a stand-alone hybrid solar PV/Fuel Cell/Diesel Generator power system for university building," *Energy*, Vol. 171, Mar. 2019, pp. 180-189

[28] P. Millet, S. Grigoriev, P. M. Diéguez, P. Millet, and S. Grigoriev, "Water Electrolysis Technologies," in *Renewable Hydrogen Technologies*, Elsevier, 2013, pp. 19-41

[29] "HOMER - Hybrid Renewable and Distributed Generation System Design Software" [Online] Available: http://www.homerenergy.com/ [Accessed: 18-Apr-2017]

[30] "Plan Maestro de Electrificación 2013- 2022, Volume 2, Pag 42, Residential load curve, Guayaquil electric," 2017

[31] "Instituto Oceanográfico de la Armada - Tabla de mareas puertos del Ecuador" [Online] Available: http://www.inocar.mil.ec/web/index.php/productos/tabla-mareas [Accessed: 29-Apr-2017]

[32] HOMER, "Maximum Annual Capacity Shortage" [Online] Available: https://www.homerenergy.com/support/docs/3.11/maximum_annual_capacity_shortage.html [Accessed: 28-Dec-2017]

New Approach to Multi-Criteria Ranking of the Copper Concentrate Smelting Processes based on the PROMETHEE/GAIA Methodology

Ivica Nikolić[1], Isidora Milošević[1], Nenad Milijić[1], Aca Jovanović[2], Ivan Mihajlović[1]

[1] University of Belgrade, Technical Faculty in Bor, Management Department, Vojske Jugoslavije 12, 19210 Bor, Serbia

[2] University Educons, Project Management College, Bože Jankovića 14, Voždovac, 11000 Belgrade, Serbia

inikolic@tfbor.bg.ac.rs, imilosevic@tfbor.bg.ac.rs, nmilijic@tfbor.bg.ac.rs, aca.jovanovic@pmc.edu.rs, imihajlovic@tfbor.bg.ac.rs

Abstract: Although the production of copper and its usage has been known for thousands of years, the search for the optimal process of its production is still in progress. The removal of the negative impacts of a certain technology could cost more, over a period of time, than the initial investment in the selection of the optimal technology process which would consider these effects. Hence, the selection should be supported by the innovative and modern tools, such as the applying of the multi-criteria analysis. This paper presents the implementation of the PROMETHEE/GAIA methodology for the ranking of the appropriate technological pyrometallurgical smelting process for the copper concentrate based on the eleven parameters which recognize the economic, ecological and technical aspects of the technological process. The implementation of the multi-criteria decision-making presented in this paper can be deemed as a contribution to the decision making tools, that is, to the one who makes a decision on the selection of the appropriate technological facility for smelting of the copper concentrates. The decision-makers faced with the practical need of evaluating and selecting the most appropriate technological process for pyrometallurgical copper extraction will get the greatest benefit from this multi-criteria model. The innovative contribution of this paper is also presented in the obtained model which systematically analyzes the ecological, economic and technical parameters of the copper extraction process.

Keywords: technological process; pyrometallurgical processes; ranking, PROMETHEE/GAIA

1 Introduction

Copper is one of the most important raw materials in the 21st Century. The demand for copper and its production shows a countinuous growth year after year. Therefore, there is a constant need for the development of better scientific tools and new technologies for copper extraction. Therefore enabling the preservation of the beauty of nature and its diversity as well as to increase copper production and consumption [1, 2]. These are the reasons for an evolutionary development of the technological processes in extraction of non-ferrous metals, especially copper extraction, in the last 50 years [3-5].

Copper can be produced by two basic procedures: pyrometallurgical and hydrometallurgical processes. In the Earth's crust, copper is usually presented in the form of copper – iron sulphide ($CuFeS_2$ - chalcopyrite; Cu_5FeS_4 - bornite) and copper sulphide (CuS - covellite, Cu_2S -chalcocite). It indicates that the use of pyrometallurgical processes is significantly higher than the use of hydrometallurgical treatments in obtaining copper. About 80% of copper, nowadays, has been produced by pyrometallurgical process [2, 6]. Also, numerous improvements that have occurred in the previous periods, have led to increasing production infrastructure and capacity as well as to reducing negative impact on the environment [7-9].

Copper smelters around the world currently implement a great number of different pyrometallurgical processes for copper extraction [6, 10, 11]. The authors of this study believe that, in the course of the ranking and selection of the appropriate technological process for the pyrometallurgical copper extraction, it is necessary to make the selection based on the multiple parameters considered simultaneously. Systomaticly resulting in obtaining an insight into the economic benefits of the process as well as the environmental and technological benefits. Some technological procedures, apart from their economic performances, may have bad ecological indicators and impacts which are not in accordance with the norms of the World Health Organization (WHO) [12, 13], which should be also considered.

This study has considered eight technological processes with the greatest application in pyrometallurgical copper extraction in the world [6, 11, 14]. Based on the largest representation of the available technology, the following technological procedures have been considered: Outokumpu flash smelting (in contemporary practice known as Outotec flash smelting); Ausmelt/Isasmelt lance; Inco Flash; Mitsubishi; Noranda; El Teniente; Vanyukov and Reverberatory process.

The motive for this research derived from the detailed investigation of the literature, based on which it was concluded that there is no research in the available literature analyzing systematically the ecological and technical parameters. Also, the methodology applied in this paper is not sufficiently represented in this field. The aim of this paper is to perform prioritization and

selection of overall ranking indicators, for each of the given technologies, which have been done on the basis of the eleven selected relevant process parameters. The following parameters have been considered: concentrate amount in charge (t/day); Cu content range in the concentrate, i.e. the difference between maximum and minimum possible content in the concentrate that the chosen technology can process (%); Cu content in copper matte (%); Fe content in copper matte (%); production of waste slag (t/day); production of copper matte (t/day); campaign life (years); sulfur recovery (%); copper recovery (%); Cu content in waste slag (%), and minimal Cu content in the concentrate.

The structure of this paper is the following: Section 2 presents a literature review of analysed methodologies; Section 3 gives summary of considered technological processes; Section 4 describes applied methodology and the obtained results; Section 5 provides the discussion of the results; the final section, Section 6, presents the conclusions with indicated contributions and plans for further research.

2 Literature Review

Based on a broad literature review, the authors have come to the conclusion that since 1950s and 1960s, when foundations of Multicriteria Decision Making Methodology (MCDM) methods have been developed, MCDM have been applied in many areas [15-20]. Many studies attempt to develope new MCDM models and techniques and in the past decades the development of MCDM techniques have accelerated and seem to continue growing exponentially [20]. One of the most used MCDM technique is Analytic Hierarchy Process (AHP) developed by Saaty (1980) [21]. Main advantages of this method are usability, flexibility, dealing with tangible and non-tangible attributes and comparing alternatives with relative ease [22]. However, the main flaws of the AHP method is that it does not include appropriately, a decision-maker's pattern of thinking. For that reason Van Laarhoven and Pedrycz (1983) proposed fuzzy AHP which appropriately include decision-maker's thinking, but fuzzy numbers are not a solution when decision makers are hesitant in defining membership functions [22, 23]. Later, various variants of the FAHP and AHP methods were developed. For example, Zhu et al., (2015) presented AHP method based on rough numbers to determine the weight of each evaluation criterion, while Radwan et al. (2016) extended the AHP method via the neutrosophic set because neutrosophic logic is able to deal with contradictions which are true and false at the same time and also might be capable of simulating the human thinking [22, 24]. PROMETHEE method was developed by Jean-Pierre Brans (1982). Group of PROMETHEE methods includes PROMETHEE I for partial ranking, the PROMETHEE II for complete ranking, PROMETHEE III for ranking based on interval, PROMETHEE IV for complete or partial ranking of the alternatives when the set of viable solutions is continuous,

the PROMETHEE V for problems with segmentation constraints, PROMETHEE VI for the human brain representation, the PROMETHEE GDSS for group decision making, PROMETHEE TRI for dealing with sorting problems and the PROMETHEE CLUSTER for nominal classification [25, 26].

Also, its application in the selection of technological processes in the industry is significant [27-30]. On the other hand, the application of the PROMETHEE/GAIA methodology in the field of pyrometallurgy is limited and can be seen through the work of the group of authors, Nikolic D. et al. (2009), which were rankingcopper concentrates according to their quality [5]. However, the application of this methodology in the field of pyrometallurgical process selection is very scarce. Especially when viewed through an adequate understanding of the technical and ecological aspects of the observed technological processes, at the same time.

3 Summary of Considered Technological Processes

A brief presentation of the main characteristics of each considered technological process for pyrometallurgical copper extraction, currently in use in smelters all over the world will be provided in this section.

3.1 Outokumpu Flash Smelting (Outotec)

Outotec process has had very successful development in the last 50 years, and today it is the most widely used process in copper and nickel production [6, 35-37]. About 50% of total world copper and nickel production is obtained by this technological procedure. Outotec smelting technology has the leading position in the copper production based on its economy, adaptability, low energy consumption and high sulfur recovery. Sulfur recovery in this technology ranges from 94% to 99%. Autogenous smelting process of the copper concentrates in this procedure, compared to the reverbatory furnace, which were mostly presented in the world plants until 1970s, has significantly better technical and economical indicators: more efficient utilization of sulphide energy from concentrate, higher metal and sulfur utilization and far better protection from SO_2 and other harmful substances pollution [6, 10, 38, 39]. According to available data, currently 21 smelting plants in the world use this technology [40].

3.2 Ausmelt/Isasmelt Lance

Ausmelt/Isasmelt is a simple and highly efficient production process of the ferrous metals. This process of continuous smelting of sulfide, copper and other

concentrates and materials is a newer autogenous smelting process and is present in nine world smelters [10, 40, 41]. It was developed in Australia by Mount Isa Mines Limited and Australian Commonwealth Scientific and Research Organisation (CSIRO), and was introduced in commercial use in 1992 [6]. The basic smelting technology of Isasmelt/Ausmelt consists of a lance submerging into the melt from the furnace top (TSL-Top Submerged Lance). This technology occupies one of the leading position in the world for low production cost and it meets strict environmental standards [42].

3.3 Inco Flash

Commissioning INCO smelting furnace in 1952, the deficiencies of the dominant reverbatory furnaces were eliminated, in order to better utilize sulfide mineral from concentrate. It helped in reducing the energy consumption, improving ecological environment conditions and increasing copper utilization [2, 11, 36, 43]. Introducing this process in the industrial application unifies the stages of roasting and smelting. Also, by replacing oxygen with technical oxygen the amount of gases generated in the process is reduced up to 40 times, compared to the reverbatory furnace, improving the economy of the operation and the protection of the atmosphere. However, the main disadvantage of the technology, beside expensive and complex charge preparation, is the great consumption of electricity, required for the oxygen production, with 50% participation in the total costs. Other significant ecological parameters are: sulfur utilization of average 93.6%, production dust of 95 to 230 tons per day, quantity of waste gases of 35000 Nm^3/h, whereas SO_2 in waste gases is 70% [6, 44, 45].

3.4 Mitsubishi

Mitsubishi process is a process of autogenous continuous smelting of copper sulfide concentrates. This technological procedure is characterized by high SO_2 utilization, which is removed through smelting and converting process and makes about 99.5%, and is further directed into the sulfuric acid or liquid SO_2 production, via electrostatic precipitator [46-48] Emission of harmful gases is also reduced as the transport of melted material (melt), from one aggregate into the other, is not performed in the smelting pots. Mitsubishi process of continuous copper concentrate smelting has been continuously modernized and improved which has led to increased environment protection. This procedure is used in smelting plant in Naoshima, Japan, with the capacity of 240000 tons of copper per year and Kidd Creek, Canada, with the capacity of 120000 to 150000 tons of copper per year. The advantage of this procedure is its flexibility related to the conventional copper concentrate smelting methods, due to its ability to process reversal and secondary materials of different kinds and composition [49, 50]. Beside reduced electricity consumption, other process parameters that have impact

on environment are: production dust ranging from 60 to 67 t/day, amount of waste gases of 500 Nm^3/h, as well as SO_2 in waste gases making 25 to 30% [6, 10].

3.5 Noranda

Noranda Inc. is a mining metallurgical company located in Rouyn-Noranda, Quebec, Canada. Noranda process has been constantly improved. By using 34% of oxygen enriched air, they have almost achieved autogenous process with the use of small amounts of fuel, where as the oxygen grade of 40% achieved fully autogenous smelting. However, this oxygen enrichment is the upper limit at which the rapid wearing of refractory lining of reactor was observed that threatened the stability of the entire aggregate [51, 52]. Significant parameters of this technological process that have an impact on an ecosystem are: specific consumption of fuel heat making 2321-2954 MJ/t of concentrate, sulfur utilization ranging up to 94%, production dust 70-100 t/day, waste gases of 55000 Nm^3/h, as well as SO_2 grade in waste gases ranging from 16-20% [6, 10, 36].

3.6 El Teniente

Increasing of the mining capacities in Chile required increasing of smelting capacities, better energy utilization, modernization of smelting process and more economical production. Therefore, Teniente, presents an important technology for copper concentrate smelting and processing. Control of work process is more complex than with the other technologies. The complexity is caused by technological process characteristics [53]. Smelting process products that have great impact on environment are: sulfur utilization ranging from 90%-98%, production dust of 50 t/day, amount of wast gases of 60000 Nm^3/h, SO_2 in outlet gases ranging 12%-25% [10, 36, 41, 54-56].

3.7 Vanyukov

Vanyukov process is an intensive autogenous process of copper sulphide concentrates in a bubbling bath. After the extensive testing in semi-industrial and industrial conditions, the first furnace was commissioned in 1982, in smelter in Norilsk (Russia). Six furnaces of this kind were installed in the ex Soviet Union. The process was named after its author, academician, professor A. V. Vanjukovu, in 1988. Some characteristic of the technology are the following: sulfur utilization 90%, production dust 0.5 to 0.9% per ton of charge, amount of waste gases 35000-55000 Nm^3/h, SO_2 in waste gases 25-40% [6, 10]. Also, the advantage of this smelting process is a possibility of autogenous or semi-autogenous smelting of poor and rich copper concentrates with different additives (humidity 6-8%), coarse and selective excavated rich copper ore, reversal material (coarseness to 50 mm),

fuel (lumpy coal) and flux, so the procedure is far more flexible than smelting processes in floating conditions [11, 41].

3.8 Reverberatory Furnace

Although this copper extraction technology has been mostly replaced in almost entire world by some of the above described ones, the authors of the study believed that it should also be taken into consideration, as it used to be the basic process for the pyrometallurgical Cu production. This traditional way of sulfide concentrates smelting is still used in the countries such as China (Changzhou; Jiangsu; Shuikoushan; Hunan; Wuhu; Anhui); Germany (Hettstedt); Hungary (Csepel); Iran (Sar Chesmeh), Romania (Zlatna); etc. [11, 40, 57, 58]. Mixing copper concentrate with fluxes in predetermined ratio which is calculated in advance forms the charge. Charge formed in this way, requires previous roasting process. Batch roasting process is an exothermic process and represents a partial oxidation of sulphides, wherein the amount of sulfur is reduced to the limit that ensures smelting of the rich copper matte. In the course of this process, at temperature of 650 ºC - 700 ºC reaction of dissociation of higher sulphides into lower ones occurs, oxidation of sulfuric vapors into SO_2 and partly oxidation of lower sulphides into oxides. [56, 59]. Oxdiation level and roasting speed depend on the excess air. The success of roasting process is measured by the degree of desulfurization which represents the ratio of removed sulfur related to the total sulfur grade of the charge. Roasting process products are calcine which further goes into the smelting process and gases containing 8%-9% of SO_2 which are directed from reactor to cooling (spray cooling), cleaning (electrofilters), and, then, through pipelines to sulfuric acid plant. Previously roasted charge (calcine) is loaded telescopically into the reverbatory furnace. Smelting process is performed at high temperatures (1550 ºC in the focus of the flame) in more or less oxidizing atmosphere, and the smelting products are copper matte containing 35%-45% of copper, slag with 0.5%-1% of copper and gases with 0.5%-1.5% of sulfur dioxide [41, 56, 60, 61].

4 Methodology and Results of Research

4.1 Data Collection

Data used for analyzing and ranking technology presented in this study have been collected from the relevant literature [6, 10, 11, 36, 41, 44, 56, 62, 63]. Based on the analysed data from literature, the following criteria for ranking technologies have been selected: quantity of concentrate in charge, Cu content range in the

concentrate, Cu content in the copper matte, Fe content in the copper matte, production of waste slag, production of copper matte, campaign life, sulfur utilization, copper utilization, Cu content in waste slag and minimal Cu content in concentrates.

The average value for each of the eleven parameters was taken for the analysis, from a large number of smelting plants operating these technologies. Multicriteria Decision Making Methodology (MCDM) was used for the ranking of the analized technological processes for the smelting of the copper concentrate, according to eleven parameters, applying Decision Lab 2000 software [6, 31-34]. The PROMETHEE method has been used in the study, within multi-criteria analysis, for the selection of the optimal technology, based on pre-defined criteria. Application of GAIA plane has provided the graphical interpretation of the PROMETHEE method, thus supporting the analysis of the given decision and selection problem visually.

MCDM methods provide mathematical models for ranking alternatives based on the selected criteria, clearly and transparently presenting ranking alternative results and synthesis of final results [64, 65]. PROMETHEE (Preference Ranking Organization METHod for Enrichment Evaluation) is a kind of MCDM method developed by J. P. Brans and its basic point is comparison based on alignement [66-68]. The main objective of the method selecting for ranking copper producting technologies is the fact that PROMETHEE takes into account inner relationships of all evaluation factors in the decision making process as well as alternatives valuating according to each determined criterion. In addition to all selected ranking criteria are quantitative, PROMETHEE method enables consideration of multiple criteria in incommensurable units [69], which is additional reason for this method selection. Ranking results obtained by PROMETHEE methodology are followed by graphic display of alternatives applying GAIA plane. Hence, the technology has been named in contemporary literature PROMETHEE GAIA. The number of the studies applying PROMETHEE GAIA method in their research is increasing year after year, which is confirmed by numerous publications [64, 65, 67, 68, 70-76].

4.2 Multicriteria Analysis for Copper Obtaining Technologies

Data collected on the base of the available technical characteristics from analysed technological processes, have been found in the literature resources [6, 10, 11, 36, 41, 44, 56, 62]. Given data present basis for multicriteria analysis of the actual technologies for copper obtaining. Collected values are results of eleven considered technical characteristics (criteria) analysed for each individual technology (alternative), and they form collection of baseline data for PROMETHEE calculations (Table 1).

None of the parameters (criteria) used in this study, was given advantage related to other parameters, therefore weight coeffecients have been excluded as a part of multicriteria analysis to avoid subjectivity assignment and give preference to one technology over another [70-73].

PROMETHEE METHOD includes six potential preference functions enabling user to express differences on the base of minimal dissents [33, 34]. Research presented in this study, has used type 1 function (usual). Preference function usual has been selected as the best solution for describing the analysed data (all data are quantitative) [77]. Min/Max values orientations are based on the context of each considered technology characteristics and their potential impact on researched technologies (Table 2). Thus, evaluation matrix has been given in Table 2.

Table 1

Average values of parameters used for ranking copper extraction technologies

Criterion / Alternatives	Concentrate amount in charge (t/day)	Cu content range in the concentrate (%)	Cu content in copper matte (%)	Fe content in matte (%)	Production of waste slag (t/day)	Production of copper matte (t/day)	Capmaign life (year)	Sulfur recovery (%)	Copper recovery (%)	Cu content in waste slag (%)	Minimal Cu content in the concentrate (%)
Outokumpu flash	2750	5	65	11.5	2025	1500	9	96	97	0.65	26
Ausmelt / Isasmelt lance	2250	4	67	14	1210	1200	2.1	97	97	0.6	25
Inco Flash	3000	9	50	15	1350	935	15	93.6	97.5	0.65	20
Mitsubishi	2150	6	71.5	7.75	1375	1209	3	99.5	97	0.75	28
Noranda	2250	9	72.5	4.5	1500	975	1.75	94	95	0.75	28
El Teniente	2300	7	73	4.5	1725	962.5	1.5	90	96	0.325	26
Vanyukov	2150	8	59.5	10.5	1750	1300	4.5	90	98	0.6	26
Reverberatory	2000	16	40	27.5	1050	1450	3.5	50	93	0.75	19

Table 2
Preferences functions and Min/Max values orientations

Criterion	Concentrate amount in charge (t/day)	Cu content range in the concentrate (%)	Cu content in copper matte (%)	Fe content in matte (%)	Production of waste slag (t/day)	Production of copper matte (t/day)	Capmaign life (year)	Sulfur recovery (%)	Copper recovery (%)	Cu content in waste slag (%)	Minimal Cu content in the concentrate (%)
Preferences functions	Usual	Usual	Usual	Usual	Usual	Usual	Usual	Usual	Usual	Usual	Usual
Min/Max	Max	Max	Max	Min	Min	Max	Max	Max	Max	Min	Min

Character of each criterion may be Max or Min and it is defined according to pre-set goals. For example, criterion concentrate quality in charge has been defined as Max, as greater values of given parameter realize economical production and have significant impact on better business results. Greater range of concentrate content flexibility will also provide better recovery, so character maximum value has also been used for this parameter. If the content range that can be technologically processed is greater, it implies that concentrates poor in ore could also be processed and the advantage is given to technologies that can melt concentrates with greater variation of copper grade. Copper grade in copper matte is defined as a parameter requiring higher value, hence it is also defined Max. Also, reactor lining repair (campaign life) is also defined max, as better lining cooling and longer campaign life are increasing economy of the process. Sulfur and copper recovery are also defined as max, from economical reasons as well as from ecological reasons. Character of other parameters has been defined as Min. The tendency is to keep Fe content in copper matte as small as possible to get the higher copper content, which requires less scope of further rafination. Also, Cu content in slag should be at the lowest possible level to obtain the best possible copper recovery Priority is given to smelters with possibility to melt poor concentrates, hence parameter of minimal Cu content in concentrate has been defined as Min. [5, 78, 79]. Multicriteria comparative analyses of copper obtaining technologies has been performed by using the software package Decision Lab 2000 [5, 34, 79]. Main screen of the software package Decision Lab 2000 is presented in Figure 1.

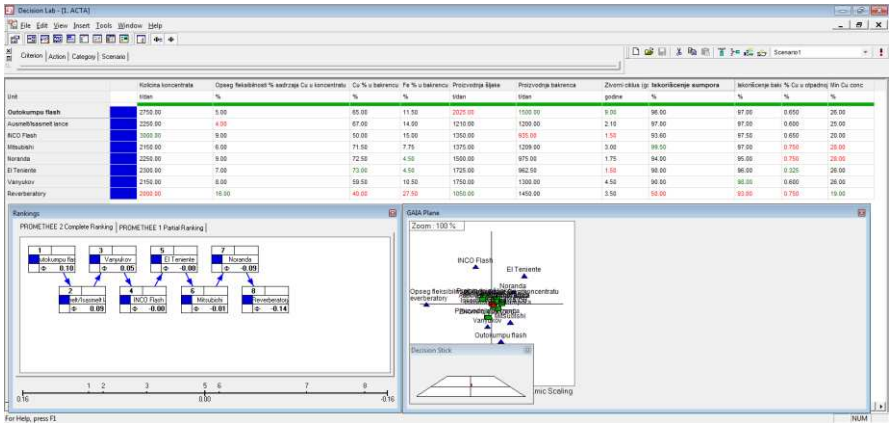


Figure 1
Main screen of the software package Decision Lab 2000

PROMETHEE method is based on determination of positive (ϕ+) and negative flow (ϕ-)for each alternative. Positive preferance flow shows to what extent certain alternative dominates over the other alternative.The higher the value is (ϕ+→1), the alternative is more significant. Negative preferance flow shows to what extent certain alternative is preferred by other alternatives. Alternative is more significant if the flow value is less (ϕ-→0). Entire ranking within PROMETHEE II is based on calculating netto flow (ϕ), which presents difference between preference positive and negative flow. Alternative with the highest netto flow value is best rated, and the one with the lowest netto flow is the worst rated [69, 80, 81]. Complete ranking (PROMETHEE II) of eight copper obtaining technologies was performed on the base of the given data and alternative values (technical characteristic for each researched copper obtaining technology, given in Table 1). Obtained results are given in Table 3 and Figure 2.

Table 3
Results of complete ranking of copper obtaining technologies based on PROMETHEE II MCDM method

Rang	Alternative	Φ^+	Φ^-	Φ
1	Outokumpu flash	0.5195	0.4156	0.1039
2	Ausmelt / Isasmelt lance	0.5195	0.4286	0.0909
3	Vanyukov	0.4935	0.4416	0.0519
4	Inco Flash	0.4805	0.4805	0.0000
5	El Teniente	0.4675	0.4675	0.0000
6	Mitsubishi	0.4545	0.4675	0.0130
7	Noranda	0.4156	0.5065	-0.0909
8	Reverberatory	0.4156	0.5584	-0.1429

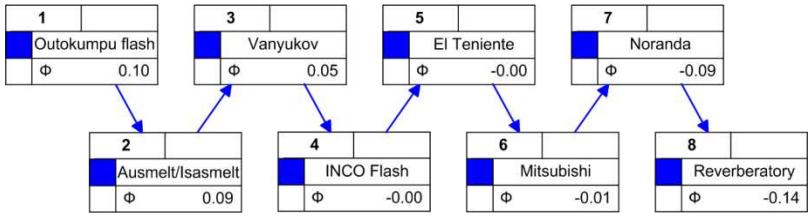


Figure 2

PROMETHEE II complete alternative ranking (technologies for pyrometallurgical copper extraction)

Very significant benefit of applying PROMETHEE methodology using Decision Lab, is visualization of obtained results, i.e. solution ranking by GAIA (Geometrical Analysis for Interactive Aid) plane. GAIA plane and results obtained by it, facilitate evaluation of obtained solutions as well as interpretation of significance of the individual variables. GAIA analysis provides important information on ranking within two-dimensional space, obtained by PCA extraction. This way, it is possible to display issues of conducted ranking graphically, to determine specific relationship characteristic between the selected alternatives. Positions of alternatives considered (triangles) determine strengths or weaknesses of activity properties related to selected criteria, determining future result of conducted final ranking. As the alternative is closer to criterion vector direction, it has a better value for the criterion [80, 82]. Figure 3 shows position of the considered alternative on GAIA plane.

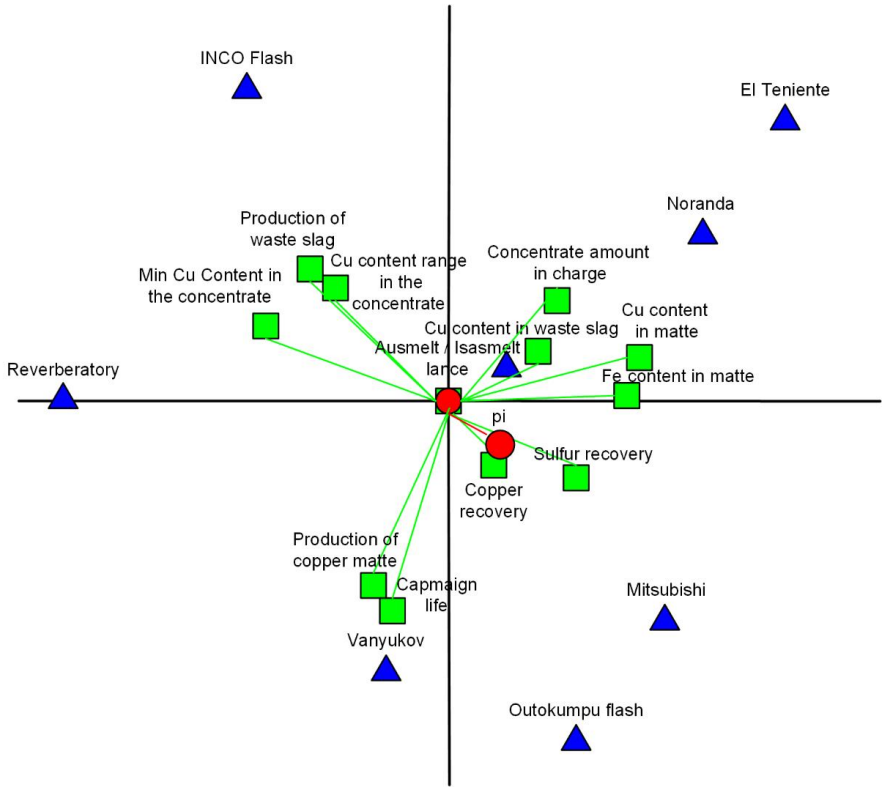


Figure 3
GAIA plane selection of the most appropriate alternative (copper obtaining technology)

5 Results Discussion

Based on the collected technical characteristics of copper obtaining technologies considered, it is obvious that it is possible to perform ranking of these characteristics (criteria) simultaneously. It means that application of multicriteria analysis enables multi-criteria ranking, list of priorities and thorough analysis of the problem. Data given in Table 1 were analysed that way, applying software package Decision Lab 2000. The results of the complete PROMETHEE II ranking are shown in the Table 3 and Figure 2. Visual presentation of the ranking is presented in Figure 3, providing comparison of all alternatives per a criterion–GAIA plane. Percentage of the data collecting on GAIA plane, i.e. reliability of graphic interpretation is greater than 60% (Φ:70.19%), which is considered as very acceptable [80].

Position of the alternative (triangles on GAIA plane) determines its strengthen or weakness related to the criteria. If alternative is closer to the axis direction of a criterion, the alternative is a better option for the given criterion. Hence, Outokumpu flash technology has the best performances, as the alternative is the closest to the axis direction of the criterion with the greatest impact and positioned very close to axes of other criteria as well, unlike the most other alternatives. Vector pi (stick decision) is presented in the form of axe ending in circle and is an optimal solution according to given weight criteria. The best decision is the one closest to stick decision (Outokumpu flash), whereas the worst alternative is Reverberatory which is good at no criteria (except the possibility of processing poor concentrates), and it is located opposite of the stick decision direction.

Observing eccentricity of criteria positions (axes ending in square), i.e. their distance from the coordinate origin, it is obvious that criteria Utilization of Cu, Utilization of SO_2 and Fe in matte have greater impact related to other eight criteria. Additionally, they are closest to stick decision pi, which is confirmed by their greatest impact at alternative ranking. Taking into consideration that relatively great number of criteria have been used, their quality may be assessed as a satisfactory, as there are no expressed conflicts between them (there are no two criteria positioned opposite one another on GAIA plane.

As the ranking of technological process in this study, were based on the technical criteria, such result could have been expected, based on the collected data set. Outokumpu technology (ranked on the top) is characterized by autogenous continual smelting of sulfide copper and other concentrates and materials, and belongs to the group of the latest autogenous smelting processes. This technology is one of the leading technologies in the world for the low production costs and meets strict environmental standards. Opposed to it, Reverberatory technology has been ranked at the bottom for its poor technical characteristics (Cu in matte=56.50%; Fe in matte= 15.00%, Cu in slag=1.20%). Therefore, these are the reasons why it has been almost completely suppressed, and is still being used in just few smelters in the world.

Noranda technology also has very poor performances whereas Mitsubishi, INCO Flesh and El Teniente, have no significant positive performances (Noranda has significantly negative and Mitsubishi negative flow value Φ, whereas INCO Flesh and El teniente has neto flow value Φ near to 0 - Table 3 and Figure 2). The primary reason for such ranking of Norande and Mirtsubishi technologies should be found in the fact that they mostly process concentrates with high Cu content and create waste dross with increased Cu content. On the other hand, Inco flesh has been relatively poorly ranked for low Cu content in copper matte, low productivity of copper matte and short operation life. The similar situation is with El teniente technology which additionally has low total Cu recovery.

Based on their technical characteristics, Ausmelt/Isasmelt and Vanyukov technologies are ranked relatively satisfactory (positive netto flow values Φ).

Ausmelt/Isasmelt technology is ranked (positioned) positively according to the following criteria: Cu in slag, Utilization of Cu, Concentrate Quantity in charge, Cu in matte and Fe in matte. They, actually, explain satisfactory performances of those technical characteristics at Ausmelt/Isasmelt technology. Besides, the possibility for improving technical performances of this technology for copper obtaining is in the fields Campaign life, Matte production and Min Cu in concentrate. Based on the obtained results, the very interesting data is that, contrary to Ausmelt/Isasmelt technology, Vanyukov technology has been ranked (positioned) positively according to the criteria Campaign life and Matte production. Slag production, Range of Cu content flexibility in concentrate, Min Cu in concentrate and Content Quantity in charge are potential fields for improving performances of Vanyukov technology.

Finally, El teniente and Mitsubishi alternative position on GAIA plane, indicate that it is possible to achieve improvement of the technical characteristic performances of these technologies by improving any of considered technical characteristics for their low negative netto flow falue Φ. On the other side, improving positions of Reverberatory furnace technology, or even Noranda technology, would require entire modification of values for almost all considered operational parameters, as negative netto flow values Φ are significantly great for these two technologies.

Conclusion

Copper production presents one of the primary activities in the sector of industry. It is energetically intensive and therefore it is positioned on the third place according to consumption of the specific energy, considering production of five base metals [83-85]. For this reason, it is necessary to conduct performing analysis of the available technologies for copper production and propose their further improvement. The fact is that technologies for pyrometallurgical copper extractions from sulfide type concentrates have evolved and are still evolving with the aim to optimize technical and economical as well as ecological process parameters. That is the reason why this study has analysed current copper extraction technologies and performed their multicriteria ranking.

The study is based on prioritization and selection of observed technologies whereas the selection has been performed according to the average values of the chosen eleven parameters. Most technologies considered in the study are based on copper concentrate autogenous smelting process starting from requirement for integration of particular processing stages (roasting, smelting, converting). Many leading companies dealing with copper production, have developed their own copper concentrate autogenous smelting processes which differ technologically and operationally. Also, there is a small number of companies still operating with technical processes which are not autogenic (as a reverberatory furnace). Those are the reasons why both kind of processes have been analysed.

Multicriteria comporative data analysis used in this study, has provided significant conclusions on ranking and possible selection of the optimal technology for copper concentrate smelting, performed on the base of the eleven selected parameters belonging to the group of technical and ecological indicators. Moreover, multicriteria analysis explicitly emphasizes all key elements, that mostly influence the final prioritization result. The best ranked Outokumpu flash technology has totally best performances. If only individual ranking parameters were taken into account, this technology would have advantage over all others for the great quantity of produced copper matte and long campaign life. However, it has been the best ranked technology considering impact of all factors simultaneously, which is the aim of multicriteria analysis presented in this study.

Contribution of this research is in the systematical approach, in analysing of actual smelting technologies, by using MCDM methodology. Also, another contributioin is in developing the apropriate basis for development of the MCDA approach, that can be used in selection of the optimal smelting technology, based on the combination of input parameters, which is also partially filling the research gap in this field. This way, besides scientific contribution, presented methodology can be useful to decision makers, i.e. representatives of companies dealing with copper smelting process, as a good tool in the process of selecting the optimal technologies, based on different criteria they want to take into consideration.

The essence of this work is to present possibility of applying method MCDM as a useful support in decision making process, to decision maker dealing with ranking and selecting copper extracting technologies. Normally, the study has presented average parameter values from different world smelters, used by the given technologies. However, the same method could be used in realistic business environment in decision making processes based at real technological process parameters. Besides, the study implies that all criteria has the same weight coefficient value. But, if potential decision makers in the specific actual conditions consider that, e.g. ecological parameters are of greater importance (such as SO_2 utilization), they can appoint this parameter with greater weight coefficient than other parameters. That will be the subject of further research that will include interviewing experts in the respected field on importance of each of 11 analysed parameters, and weight parameters of new ranking will be appointed according to the obtained opinions. There is also an interesting example from RTB Company Bor (Serbia) [86], when, in 2013 a group of experts from respected field of metallurgical sciences were included in selection of a new technology to replace the reverberatory furnace that had been in operation until then. Outokumpu technology was selected according to the opinion of experts clearly demonstrating that presented selection and results of multicriteria analysis given in this study has practical stand-fast.

Acknowledgement

The authors feels indebted to the company Visual Decision Inc. Montreal, Canada; for software package Decision Lab 2000 provided to them free of charge.

Part of this research is financially supported through the project of the Ministry of Education, Science and Technological Development of Serbia – TR34023

References

[1] Sievers H, Meyer FM. Parameters influencing the efficiency of copper extraction. Erzmetall. 2003;56(8): pp. 420-425

[2] Požega E, Gomidželović L, Trujić V, Živković D. Analysis of advanced technologies in copper metallurgy. Copper. 2010;35(1): pp. 15-24

[3] Herreros O, Quiroz R, Manzano E, Bou C, Vinals J. Copper extraction from reverberatory and flash furnace slags by chlorine leaching. Hydrometallurgy. 1998;49(1-2): pp. 87-101

[4] King GM. The evolution of technology for extractive metallurgy over the last 50 years – is the best yet to come? JOM. 2007;59(2): pp. 21-27

[5] Nikolic D, Jovanovic I, Mihajlovic I, Zivkovic Z. Multi-criteria ranking of copper concentrates according to their quality – An element of environmental management in the vicinity of copper – Smelting complex in Bor, Serbia. J. Environ. Manage. 2009;91(2): pp. 509-515

[6] Schlesinger M, King M, Sole K, Davenport W. Extractive Metallurgy of Copper (Fifth edition). Amsterdam (NL): Elsevier; 2011

[7] Franzin WG, McFarlane GA, Lutz, A. Atmospheric fallout in the vicinity of a base metal smelter at Flin Flon, Manitoba, Canada. Environ. Sci. Technol. 1979;13(12): pp. 1513-1522

[8] Filipou D, St.German P, Grammatikopolus T. Recovery of metal values from copper – arsenic minerals and other related resources. Miner. Process. Extr. Metall. Rev. 2007; 28: pp. 247-298

[9] Aznar JC, Richer-Lafleche M, Cluis D. Metal contamination in the lichen Alectoria sarmentosa near the copper smelter of Murdochville, Quebec. Environ. Pollut. 2008;156(1): pp. 76-81

[10] Davenport W, King M, Schlesinger M, Biswas A. Extractive Metallurgy of Copper (Fourth edition). Amsterdam (NL): Elsevier, 2002

[11] Moskalyk R, Alfantazi A. 2003. Review of copper pyrometallurgical practice: today and tomorrow. Miner. Eng. 2003;16: pp. 893-919

[12] Nikolic D, Milošević N, Mihajlović I, Živković Ž, Tasić V, Kovačević R, Petrović N. Multi-criteria Analysis of Air Pollution with SO2 and PM10 in Urban Area Around the Copper Smelter in Bor, Serbia. Water, Air & Soil Pollution. 2010;206: pp. 369-383

[13] WHO (World Health Organization), 2015. WHO Expert Consultation: Available evidence for the future update of the WHO Global Air Quality Guidelines (AQGs) Regional Office for Europe, Bonn, Germany

[14] Jovanović I, Stanimirović P, Živković Ž. Environmental and economic criteria in ranking of copper concentrates. Environ. Model. Assess. 2013;18(1): pp. 73-83

[15] Badi I, Ballem M,Supplier selection using the rough BWM-MAIRCA model: A case study in pharmaceutical supplying in Libya. Decis. Mak. Appl. Manag. Eng.2018;1(2):pp. 16-33

[16] Pamučar D, Stević Ž, Sremac S, A New Model for Determining Weight Coefficients of Criteria in MCDM Models: Full Consistency Method (FUCOM). Symmetry. 2018;10(9): p. 393

[17] Liu F, Aiwu G, Lukovac V, Vukic M, A multicriteria model for the selection of the transport service provider: A single valued neutrosophic DEMATEL multicriteria model. Decis. Mak. Appl. Manag. Eng. 2018;1(2): pp. 121-130

[18] Bojanić D, Kovač M, Bojanic M, Ristic V, Multi-criteria decision making in defensive operation of guided anti-tank missile battery: An example of hybrid model fuzzy AHP-MABAC. Decis. Mak. Appl. Manag. Eng. 2018;1(1): pp. 51-66

[19] Pehlivan NY, Şahin, A., Zavadskas, E. K., & Turskis, Z. A comparative study of integrated FMCDM methods for evaluation of organizational strategy development. J. Bus. Econ. Manag.2018;19(2):pp. 360-381

[20] Zavadskas EK, Turskis Z, Simona Kildiene S. State of art surveys of overviews on MCDM/MADM methodsTechnol. Econ. Dev. Eco. 2014;20(1): pp.165-179

[21] Saaty TL. The analytic hierarchy process. New York (US): McGraw-Hill; 1980

[22] Radwan NM, Badr Senousy M, Riad AEDM. Neutrosophic AHP Multi Criteria Decision Making Method Applied on the Selection of Learning Management System. Int. J. Adv. Comput. Technol. 2016;8(5): pp. 95-105

[23] Van Laarhoven PJM, Pedrycz W. (1983). A fuzzy extension of Saaty's priority theory. Fuzzy. Sets. Syst.1983;11: pp. 229-241

[24] Zhu GN, Hu J, Qi J, Gu CC, Peng YH. An integrated AHP and VIKOR for design concept evaluation based on rough number. Adv. Eng. Inform.2015;29(3): pp. 408-418

[25] Brans JP. L'ingénièrie de la décision; Elaboration d'instrumentsd'aide à la décision. La méthode PROMETHEE. In: Nadeau R and Landry M (eds). L'aide à la décision: Nature,Instruments et Perspectives d'Avenir. Presses de l'Université Laval: Quebec, 1982; pp. 183-214

[26] Behzadian M, Kazemzadeh RB, Albadvi A, Aghdasi M. PROMETHEE: A comprehensive literature review on methodologies and applications. Eur. J. Oper. Res. 2010;200(1): pp. 198-215

[27] Vasiljević M, Fazlollahtabar H, Stević Ž, Vesković S. A rough multicriteria approach for evaluation of supplier criteria in automotive industry. Decis. Mak. Appl. Manag. Eng. 2018;1(1): pp. 82-96

[28] Stević Ž, Pamučar D, Vasiljević M, Stojić G, Korica S. Novel integrated multi-criteria model for supplier selection: Case study construction company. Symmetry. 2017;9(11): p. 279

[29] Liu D, Yuan Y, Liao S. Artificial neural network vs. nonlinear regression for gold content estimation in pyrometallurgy. Expert Systems with Applications, 2009;36(7): pp. 10397-10400

[30] Nikolic D, Milosevic N, Zivkovic Z, Mihajlovic I, Kovacevic R., Petrovic, N. Multi-criteria analysis of soil pollution by heavy metals in the vicinity of the Copper Smelting Plant in Bor (Serbia). J. Serb. Chem. Soc. 2011;76(4): 625-641

[31] Azadeh A, Izadbakhsh HR. A MULTI-VARIATE/MULTI-ATTRIBUTE APPROACH FOR PLANT LAYOUT DESIGN. Int. J. Ind. Eng.Theory Appl. Pract. 2008;15(2): pp. 143-154

[32] Soota T, Singh H, Mishra RC. Selection of Curricular Topics Using Framework for Enhanced Quality Function Deployment. Int. J. Ind. Eng.Theory Appl. Pract. 2009;16(2), pp. 108-115

[33] Nikolić D, Milošević N, Živkovic Ž, Mihajlović I, Kovačević R, Petrović N. Multi-criteria analysis of soil pollution by heavy metals in the vicinity of the Copper Smelting Plant in Bor (Serbia). J. Serb. Chem. Soc. 2011;76(4): pp. 625-641

[34] Milijić N, Mihajlović I, Nikolić D, Živković Ž. Multicriteria analysis of safety climate measurements at workplaces in production industries in Serbia. Int. J. Ind. Ergon. 2014;44: pp. 510-519

[35] Higgins DR, Gray NB, Davidson MR. Simulating particle agglomeration in the flash smelting reaction shaft. Miner. Eng. 2009;22: pp. 1251-1265

[36] Vračar R. Theory and practice of non-ferrous metals. Belgrade (RS): Association of metallurgical engineers of Serbia; 2010 (In Serbian)

[37] Jian-hua L, Wei-hua G, Yong-fang X, Chun-hua Y. Dynamic modeling of copper flash smelting process at a Smelter in China. Appl. Math. Modell. 2014;38: pp. 2206-2213

[38] Outokumpu: http://www.outokumpu.com/en/company/history/Pages/default.aspx

[39] Outotec: http://new.outotec.com

[40] USGS: https://mrdata.usgs.gov/mineral-resources/copper-smelters.html

[41] Najdenov I, Rai K, Kokeza G. Aspects of energy reduction by autogenous copper production in the copper smelting plant Bor. Energy. 2012;43: pp. 376-384

[42] Isasmelt: http://www.isasmelt.com/EN/technology/Pages/Technology.aspx 2017

[43] Queneau PE, Marcuson SW. 1996. Oxygen pyrometallurgy at copper cliff– –a half century of progress. JOM-J MET. 48(1): pp. 14-21

[44] Kapusta JPT. JOM World Nonferrous Smelters Survey, Part I: Copper. JOM. 2004;56(7): pp. 21-27

[45] Inco: http://www.inco.com.tr/about.php 2017

[46] Shibasaki T, Hayashi M, Nishiyama Y. 1993. Recent operation at Naoshima with a larger Mitsubishi furnace line, in: C. Landolt (Ed.). Extractive Metallurgy of Copper, Nickel and Cobalt (the Paul E. Queneau International Symposium). Volume II: Copper and Nickel Smelter Operations TMS, Warrendale, PA, pp. 1413-1428

[47] Asaki Z, Taniguchi T, Hayashi M. Kinetics of the reactions in the smelting furnace of the Mitsubishi process. JOM. 2001;53(5): pp. 25-27

[48] Wang JL, Chen YZ, Zhang W, Zhang CF. Furnace structure analysis for copper flash continuous smelting based on numerical simulation. Trans. Nonferrous Met. Soc. China. 2013;23(12): pp. 3799-3807

[49] Iida O, Hayashi M, Goto M. Process designs on new smelter projects of the Mitsubishi continuous copper smelting and converting process. In: Proceedings of the Nickel–Cobalt 97 International Symposium, vol. 3, 1997 August 17-20, Sudbury, Canada: pp. 499-511

[50] Fthenakis V, Wang W, Kim HC. Life cycle inventory analysis of the production of metals used in photovoltaics. Renewable Sustainable Energy Rev. 2009;13: pp. 493-517

[51] Veldhuizen H, Sippel B. Mining discarded electronics. Industry and Environment. 1994;17(3): pp. 7-11

[52] Cui J, Zhang L. Metallurgical recovery of metals from electronic waste: A review. J. Hazard. Mater. 2008;158: pp. 228-256

[53] Schaaf M, Gómez Z, Cipriano A. Real-time hybrid predictive modeling of the Teniente Converter. J. Process Control. 2010;20(3): pp. 3-17

[54] Bergh LG, Chacana P, Carrasco C. Control strategy for a Teniente Converter. Miner. Eng. 2005;18 (11): pp. 1123-1126

[55] Valencia A, Rosales M, Paredes R, Leon C, Moyano A. Numerical and experimental investigation of the fluid dynamics in a Teniente type copper converter. Int. Commun. Heat Mass Transfer. 2006;33 (3): pp. 302-310

[56] Najdenov I. Managing copper smelting and rafination processes for improving energy efficiency and economic feasibility. Belgrade (RS): University of Belgrade, Faculty of Technology and Metallurgy; 2013 (In Serbian)

[57] Ullmann F. Ullmann's Encyclopaedia of Industrial Chemistry, 7th ed., Wiley-VCH, Weinheim. 1995; pp. 471-524

[58] Mohagheghi M, Askari M. 2016. Copper recovery from reverberatory furnace flue dust. Int. J. Miner. Process. 2016;157: pp. 205-209

[59] Stanković Ž. Management of technological innovations in metallurgy of heavy non-ferrous metals. Bor: RTB Bor and Mining and Metallurgy Institute Bor; 2000

[60] Diaz C, Landolt C, Luraschi A, Newman CJ. Pyrometallurgy of copper. Volume IV. New York (US): Pergamon Press; 1991

[61] Davidović A, Najdenov I, Husović Volkov T, Raić TK. Induction Furnace without Core: Design, Operating Parameters and Applications. Livarstvo. 2009;48(2): pp. 12-23

[62] Biswas AK, Davenport WG. Extractive Metallurgy of Copper (Third edition). Great Britain: British Library; 1994

[63] Davenport WG, Jones DM, King MJ, Partelpoeg EH. Flash Smelting: Analysis, Control and Optimization. TMS (The Minerals, Metals and Materials Society), Warrendale. PA; 2001

[64] Kazem S, Hadinejad F. PROMETHEE technique to select the best radial basis functions for solving the 2-dimensional heat equations based on Hermite interpolation. Eng. Anal. Boundary Elem. 2015;50: pp. 29-38

[65] Bagherikahvarin M, De Smet Y. A ranking method based on DEA and PROMETHEE II (a rank based on DEA & PR.II). Measurement. 2016;89: pp. 333-342

[66] Figueira J, Greco S, Ehrogott M. Multiple Criteria Decision Analysis, State Of The Art Survey. New York (US): Springer Science; 2005

[67] Abedi M, Torabi SA, Norouzi GH, Hamzeh M, Elyasi GR. PROMETHEE II: a knowledge-driven method for copper exploration. Comput. Geosci. 2012;46: pp. 255-263

[68] Kadziński M, Ciomek K. Integrated framework for preference modeling and robustness analysis for outranking-based multiple criteria sorting with ELECTRE and PROMETHEE. Information Sciences. 2016;352 (C): pp. 167-187

[69] Roy B, Vincke P. Multicriteria analyses: survey and new directions. Eur. J. Oper. Res. 1981;8 (3): pp. 207-218

[70] Vetschera R, de Almeida AT. A PROMETHEE-based approach to portfolio selection problems. Computers & Operations Research. 2012;39 (5): pp. 1010-1020

[71] Peng AH, Xiao XM. Material selection using PROMETHEE combined with analytic network process under hybrid environment. Matererials and Design. 2013;47: pp. 643-652

[72] Tavana M, Behzadian M, Pirdashti M, Pirdashti H. A PROMETHEE-GDSS for oil and gas pipeline planning in the Caspian Sea basin. Energy Econ. 2013; 36: pp. 716-728

[73] Yu X, Xu Z, Ma Y. Prioritized multi-criteria decision making based on the idea of PROMETHEE. Procedia Comput. Sci. 2013;17: pp. 449-456

[74] Zhao H, Peng Y, Li W. Revised PROMETHEE II for Improving Efficiency in Emergency Response. Procedia Comput. Sci. 2013;17: pp. 181-188

[75] Amaral TM, Costa APC. Improving decision-making and management of hospital resources: An application of the PROMETHEE II method in an Emergency Department. Operations Research for Health Care. 2014; 3: pp. 1-6

[76] Veza I, Celar S, Peronja I. Competences-based Comparison and Ranking of Industrial Enterprises using PROMETHEE Method. Procedia Eng. 2015;100: pp. 445-449

[77] Vego G, Kučar-Dragičević S, Koprivanac N. 2008. Application of multi-criteria decision-making on strategic municipal solid waste management in Dalmatia, Croatia. Waste Manage. 20008;28: pp. 2192-2201

[78] Ilić I, Bogdanović D, Živković D, Milošević N, Todorović B. Optimization of heavy metals total emission, case study: Bor (Serbia). Atmos. Res. 2011;101: pp. 450-459

[79] Savic M, Djordjevic P, Mihajlovic I, Zivkovic Z. Statistical modeling of copper losses in the silicate slag of the sulfide concentrate smelting process. Polish Journal of Chemical Technology. 2015;17 (3): pp. 62-69

[80] Brans JP, Mareschal B. The PROMCALC and GAIA decision support system for multicriteria decision aid. Decision Support Systems.1994;12: pp. 297-310

[81] Anand G, Kodali R. Selection of lean manufacturing systems using the PROMETHEE, Journal of Modelling in Management. 2008;3(1): pp. 40-70

[82] Ishizaka A, Nemery P. Selecting the best statistical distribution with PROMETHEE and GAIA. Comput. Ind. Eng. 2011;61(4): pp. 958-969

[83] Djordjević P, Nikolić D, Jovanović I, Mihajlović I, Savić M, Zivković Ž. Episodes of extremely high concentrations of SO2 and particulate matter in the urban environment of Bor, Serbia, Environ. Res. 2013;126: pp. 204-207

[84] Nikolić I, Jovanović I, Mihajlović I, Miljanović I. 2015. Analysis of copper concentrate production by sistemic approach. Copper. 2015;40(2): pp. 33-50

[85] Nikolić I, Milošević I, Milijić N, Mihajlović I. Impact on the environment on selection of adequate technology for the copper smelting, Environmental awareness as a universal European Value 2016, Bor, University of Belgrade, Technical Faculty in Bor, Engineering Management Department (EMD). 2016; pp. 168-177

[86] RTB Bor: http://rtb.rs

Impact Assessment of Eight Year Application of the SOL Safety Event Analysis Methodology in a Nuclear Power Plant

Lajos Izsó[a], Miklós Antalovits[a], Sándor Suplicz[b]

[a]Department of Ergonomics and Psychology, Budapest University of Technology and Economics, Magyar tudósok körútja 2, 1117 Budapest, Hungary, izsolajos@erg.bme.hu, antalovits@erg.bme.hu

[b]Ágoston Trefort Centre for Engineering Education, Óbuda University, Népszínház utca 8, 1081 Budapest Hungary, suplicz.sandor@tmpk.uni-obuda.hu

Abstract: The objective of this paper was to summarize the measurable indicators of the impact of eight year application of the SOL safety event analysis methodology in the period of 2007 – 2015 in a nuclear power plant in Hungary. The theoretical framework of this paper consists of the (1) "Swiss-Cheese Model", (2) the "socio-technical system model", (3) the organizational learning approach, and (4) the concept of safety culture. The selected broad spectrum of methods corresponds to the approach of progressing from the actual state of the safety culture – via covering the SOL related experiences and opinions of the most involved employees, middle and top managers, and training experts as well – towards the whole community of the NPP. As the results of widespread questionnaire surveys, focus group interviews and anonymous intranet-based inquiry methods it can be stated that the overwhelming majority of the respondents considered the application of the SOL methodology as useful and supporting the safety-related organizational learning. It was also found, however, that in the respondents' opinion the utilization of the – otherwise correct and deeply penetrating – results of SOL analyses is still to be improved.

Keywords: SOL; safety event analysis; nuclear power plant

1 Introduction

1.1 Background

This paper is the second of two related papers providing fundamental information on the experiences gained during applying the SOL safety event analysis methodology in the MVM Paks Nuclear Power Plant Ltd. (hereafter - Paks NPP) in Hungary.

The first paper, entitled '*Factual results of eight year application of the SOL safety event analysis method in a nuclear power plant*', dealt with general factual findings. The goal of the present paper is to present the impact of introducing the SOL methodology on the safety culture of the Paks NPP. The fundamentals of the SOL methodology have already been published elsewhere in many journal articles and books, e.g. refer to [4], [5], and in [3]. Some IAEA (*International Atomic Energy Agency*) and EC (*European Commission*) technical documents also review the SOL, refer, e.g. to [6], [8] and [2]. More details about applying the SOL at Paks NPP can be found in our first paper.

1.2 Research Questions

Based on the demands from the top management of the NPP and also on our earlier experiences, the following main research questions have been selected for studying.

- What is the general opinion of the employees about the usefulness of the SOL methodology in this NPP?
- What are the added values of SOL analyses compared with the routine event investigation methodology in the opinion of the employees?
- Who are the main beneficiaries of SOL analyses in the opinion of the employees?
- To what degree do they consider the utilization of the results of SOL analyses satisfactory?
- How all the opinions above depend on the position and professional areas of the respondents?

2 Methods

2.1 Approach

The theoretical framework of this paper consists of (1) the "Swiss-Cheese Model", (2) the "socio-technical system model", (3) the organizational learning approach, and (4) the concept of safety culture. Since (1), (2) and partly (3) accident causation models are touched in our first paper, here we focus partly on (3) and mainly on (4).

We accept the definition of the [1]: organizational learning is an organization-wide continuous process that enhances its collective ability to accept, make sense of, and respond to internal and external change. All organizations learn, in the

sense of adapting as the world around them changes. The big differences between organizations are, however, that some organizations are faster and more effective learners. Concerning the operational teams, as smaller units of organizations, similarly, the big safety-relevant differences between them are that some teams are better cooperating and more adaptive learners (refer to [15] and [16]).

The term "safety culture" was first introduced by the International Nuclear Safety Advisory Group (INSAG) in 1986 as a response to the Chernobyl disaster. The INSAG later introduced the presently used following definition of safety culture in its [10, page 4] report: "Safety Culture is that assembly of characteristics and attitudes in organizations and individuals which establishes that, as an overriding priority, nuclear plant safety issues receive the attention warranted by their significance." For other aspects of safety culture in nuclear installations refer to [9] and [10], [11].

2.2 Applied Methods

2.2.1 Studying the Results of Safety Culture Assessments

An analysis of documents and reports on different safety culture assessments since 1999 carried out in the Paks NPP using the basic questionnaire-based assessment methodologies proposed by the IAEA [7] has been completed.

2.2.2 Questionnaire Survey among Employees Who had already Taken Part in SOL Analyses

Within our whole target period of interest (2007 March – 2015 May, totaling up to about 8 years) there were four two-year sub-periods (2007-08, 2009-11, 2011-13, 2013-15) for each of which a separate SOL meta-analysis was carried out thus covering altogether 27 individual SOL event analyses.

In this survey the participants of SOL analyses and meta-analyses had been asked to weight the significance of problems identified during SOL analyses. The aim of this survey was to map the opinions of all the employees who had already taken part in SOL analyses concerning the most serious actual safety-related problems as a function of time (in terms of the data gained in the four subsequent meta-analyses). These opinions were considered as important reflections of the impact of the application of the SOL on the safety culture of the NPP.

These opinions were asked on a 3-point "seriousness" weighting scale, the anchor points of which were defined as follows:

(1) Not real problem, or already solved

(2) Problem solving in progress

(3) No progress made

Before each meta-analysis all the employees who had already taken part in SOL analyses until that time were asked to select those 5 problems that they judged as most serious, and later to weight them on the "seriousness" scale presented above. These individual weights were summed up to each problem and finally all the problems were arranged into a list of descending total weight order. The first 15 problems in this list were taken as the most "serious". Similarly, the participants of the actual SOL meta-analysis also were asked to select those 5 problems that appeared to them to be the most serious ones, and later they also had to weight these on the "seriousness" scale. These individual weights were again summed up to each problem and finally the problems were arranged into a list in descending total weight order. The first 25 problems in this list were taken as the most "serious".

Table 1
The process of identifying and weighting the safety-related problems based on SOL analyses

Time sub-period of SOL analyses	**Number of performed SOL analyses**	**List of earlier 15 most „serious problems compiled by**	**List of present 25 problems compiled by**	**Unified list of 40 most „serious problems re-weighted by**
2007-2008 1st meta-analysis	8	none (as still there were no "earlier" problems)	the participants of the 1st SOL meta-analysis performed in 2008	none (as still there were no "earlier" problems)
2009-2011 2nd meta-analysis	8	the participants of all SOL analyses performed in the period of 2007-2011	the participants of the 2nd SOL meta-analysis performed in 2011	the participants of the 2nd SOL meta-analysis performed in 2011
2011-2013 3rd meta-analysis	6	the participants of all SOL analyses performed in the period of 2007-2013	the participants of the 3rd SOL meta-analysis performed in 2013	the participants of the 3rd SOL meta-analysis performed in 2013
2013-2015 4th meta-analysis	5	the participants of all SOL analyses performed in the period of 2007-2015 (194 persons)	the participants of the 4th SOL meta-analysis performed in 2015	the participants of the 4th SOL meta-analysis performed in 2015

Finally, the two lists were unified and re-weighted by the participants of the actual SOL meta-analysis resulting in the unified and re-weighted list of the 40 most serious problems. The weights were summed up separately for managers (for group leaders and above) and subordinates (for employees below the group leader position). In order to follow with attention the changes of seriousness of the perception of safety-related problems as a function of time, the overlaps of these lists belonging to different periods of times were studied.

2.2.3 Focus Group Interviews with Middle Managers

In the frame of the 2015 year SOL meta-analysis 12 opinion-shaper middle managers were participating in a focus group discussion to find answers to the following questions:

(1) What are the added values of SOL analyses for the participants and for the Paks NPP Company compared with the routine PRCAP (*Paks Root Cause Analysis Procedure*) event investigation methodology?

(2) Who are the main beneficiaries of SOL analyses?

(3) Is it expectable during all SOL analyses that the "truth" will come out concerning the given event?

(4) Are the SOL analyses well-documented?

Text analysis method was also used to summarize the different opinions.

2.2.4 Interviews with Top Managers about the Impact of Applying the SOL Methodology on the Safety Culture

The interviewees were the 13 top managers (directors, heads of main departments and heads of departments) who had already taken part in SOL analyses, and these interviews included giving scaled/numeric answers along the following dimensions

- SOL usefulness (on five-point scale)
  (*To what degree do you judge the SOL methodology useful?*)
- SOL notoriety among top managers (on five-point scale)
  (*To what degree do you judge the SOL methodology known among top managers?)*
- SOL notoriety among the wider managerial group (on five-point scale)
  (*To what degree do you judge the SOL methodology known among middle and lower level managers?*)
- SOL notoriety in the power plant as a whole (on five-point scale)
  (*To what degree do you judge the SOL methodology known among all the employees?)*

- SOL acceptance among top managers (on five-point scale)
  (*To what degree do you judge the SOL methodology accepted among top managers?)*
- SOL acceptance among the wider managerial group (on five-point scale)
  (*To what degree do you judge the SOL methodology accepted among middle and lower level managers?*)
- SOL acceptance in the power plant as a whole (on five-point scale)
  (*To what degree do you judge the SOL methodology accepted among all the employees?)*
- Percentage of SOL problem descriptions that got to decision makers (in percentage)
  (*Percentage of SOL problem descriptions that got to decision makers? %*)
- Percentage of SOL based measures taken (in percentage)
  (*Percentage of that SOL based measures that were taken? %*)
- Percentage of realization of SOL based measures (in percentage)
  (*Percentage of SOL based measures that were realized? %)*

In addition, the respondents gave also corresponding free textual answers.

The numeric answers – as values on ordinal scales – along the different dimensions were processed by statistical methods: relevant descriptive statistics were calculated and appropriate nonparametric tests were applied. The free textual answers were processed by text analysis.

2.2.5 Questionnaire Survey among Instructors of the Training Center

Since one of the most important and most frequent types of the utilization of the results of SOL event analyses is to train employees in order to avoid in the future the recurrence of certain problems identified by the SOL, the SOL related experiences, opinions and attitudes of the training staff are essential for shaping the safety culture.

Therefore, a questionnaire survey was carried out among instructors of the Training Center with the following main questions:

- How long has the respondent been qualified instructor (years)?
- Form of teaching: basic, drilling, refresher, department level, simulator training, maintenance training, e-learning, other. All answers on 0 (no), 1 (yes) scale
- Degree of knowing SOL methodology (3-point ordinal scale):

  (1) Does not know, only heard about it

  (2) Knows its fundamentals

(3) Has already participated in SOL analysis

- Degree of knowing SOL experiences made public in the portal (4-point ordinal scale):

  (1) Has not visited SOL reports on the portal, because has not been interested

  (2) Has already visited SOL reports on the portal

  (3) Regularly follows with attention SOL reports on the portal

  (4) Regularly follows with attention both SOL reports and normal PRCAP event investigations on the portal

- Considers the present posterior SOL analysis practice as useful: 0 (no), 1 (yes) scale

- Degree of utilizing SOL experiences in teaching practice (4-point ordinal scale):

  (1) Not yet

  (2) Refers to SOL analyses, but does not go into details

  (3) Presents some experiences of SOL analyses as convincing examples

  (4) Studies the experiences of SOL analyses in more details and organically builds them into the teaching as case studies

54 training staff members were directly asked to fill in the questionnaire.

2.2.6 Anonymous Intranet-based Questionnaire Survey about Applying the SOL Methodology

This and the following last method (section 2.2.7) are intranet-based approaches by the help of which it was hoped that a large part of the whole NPP community could be reached. The questionnaire survey finally involved 642 respondents, and asked the following main questions:

- Which directorate do you belong to?
  Production, Maintenance, Technology, Safety, Human Resources, Economic

- To what degree do you know the principles of SOL? (4-point ordinal scale):

  (1) Has never heard about SOL

  (2) Does not know, but has already heard about SOL

  (3) Knows the essence of SOL

  (4) Has already taken part in SOL analysis session

- Do you consider the SOL as useful? 0 (no), 1 (yes) scale

- Why do you consider the SOL as useful? (free text answer)

- Why do you consider the SOL as not useful? (free text answer)
- Do you know concrete measures that were taken based on the SOL?
  0 (no), 1 (yes)
- Comments and proposals concerning the use of SOL: (free text answer)

2.2.7 Log File Analysis of viSitors' Activity Concerning the Results of SOL Event Analyses Available on a Dedicated Portal of the Intranet of the Paks NPP

The number of SOL-related downloads from the portal of the NPP intranet was studied during the latest 15 months of the whole target period of interest.

For the date period of 2014.01.28 – 2015.04.30, the file request statistics were analyzed, as parameters characteristic for visitors' activity concerning the results of SOL event analyses.

3 Results

3.1 Safety Culture Assessments Carried Out by the Paks NPP

All the following assessments in this sub-section reflect the levels of safety culture perceived subjectively by the respondents. Therefore, these can be regarded as "opinions" collected in a methodologically appropriate way, rather than the "real" or "absolute" levels.

In 1998/99 and 2000 two of the authors [12] conducted two safety culture assessments at the Paks NPP based on questionnaires and interviews comparable with the methodology proposed by the IAEA [7] based on 26 sub-dimensions. The first assessment involved 153 subordinates, the second 63 managers. Although there were some minor differences in the results of these two assessments in certain sub-dimensions, the overall level actually was the same in these two samples (77% for the subordinates and 76% for the managers).

Later four more safety culture surveys were completed by the Aon Hewitt method in 2005, 2009, 2013 and 2015, involving also both subordinates and managers [13], [14]. The Aon Hewitt method – which usually produces slightly lower overall percentage levels than the IAEA method – is based on an anonym and voluntary questionnaire survey and the results comprise seven indices expressed in percentages and their average as one main summary index (overall percentage level). These indices (sub-dimensions) are: (1) commitment to safety, (2) procedure usage, (3) conservative decision making, (4) reporting culture, (5) treating unsafe activities and conditions, (6) organizational learning, (7) communication, clear priorities and responsibilities and transparent organization.

Of the above, the (6) organizational learning index directly relates to the values that the SOL also promotes and aims at developing, while all the others also relate to them but only rather indirectly.

It was found that from 2005 via 2009 to 2013 all the main summary indices showed slight increases or remained at constant levels in the range of 72-74%. Since the main summary indices of our 1998/99 and 2000 safety culture assessments by the IAEA method fit to this series, it can be taken that in the whole range of about 2000 to 2013 the level of safety culture increased only slightly or stayed constant. From 2013 to 2015, however, there was a radical 10% increase in the main index (from 74% to 84%). In this period the SOL related organizational learning index (sub-dimension) also jumped from 69% to 78%. This marked increase, among many others, may – or may not – be attributed to the influence of introducing the SOL method and disseminating its results in the NPP.

Studying the differences in the main index between directorates, positions and the time spent employed at the NPP revealed the following relationships:

- Concerning directorates, the highest differences were between the opinions of employees belonging to the Production Directorate and to the Maintenance Directorate. The opinions of the production staff were much more positive than that of the maintenance staff, who most often are facing with different unforeseen problems and deficiencies.
- Concerning positions, the highest differences were between the opinions of middle managers and the operative managers. The opinions of middle managers were much more positive than that of operative managers. Based on the results of a targeted focus group session, this effect could probably be interpreted by the fact that the operative managers are continuously working "between two fires": they are responsible for operative work, but simultaneously they also have to strictly observe all the related safety rules.
- Concerning the time spent at the NPP, the highest differences were between the opinions of most newer and most senior employees. The opinions of newer employees were very positive, but this value gradually decreased with the years spent at the NPP. Our interpretation of this finding is that while the most newer employees have an idealistic, a little bit still naïve, unrealistically positive overall picture about safety, the more experienced senior employees, on the contrary– based on their own occasional frustrations – might be slightly disappointed and may have an even more pessimistic view than the reality. One possible way to prevent this harmful mental process is to systematically and continuously show via many examples of how certain safety critical issues (identified e.g. by the SOL) are treated to forestall their serious consequences.

3.2 Questionnaire Survey among Employees Who had already Taken Part in SOL Event Analyses

For identifying and weighting the safety-related problems by employees who had already taken part in SOL event analyses the process presented in sub-section 2.2.2 (especially in Table 1) was applied. Since these results are very voluminous, here we only declare that the resulted rank order of the identified concrete particular problems was very useful for the managers and contributed to deeper understanding of the actual – both obvious and latent – risks. Three illustrating examples from this long ordered list (without their corresponding seriousness weights):

- The financial and human resources are not always matched to the tasks.
- Technological changes are often carried out under strong time pressure.
- The NPP cannot always properly provide the contractors and sub-contractors with the necessary training.

Another general experience was that although there is a moderate correlation between the weights given by the managers and subordinates (Spearman correlation coefficient r_S = 0,718; p=0,003), there are certain problems the seriousness of which are quite differently judged by managers than by subordinates. Analyzing and interpreting the details of this finding has also been proven useful for better understanding the managers' and subordinates' view.

3.3 Focus Group Interviews with Middle Managers

The interviewees were 12 opinion-shaper middle managers, who – in the frame of a focus group discussion as part of the 2015 year SOL meta-analysis – gave the following main groups of answers to the four predefined broad questions.

(1) What are the added values of SOL analyses for the participants and for the NPP compared with the routine PRCAP event investigation methodology?

- The SOL takes into account many aspects of events simultaneously, while PRCAP event investigations cannot do that.
- A big advantage of the SOL is that the participants can get to know other professional areas and their representatives, which is not true for PRCAP.
- The PRCAP practically always reveals equipment, system or human failures, while the SOL can identify organizational, leadership and procedure-related problems and thus can delve deeper into underlying causes.

(2) Who are the main beneficiaries of SOL analyses?

- First of all, the participants themselves, because they can get to know other professional areas and their representatives in more details.

- Provided that the utilization of results via managerial commitment will further be built and sustained, beneficiaries could be the wider professional areas.
- Since the problems identified by the SOL are mostly global NPP level malfunctions, the NPP as a whole could benefit from it.

(3) Is it expectable during all SOL analyses that the truth will come out concerning the given event?

- Although there have already been several cases in which the climate of SOL analyses was not honest, the truth usually still comes out.
- If the participants arrive at the SOL analysis „prepared", the chance is smaller.
- If the SOL analysis is conducted properly, the truth must come out, independently of the participants.

(4) Are the SOL analyses well-documented?

- The SOL documentation is appropriate, but in order to better utilize the results it is necessary to compile shorter targeted abstracts and circulate them.
- The reports contain the opinions of the participants, which is not necessarily the truth. Since even if the participants understand the problems correctly, the solution of them is already not their competence.
- Therefore, later an after-processing of the results would be necessary, and based on it, new, corrected documents have to be produced.
- The SOL documentation would be even better, if such experts were always participating in SOL analyses who can identify process-level problems.

3.4 Interviews with Top Managers about the Impact of Applying the SOL Methodology on the Safety Culture

The 13 top managers were interviewed by the authors. Parts of the results of the interviews were given in the form of scores on five-point scales and in estimated percentages, which made possible some simple quantitative statistical data processing. The main results are as follows.

Table 2

The main descriptive statistics of the scaled interview answers given by the 13 top managers about the impact of applying the SOL methodology on the safety culture

Interview question(answers on five-point scales)	**Mean**	**SD**
To what degree do you judge the SOL methodology useful?	4,15	0,899
To what degree do you judge the SOL methodology known among top managers?	4,92	0,277
To what degree do you judge the SOL methodology known among middle and lower level managers?	4,35	0,747

To what degree do you judge the SOL methodology known among all the employees?	2,96	0,967
To what degree do you judge the SOL methodology accepted among top managers?	4,27	0,599
To what degree do you judge the SOL methodology accepted among middle and lower level managers?	3,77	0,927
To what degree do you judge the SOL methodology accepted among all the employees?	3,33	0,888
Percentage of SOL problem descriptions that got to decision makers? (%)	68,08	27,729
Percentage of SOL based measures that were taken? (%)	42,46	31,853
Percentage of SOL based measures that were realized? (%)	30,38	26,037

Already these descriptive statistics in themselves are very informative. Emphasizing the most important ones, it can be seen, that respondents judged that:

- the application of SOL methodology is very useful (mean = 4,15; SD = 0,90),
- among top managers the SOL methodology is already almost perfectly known (mean = 4,92; SD = 0,28), and its acceptance is also rather high (mean = 4,27; SD = 0,60),
- among all the employees the notoriety and acceptance of the SOL methodology is already much lower (mean = 2,96; SD = 0,967 and mean = 3,33; SD = 0,89; respectively),
- the percentage that the problems identified by the SOL get to decision makers is 68% on the average, and these opinions have a relatively high dispersion (SD = 27,73),
- the percentage that the problems identified by the SOL result in measures is 42% on the average, and these opinions have an even higher dispersion (SD = 31,85),
- the percentage that the measures taken for solving the problems identified by the SOL is also realized is 30% on the average, and these opinions have again a very high dispersion (SD = 26,04).

Concerning the latest two points, these percentages could be increased by the continuous commitment of top managers to utilize the SOL results in practice.

Statistical comparison of the interview answer scores of the 13 interviewees by the Kruskal-Wallis test using the directorates where the interviewees belonged to (Production, Maintenance, Technology, Safety, Economic) as grouping variable, resulted in no significant differences. The comparison by the positions as grouping variable, however, produced two significant differences by the pair-wise Mann-Whitney test (Figure 1).

As Figure 2 shows, there was also a significant correlation between the perceived acceptance of the SOL methodology by all the employees and its perceived usefulness.

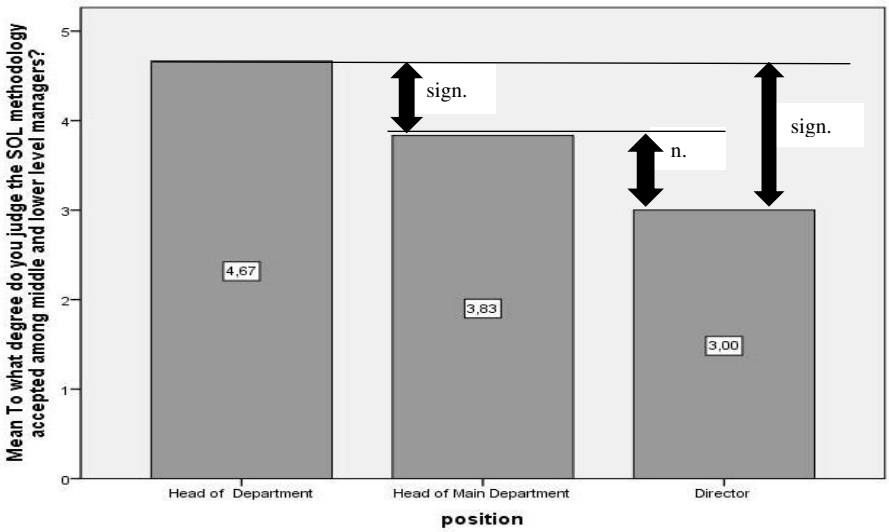


Figure 1

Acceptance of the SOL methodology among middle and lower level managers as a function of the position of the interviewees. The differences indicated by "sign." are significant by the Mann-Whitney test. Since the variable "To what degree do you judge the SOL methodology accepted among middle and lower level managers?" presented on the vertical axis is measured on an ordinal scale, its means are displayed only for visual demonstration purposes.

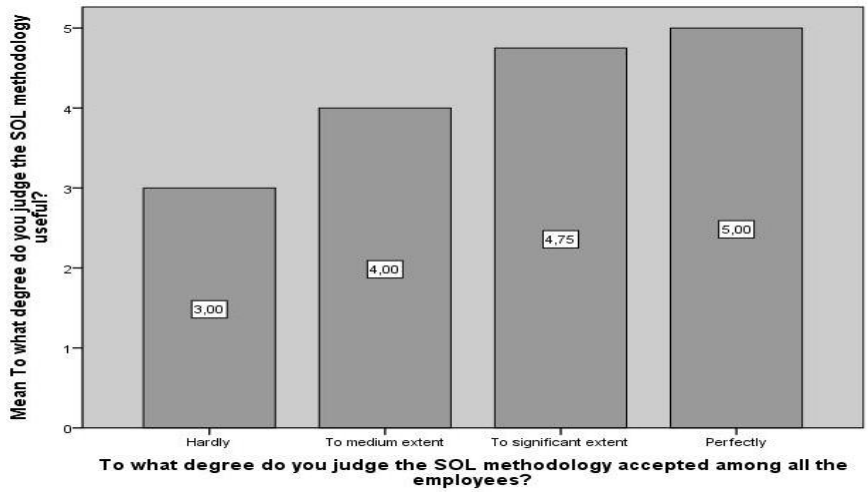


Figure 2

The degree of perceived usefulness as a function of the perceived acceptance of the SOL methodology among all the employees (Spearman correlation coefficient $r_S = 0{,}715$; $p=0{,}009$). Since the variable "*To what degree do you judge the SOL methodology useful*?" presented on the vertical axis is measured on an ordinal scale, its means are displayed only for visual demonstration purposes.

Of the free text answers the following two were both the most important and most frequently mentioned: (1) "*Very useful methodology, but the utilization of results is still incomplete*"; (2) "*SOL really should be about learning and not finding someone to blame*".

3.5 Questionnaire Survey among Instructors of the Training Center

54 instructors filled in the questionnaire. Concerning the respondents' opinion about the usefulness of the present SOL analysis practice the majority (83,3%, 45 persons) answered "yes", the minority (16,7%, only 9 persons) answered "no".

The mean time spent in the "*qualified instructor*" position by this 16,7% minority was 10,11 years, which is significantly longer by the Mann-Whitney test (p=0,045), than that of the majority (6,96 years), who consider the present SOL practice as useful. This finding is interpreted by the well-known experience that the older instructors are less open for such new approaches like the SOL.

Among instructors who consider the present SOL practice as useful, the percentage of those who are conducting refresher training is 80%, while this percentage among instructors who consider the present SOL practice as not useful is only 44%. This difference is significant by the Mann-Whitney test (p=0,028). We interpret this result by taking into account that the refresher training is focused on actual daily problems for which the SOL analyses usually provide support. In addition, for instructors conducting refresher training to be familiar with the newest SOL results is a definite expectation.

Figure 3 shows the variable „*Form of instruction: department level*" as a function of the variable "*Degree of knowing SOL experiences made public in the portal*" on ordinal scale.

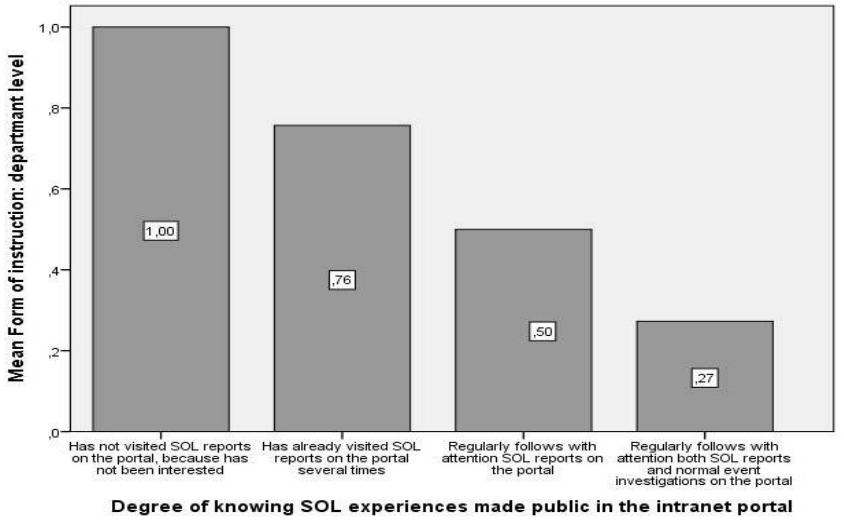


Figure 3
The mean of the variable „*Form of instruction: department level*" on dichotomous no(0)/yes(1) scale as a function of the variable "*Degree of knowing SOL experiences made public in the portal*" on ordinal scale. Both the Spearman correlation coefficient

(r_S = - 0,451; p=0,001) and the Kruskal-Wallis test (p=0,011)indicates a significant relationship.

It can be seen that in the category of lowest level knowledge about SOL experiences (*1: Has not visited SOL reports on the portal, because has not been interested*) there are exclusively instructors conducting department level training (100%), and their ratio is gradually decreasing with the growing levels of knowledge about SOL experiences (75,7%, 50%, 27,3%, respectively). We interpret this result by taking into account that the targets of department level instruction are relatively local problems and the instructors conducting department level instruction do not belong to the Training Center. There is no formal expectation toward them to follow with attention the newest SOL results, therefore only a smaller part of them was interested enough to be informed about recent SOL results.

Figure 4 shows the variable "*Form of instruction: simulator training*" as a function of the variable "*Degree of knowing SOL experiences made public in the portal*" on ordinal scale.

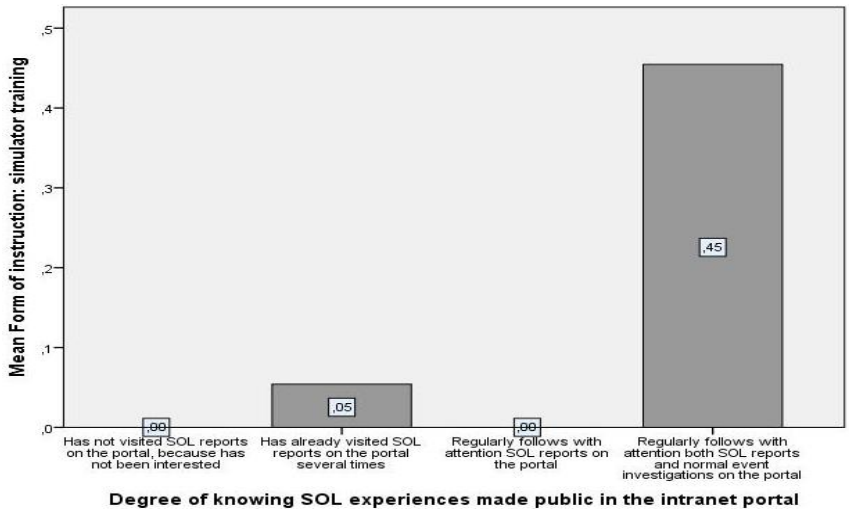


Figure 4

The mean of the variable "*Form of instruction: simulator training*" on dichotomous no(0)/yes(1) scale as a function of the variable respondents' "*Degree of knowing SOL experiences made public in the portal*" on ordinal scale. The Kruskal-Wallis test (with some additional procedures) indicates (p=0,005) that the proportion of simulator training instructors is significantly higher in the category of "*Regularly follows with attention both SOL reports and normal event investigations on the portal", than in other categories*.

It means that in the category of lowest level knowledge about SOL experiences (1*: Has not visited SOL reports on the portal, because has not been interested*) there are no instructors conducting simulator training at all (0%), and their ratio remains about zero with the growing levels of knowledge about SOL experiences till level 3 (3: *Regularly follows with attention SOL reports on the portal*). In the category of highest level knowledge about SOL experiences (4: *Regularly follows with attention both SOL reports and normal event investigations on the portal*), however, there are instructors conducting simulator training in a relatively high percentage (45,5%). This case is quite the contrary of the instructors conducting department level instruction, since toward the simulator instructors being informed concerning the latest SOL results is a definite expectation.

3.6 Anonymous Intranet-based Questionnaire Survey about Applying the SOL Methodology

Altogether 642 employees filled in the questionnaire, of which 489 respondents knew – or at least have already heard about – the SOL methodology (76%).

The distribution of all the 642 respondents along the directorates was the following: Production (222), Maintenance (140), Technology (138), Safety (64), Economic (51), and Human Resources (27).

The distribution of all the 642 respondents along the "*Degree of knowing SOL*" was: "*Has never heard about SOL*" (153), "*Does not know, but has already heard about SOL*" (170), "*Knows the essence of SOL*" (232), "*Has already taken part in SOL analysis session*" (87). Since this intranet-based questionnaire survey was anonymous, there could be certain overlaps with the other types of surveys applied and therefore cannot be taken as involving a strictly independent sample. However, because of its large sample size (642 persons); we are convinced that this is still a very valuable source of information.

The main results are summarized in the following.

Of the 489 employees who knew – or at least have already heard about – the SOL methodology 463 gave answers about the usefulness of the methodology. From these 463 persons 419 (90,5%) considered it useful and only 44 persons (9,5%) considered it not useful.

Near half of the respondents, (49.5%: 227 out of 458) knew cases where safety measures were taken based on the results of SOL analyses. The following figure provides the frequency distribution of the altogether 869 mentions along the four identified mentioning categories.

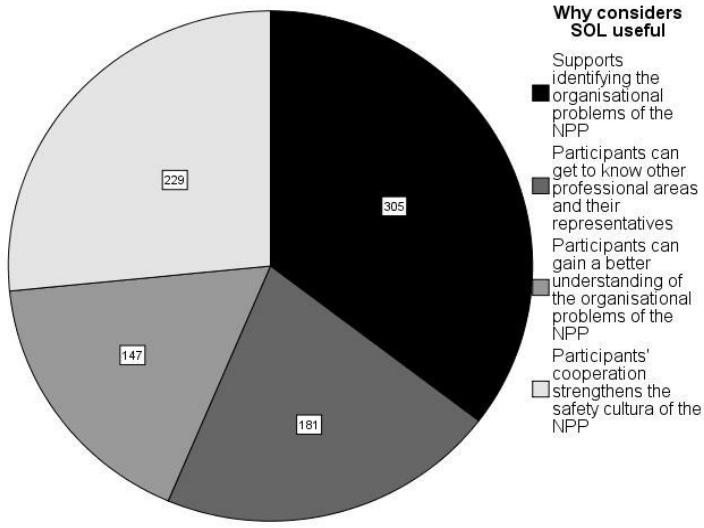


Figure 5

The distribution of the 862 mentions of the 419 respondents who considered the SOL methodology useful along four mentioning categories. Since multiple mentions were allowed the number of mentions is greater than the number of respondents.

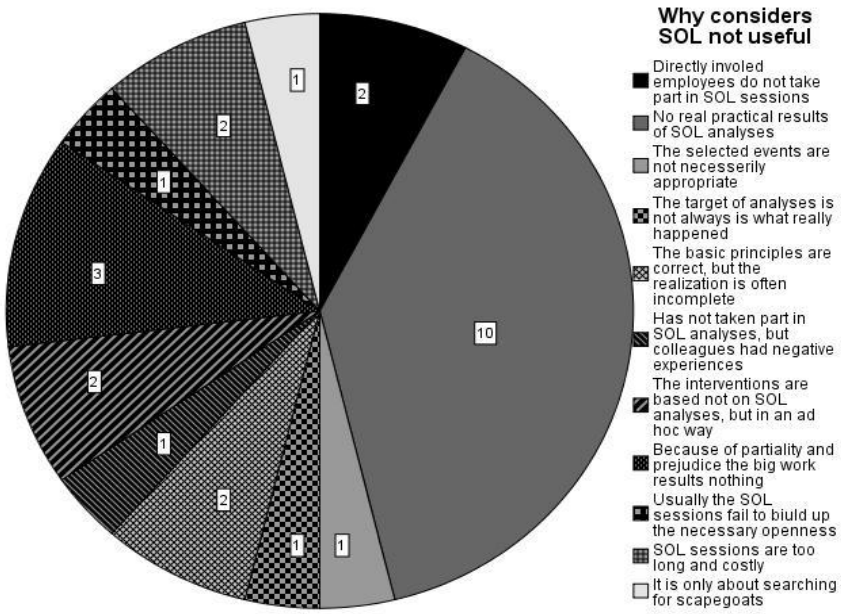


Figure 6

The distribution of the 26 mentions of the 44 respondents who considered the SOL methodology not useful along eleven mentioning categories. Although multiple mentions were allowed the number of mentions is lower than the number of respondents because the majority of these respondents gave no answer at all.

The following table summarizes the most frequent free text comments in descending order.

Table 3

The free text comments mentioned at least two times and their frequency of mentioning

The comments and proposals	**Frequency**
Very useful methodology, but the utilization of results is still incomplete.	8
SOL really should be about learning and not finding someone to blame.	6
The measures taken are often merely formal.	5
Big advantage that the SOL analysis goes deep.	5
A special education would be necessary about the SOL.	4
The SOL analysis should be a kind of "judge" that decides in debates.	3
Too long and costly.	2
The SOL results should be parts of future procedures.	2
Despite the 3 days duration, a SOL analysis is too intense, demanding and stressful.	2

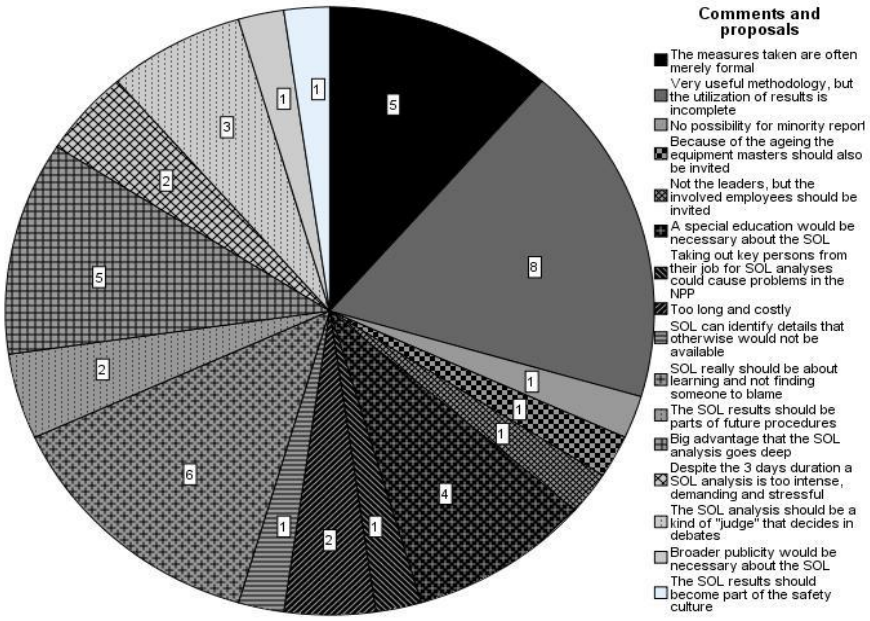


Figure 7

The comments and proposals of the 489 employees who knew – or at least have already heard about – the SOL concerning the SOL methodology

3.7 Log File Analysis of Visitors' Activity concerning the Results of SOL Event Analyses

The number of SOL-related downloads for the period of 2014-01-28 –2015-04-30, was first analyzed with the temporal resolution of one week, but as there was found no tendency along the time, only summary statistics are presented here in a simplified form and using English intranet link titles instead of the original Hungarian.

Table 4

The number of SOL-related downloads for the date period of 2014-01-28 – 2015-04-30.
(Only links downloaded more than20 times are indicated)

No	Intranet link	Number of file downloads
1	production_experiences/sol analyses	561
2	production_experiences/sol analyses/2015_1_sol_s31412.docx	120
3	production_experiences/sol analyses/2014_1_sol_s21321.docx	103

4	production_experiences/sol analyses/2014_2_sol_s11413j.docx	101
5	production_experiences/sol analyses/2014_3_sol_b31403j.docx	58
6	production_experiences/sol analyses/sol analyses content.doc	54
7	production_experiences/sol analyses/2013_meta_sol_report.docx	32
8	production_experiences/sol analyses/2007_1_sol_s30614.doc	23

If we consider here only the number of file downloads greater than 100, we receive the following files as a kind of top list:

- 2015_1_sol_s31412.docx,
- 2014_1_sol_s21321.docx,
- 2014_2_sol_s11413j.docx.

During the period of 2014.01.28 –2015.04.30, three SOL analyses were completed and these three files just contain the reports about them. It is obvious that employees are interested mostly in the results of most recent SOL analyses.

It is another question that downloads slightly greater than 100 during 15 months can or cannot be considered a significant number compared to the altogether 800 – 1000 potentially involved employees.

Our proposal for increasing the number of downloads is to arrange and organize the SOL related materials on the portal separately for the production, maintenance, technology and safety interest-groups.

4 Discussion

First of all, it is essential to emphasize that all the production and safety parameters and characteristics of the Paks NPP are excellent on internationally accepted absolute scales. From this it follows that certain results of this research that may appear to be not so favorable – actually there are hardly any – can only be interpreted as relatively negative that still may well be quite positive on the relevant absolute scale.

The background and reference points of the research were the results of safety culture assessments completed in the period of 2000 – 2015.

The selected broad spectrum of methods has made possible to satisfactorily answer the research questions as follows.

- The general opinion of the employees about the usefulness of the SOL methodology in this NPP was definitely very positive: the big majority of all the respondents (employees who had already taken part in SOL event analyses, middle and top managers, instructors of the Training Center, anonymous employees who filled in the intranet-based questionnaires) considered the present SOL analysis practice as useful.
- The main added values of SOL analyses compared with the routine event investigation methodology as identified by the employees were
  - the SOL can identify organizational, leadership and procedure-related problems,
  - the SOL can take into account many aspects of events simultaneously,
  - the participants can get to know other professional areas and their representatives.
- The main beneficiaries of SOL analyses as identified by the employees were
  - the participants themselves,
  - wider professional areas,
  - directorates and the NPP as a whole(provided that the utilization of results will be improved).
- Almost all the respondents stated that the degree of the utilization of the results of SOL analyses is still not quite satisfactory. This is the relatively weakest point of the present application practice of the SOL. This result, however, at the same time set the course of improving the present practice by increasing the efficiency of utilization of the results.
- The bigger part of the opinions above markedly depended on the position and professional areas of the respondents.

In addition to these answers to the predefined research questions, we consider the many revealed finer details (not presented here) concerning the present SOL practice also as valuable results that support deeper understanding the organizational mechanisms and their interactions in this particular NPP.

Finally, since our basic hypothesis was that applying the SOL methodology appropriately increases the level of the safety culture, consider now the results of safety culture assessments. There is a clear coincidence of introducing the SOL in 2007 and the slightly, but continuously improving level of the safety culture from about 2010 (including the radical 10 % increase in the main index from 2013 to 2015).

Based on our empirical data presented in this paper it cannot be proven, of course, that there is a causal relationship behind this coincidence. However, it cannot be rejected either. We have good reasons to believe that introducing and continuously applying the SOL methodology in an appropriate way has been an important factor that has greatly contributed to improving the safety culture.

By the very nature of the NPP organizations all over the world – not considering here if it is good or not – efforts for keeping or increasing the level of safety culture are usually invested also in the frames of different safety campaigns. We are convinced that applying the SOL methodology correctly can also be regarded as a kind of effective and permanent – or at least long-lasting – safety culture campaign.

Acknowledgements

The authors first of all would like to acknowledge the support from the directors of Paks NPP, namely from Gábor Volent safety director, János Cziczer production director, Zoltán Takács maintenance director and Géza Pekárik technology director. We are also indebted to the 13 top managers for agreeing to being interviewed by us, and furthermore we would like to thank all the 642 anonymous respondents who filled in our intranet-based electronic questionnaire.

Special thanks go to József Gergely of the Safety Directorate who shared with us his valuable experiences gained during all the 27 SOL analyses, to István Kiss, head of Main Department for Training and to Sándor Csuha, head of Psychological Laboratory for preparing and organizing the interviews and also questionnaire surveys with invited and named respondents, and also for collecting other relevant safety related data as well. Finally, we express our gratitude to Dr. Antal Kovács communication director for his useful remarks and finally giving permission for publishing this paper.

References

[1] BusinessDictionary.com, 2016. Organizational Learning. http://www.businessdictionary.com/definition/organizational-learning.html. (Aug. 2016)

[2] European Commission (Ziedelis, S., Noel, M.), 2011. Comparative Analysis of Nuclear Event Investigation Methods, Tools and Techniques. Interim Technical Report. European Commission, Joint Research Centre, Institute for Energy

[3] Fahlbruch, B., Schöbel, M. 2011. SOL - Safety through organizational learning: A method for event analysis. Safety Science, 49 (2011) 27-31

[4] Fahlbruch, B., Wilpert, B. 1997. Event analysis as a problem solving process. In: Hale, A., Wilpert, B., Freitag, M. (Eds.) After the Event from Accident to Organizational Learning. Pergamon, Oxford, pp. 113-130

[5] Fahlbruch, B., Wilpert, B. 1999. System safety - an emerging field for I/O psychology. In: Cooper, C. L., Robertson, I. T. (Eds.) International Review of Industrial and Organizational Psychology, Vol. 14, Wiley, Chichester, pp. 55-93

[6] IAEA, 1991. International Nuclear Safety Advisory Group, Safety Culture, IAEA, Vienna

[7] IAEA, 1994. Assessment of Safety Culture in Organizations Team (ASCOT), IAEA-TECDOC-743, Vienna, Austria

[8] IAEA, 2002. Review of methodologies for analysis of safety incidents at NPPs. Final report of a co-ordinated research project 1998–2001. TECDOC-1278. Vienna, Austria

[9] IAEA, 1998. Developing Safety Culture in Nuclear Activities: Practical Suggestions to Assist Progress, Safety Reports Series No. 11, IAEA, Vienna, Austria

[10] INSAG, 1991. Safety Culture. A report by the International Nuclear Safety Advisory Group. INSAG-4. IAEA, Vienna, Austria

[11] INSAG, 2002. Key Practical Issues in Strengthening Safety Culture. A report by the International Nuclear Safety Advisory Group. INSAG-15. IAEA, Vienna, Austria

[12] Izsó, L., Antalovits, M. 2001. Lessons Learned from a Safety Culture Survey. Tenth European Congress on Work and Organizational Psychology, Prague, Czech Republic. Proceedings, p. S115 (16-19 May, 2001)

[13] Atomerőmű, 2016. 10 percent increase in the safety culture of the Paks NPP (in Hungarian). XXXIX/4.http://www.atomeromu.hu/hu/Documents/Atomeromu_Ujsag/2016/Atomer%C5%91m%C5%B1%202016%2004.pdf (April 2016)

[14] TEIT HÍREK 2016. 10 percent increase in the safety culture of the Paks NPP (in Hungarian) http://teit.hu/wp-content/uploads/2016/06/teit_2016_m%C3%A1jus.pdf (May 2016)

[15] Juhász, M., Soós, J. K. (2011) Human Aspects of NPP Operator Teamwork, Nuclear Power - Control, Reliability and Human Factors, Dr. Pavel Tsvetkov (Ed.), ISBN: 978-953-307-599-0, InTech, DOI: 10.5772/17046. Available from: http://www.intechopen.com/books/nuclear-power-control-reliability-and-human-factors/human-aspects-of-npp-operator-teamwork

[16] Takács, V., Juhász, M. (2018). Adaptation and Cognition of High-Risk Environment Teams in an Input-Mediator-Outcome Framework. *Periodica Polytechnica Social and Management Sciences,* [S.l.], 2017. ISSN 1587-3803. https://pp.bme.hu/so/article/view/10219

Dynamic Motion Planning Algorithm for a Biped Robot Using Fast Marching Method Hybridized with Regression Search

Ravi Kumar Mandava, Katla Mrudul and Pandu R Vundavilli

School of Mechanical Sciences, IIT Bhubaneswar, Bhubaneswar, Odisha, India-752050. E-mail: rm19@iitbbs.ac.in, km16@iitbbs.ac.in, pandu@iitbbs.ac.in

Abstract: In the past few years, studies of biped robot locomotion and navigation have increased enormously due to its ease in mobility in the terrains that are designed exclusively for the humans. To navigate the biped robot in static and dynamic environments without hitting obstacles is a challenging task. In the present research, the authors have developed a hybridized motion planning algorithm that is, fast marching method hybridized with regression search (FMMHRS) methodology. In this work, initially the fast marching algorithm has been used to observe the environment and identify the path from start to final goal. Later on, the regression search method is combined with the fast marching method (FMM) algorithm to optimize the path without hitting any obstacles. The main objective of the present research work is to generate the path for both the static and dynamic scenarios in simulation and in a real environment. To conduct the testing of the proposed algorithm, the authors have chosen an 18-DOF two legged robot that was developed in our laboratory.

Keywords: Biped robot; static and dynamic environment; fast marching method; regression search

1 Introduction

Path planning plays an important role in the navigation of autonomous vehicles and assisted systems. But, the significant property in this field is that the path planning is developed to satisfy the non-holonomic constraints raised due to its motion. During initial stages of research on path planning, investigators had only considered the length of the path as the major cost, and majority of them were worked extensively to obtain an efficient method, that can generate a collision free path. In general, path planning for mobile robots can be categorized into various classes [1]. Roadmap based methods were used to extract the network representation from the environment and then they apply graph based search algorithms to determine the path. Exact cell decomposition methods were used to

construct the non-overlapping regions that cover free space and encode cell connectivity in graph. The approximate cell decomposition method was similar to the exact cell decomposition method, but the cells were assumed to have a predefined shape and it did not exactly cover the entire free space. The potential field method [2], which is different from the other methods in which the robot was assumed to be a point robot which was moving under the influence of attractive forces generated between the goal and the pushing away from the obstacles due to the influence of repulsive forces generated between the obstacles and robot.

The essential requirement for solving the motion planning problem is the creation of appropriate terrain/environment. Once the environment is created, motion planning algorithms can be implemented in an effective manner. This section presents a brief overview of the different types of path planning methods. The available path planning algorithms are categorized into two types [3]: The first category deals with the classic approaches and the second one is focused on heuristic approaches. In classical approaches, the algorithms are designed to calculate the optimal solution, if one exists, or to prove that no feasible path exists. These algorithms are generally very expensive, computationally. But in heuristic approaches, the algorithms are anticipated to search for good quality solutions with in a short time. However, heuristic algorithms can fail to determine the good solution for a difficult problem. There exit few variations of classical methods, such as cell decomposition, roadmap, artificial potential field and mathematical programming to generate the path for the mobile robots. These methods alone and along with their combinations are often used to develop more successful paths. In the roadmap approach [4], feasible paths were mapped onto a network and searched for the desired path. However, the searched path was limited to a network and those path planning algorithms become a graph based search algorithm. Moreover, some of the researchers had developed some well-known roadmaps after using visibility graphs, voronoi diagram [5] and sub–goal networks. In [6], the visibility graph algorithm was used to calculate the optimal path between the start and goal points. In that approach, the authors did not consider the size of the mobile robot, and the mobile robot was moved very close to the vertex of the obstacles. However, the computational time for planning the path using the above method was too long. Later on, researchers developed various methods, such as the voronoi diagram [7] and sub-goal network [8] algorithms that were performed in a better manner when compared with the visibility graph.

In addition to the above approaches, Cai and Ferrai [9] proposed the cell decomposition method, which was the simplest method for planning the path for mobile robot, but the algorithm was inefficient in terms of planning time and managing the computational memory according to the cell size. However, the hybridization of roadmap method with cell decomposition method was seen to provide better efficiency and worked based on the concept of free configuration space (C-space). Due to the lack of adaptation and robustness, the conventional

approaches were not suitable to solve the motion planning problems in dynamic environments. Further, among heuristic approaches, researchers had used A-star algorithm [10] to calculate the shortest path for a given map. In general A-star algorithm was a classical heuristic search algorithm and it was applied on a C-space for planning the path of a robot. The search efficiency of the A-star algorithm was low and the planned path was optimal, when compared with the cell decomposition method. Stenz [11] used D-star algorithm, which was not heuristic, to perform search equally in all directions. When compared with the A-star algorithm, it was found to perform slowly, because this algorithm searches large areas before reaching the goal. In conjuction with, researchers also developed soft computing based approaches for obtaining the optimal path in cluttered environments. In [12], a genetic algorithm was used to obtain the best feasible path after many iterations. It happened due to the complex structure of GA, which requires a huge time to process the data. When dealing with the dynamic environments, this fact lead to the premature convergence while obtaining the optimal solution. To improve the performance of GA, some researchers had suggested different types of optimization algorithms such as, combining genetic algorithm and simulated annealing [13], ant colony optimization [14], particle swarm optimization [15], cuckoo search algorithm [16], invasive weed optimization [17], bacterial forging optimization [18] and firefly algorithm [19]. In addition to the above approaches, some researchers have also used the soft computing approaches, such as fuzzy-genetic algorithm [20] and neural network based algorithms [21] to solve the motion planning problems of biped robots.

Further, Santiago et al. [22] proposed a robust algorithm called Voronoi fast marching method for obtaining the smooth and safe path in a cluttered environment. It worked based on the phenomenon of local-minima-free planner. Lucas et al. [23] developed a fast marching tree using FMM algorithm for obtaining the optimal path in a high dimensional configuration space. Moreover, a novel path planning algorithm was introduced [24] with non-holonomic constraints for a car-like robot. In this approach, the FMM was used to investigate the geometric information of the map, and support vector machine was used to find the information related to the clearance of the obstacles. The FMM was guided by this function to generate the vehicle motion under kinematic constraints. In [25], the authors explained a detailed overview of fast marching method and also recalled the methods that is, FM2 and FM2* developed and used by the same authors in path planning applications. Garido et al. [26] applied the FMM algorithm to simulate the electromagnetic wave propagation. Here, the wave starts from a point and continuous to iterate until it reaches the end point. The generated field had only one global minima point, which was located at the center point. Petres et al. [27] developed anisotropic fast marching method, which was an improved version of FMM with higher computational efficiency than level set method. Further, Song et al. [28] proposed a novel multi-layered fast marching (MFM) method to generate the practical trajectories for the unmanned surface

vehicles while operating in a dynamic environment. To design an optimal path planning algorithm in [29], the authors developed an effective and improved artificial potential field method combined with regression search. Simulataneously, Ravi et al. [30] established a hybridized path planning algorithm for static and dynamic environments. In this work, they used a 3-point smoothing method to generate the optimal path.

The main objective of the present research is to minimize the path length subjected to constraints on different curvature properties. In order to determine the path in the global map, the authors have presented a novel hybridized path planning method (that is, FMMHRS). It is important to note that the developed FMMHRS will help in achieving the shortest path due to the inherent characteristics of regression search. Initially, FMM has been applied to solve path planning problems [31]. In certain scenarios, the path trajectories obtained are not safe because the path is very close to the obstacles. In order to improve the safety of the path trajectories calculated by using the fast marching method, it is possible to give two solutions: The first possibility is to avoid unrealistic trajectories, generated when areas are narrower than the robot. Therefore, the minimum clearance that should be maintained between obstacles and walls is at least half the size of the robot. The second possibility that has been used in this work is to enlarge the distance between walls and objects to a safe distance so that the robot will not collide with an obstacle. Therefore, initially the entire path is generated with the help of FMM from start to final goal. Once the FMM path is generated, it is split into number of equally spaced nodes. Then regression search is initiated on the nodal data by connecting the present node to the next and by checking the clearance distance between the path and obstacle. This procedure is continued until, the robot reaches the target. To the best of the author's knowledge, this combination has not been tried by any researchers in the field of motion planning of biped robots. Moreover, the proposed algorithm is implemented on static and dynamic environments in both computer simulations and in real time environment. Further, an 18-DOF biped robot has been used to tackle the real time situations. The advantages of this method are ease of implementation, speed and quality of the path. Moreover, this method can work in both 2D and 3D environments, and it can also be used in a local or global scale path planning problems.

2 The Fast Marching Method

The FMM is an efficient numerical algorithm developed by Osher and Sethian in 1988, which was used for tracking and modeling the motion of a physical wave front interface. In general, the algorithm has been applied in various research fields including medical imaging [32], computer graphics [33], image processing [34], computational fluid dynamics and computation of trajectories etc.

The wavefront interface can be modeled as a 3D surface or a flat curve in 2D. The FMM calculates the time T that a wave needs to reach every point of the space. The wave can be originated at one point or more than one point at a given time. If it is originated at more than one point, then each source point generates one wave front and all the source points are associated with time $T=0$. In the context of the FMM the authors have assumed that the wave front (Γ) grows by motion in the direction normal to the surface. But, the wave speed F which is not same everywhere and it is always non-negative. At a certain point, the motion of wavefront is designated by the Eikonal equation, which is given below.

$$1 = F(x)\left|\nabla T(x)\right| \tag{1}$$

where x indicates the position of origin, $F(x)$ represents the wave propagation speed and $T(x)$ denotes the time required by the wave interface to reach x. Further, the magnitude of the gradient of the arrival function $T(x)$ is assumed to be inversely proportional to the velocity of the wave front.

In order to understand the present research paper, it is significant to highlight the property of wave's propagation. It is important to note that the function that represent the time required by the wave interface to reach x. i.e. T(x), only represent a global minima from one single point. Further, as the wave front only expands (F>0), the locations away from the source should have greater arrival time T. Moreover, the problem of local minima will arise only if a particular point has a lesser arrival time (T) than the neighbor point which is closer to the source, which is not possible, as the wave must have already reached this neighbor before. In [31], Sethaian established a discrete solution for the Eikonal equation in a 2D area discretized using a grid map. According to [35] the discretization of gradient ∇T can be achieved with the help of the equations given below.

$$\left\{\begin{matrix} max\left(S_{ij}^{-x}T,0\right)^2 + min\left(S_{ij}^{+x}T,0\right)^2 + \\ max\left(S_{ij}^{-y}T,0\right)^2 + min\left(S_{ij}^{+y}T,0\right)^2 + \end{matrix}\right\} = \frac{1}{f_{ij}^2} \tag{2}$$

where

$$S_{ij}^{-x} = \frac{T_{i,j} - T_{i-1,j}}{h_x},\ S_{ij}^{+x} = \frac{T_{i+1,j} - T_{i,j}}{h_x},\ S_{ij}^{-y} = \frac{T_{i,j} - T_{i,j-1}}{h_y},\ S_{ij}^{+y} = \frac{T_{i,j+1} - T_{i,j}}{h_y} \tag{3}$$

In the above expression, i and j indicate the rows and columns of the grid map, respectively, h_x and h_y are the grid spacing in x and y directions, respectively. Now substitute Eq. (3) in Eq. (2) and simplify to produce Eq. 4 shown below.

$$T_h = min\ (T_{i-1,\,j},\ T_{i+1,j}) \quad and \quad T_v = min\ (T_{i,\,j-1},\ T_{i,j+1}) \tag{4}$$

The revised form of the Eikonal equation, in 2D space after solving the above quadratic equation is given in equation (5).

$$max\left(\frac{T_{ij}-T_h}{h_x},0\right)^2+max\left(\frac{T_{ij}-T_v}{h_y},0\right)^2=\frac{1}{f_{ij}^2} \tag{5}$$

It is to mention that the speed of the wave front is assumed to be positive *(F>0), T* must be greater than T_h and T_v, whenever the wave front has not already passed over the coordinates *i, j*. Subsequently, Eq. (5) can be rewritten as follows.

$$\left(\frac{T_{ij}-T_h}{h_x}\right)^2+\left(\frac{T_{ij}-T_v}{h_y}\right)^2=\frac{1}{f_{ij}^2} \tag{6}$$

In the above equation, whenever $T_{i,j}>T_h$ and $T_{i,j}>T_v$, always choose a greater value for $T_{i,j}$ when solving the Equation. (6). Further, if $T_{i,j}<T_h$ or $T_{i,j}<T_v$, the corresponding member in Equation (5) will become zero. Moreover, while solving equation. (7), if $T_{i,j}<T_h$, then the Eq. (6) is written as follows.

$$\left(\frac{T_{ij}-T_h}{h_x}\right)=\frac{1}{f_{i,j}} \tag{7}$$

Further, if $T_{i,j} < T_v$, the equation (6) will be written as follows.

$$\left(\frac{T_{ij}-T_v}{h_y}\right)=\frac{1}{f_{i,j}} \tag{8}$$

To demonstrate the execution of solution of Eikonal equation, let us consider the following two Figs. 1(a) and 1(b) in which the wave is originated at one and two source points, respectively. In both the figures, the frozen zones are indicated by red colour and their *T* values are not changed. The points for narrow band and unknown zone are marked by yellow and white colour, respectively. It is also important to note that the wave propagates concentrically in Fig. 1(a) and it propagates in Fig. 1(b). This process continuous and the cells expand as the physical wave grows. The cells that have less *T* value will expand first. If two cells have different arrival time, then the cell first addressed by the wave front will expand first.

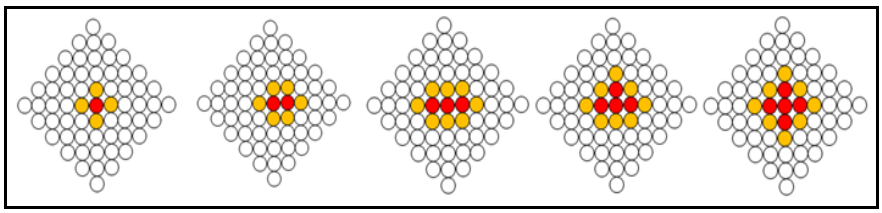


Figure 1 (a)
Wave expansion with one source point

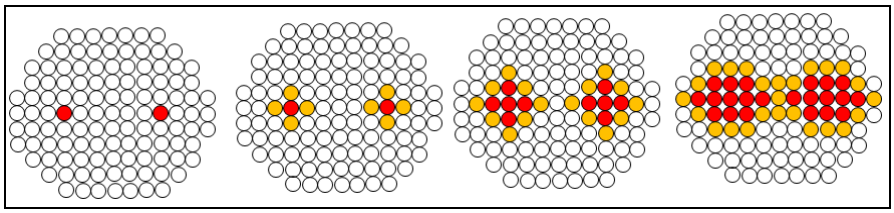


Figure 1 (b)
Wave expansion with two source points

2.1 Regression based Search Method

Although, the FMM algorithm has developed the collision free path in a cluttered environment, it will consume more time and energy of the robot to execute the path. Therefore, to optimize the obtained path in the present research work the authors have implemented a regression search method. In order to obtain the shortest path, the regression search algorithm tries to establish straight lines between the start point and goal point via interconnect points, which are connected with the latter inter points. If the straight line does not cross any obstacle, then the inter start point connects with the next later point with a new straight line until this line crosses any obstacle or the distance between the line and obstacle is less than the d_0.

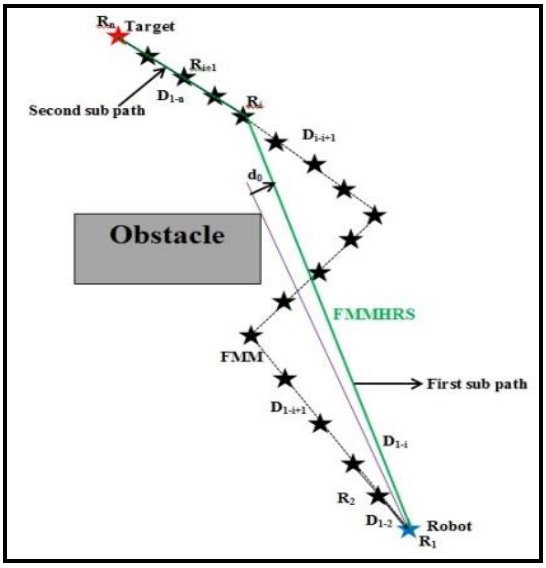


Figure 2
Schematic diagram showing the operation of regression search method

The entire proposed algorithm is given below.

Algorithm 1 Fast marching method hybridized with regression search method

**** Fast marching method****

Input : A grid map of size m × n
Input : Set a node on the grid, where the wave will be originated
Output: Set the value of T for all nodes
Initialization:
T (start point) ⟵ 0
Far ⟵ all grid points
Known ⟵ Identify all grid points with known cost
for each adjacent point k find known point
  Trail ⟵ a
  T(a) = cost update
end
while
  sort check
  the point n ⟵ point with low cost in checking
  remove n from check
  Known ⟵ n
  for each neighbor point k of n
    T(a) = cost update (a)
      If a ∈ Far then
        Remove a from Far
        Trial ⟵ a
      end
  end
end

**** Regression search based method****

R_1 *(start point) ⟵ connects with next point*
From $R_j \in \{R_2, R_3, \ldots\ldots\ldots R_n\}$
If $D_{1,j}$ *⟵ does not cross any obstacle connect next point j=j+1*
  Check the distance $D_{1,j}$ *from obstacle >* d_0 *then j=j+1*
else
  previous point is the next start point
  From $R_k \in \{R_{j+1}, R_{j+2}, \ldots\ldots\ldots R_n\}$
  Check the distance $D_{j,k}$ *from obstacle >* d_0 *then k=k+1*
End
Obtain the optimal path

The schematic diagram showing the principle of operation of regression search method is shown in Fig. 2. Let us consider that the set of sequential points generated by FMM are assumed as R_1, R_2, ... R_i, R_{i+1}, ... R_n. If the regression search method is applied on the said points, the algorithm tries to connect the initial point (that is, inter start point) R_1 with the next sequential point R_2 with the

help of a straight line forming $D_{1\text{-}2}$. Then the algorithm tries to determine that whether $D_{1\text{-}2}$ is crossing any of the obstacles existing in the terrain or not. Once it determines this, the algorithm finds the shortest distance between the line $D_{1\text{-}2}$ and the obstacles. If $D_{1\text{-}2}$ is not crossing any obstacle or the shortest distance is greater than d_0, then the algorithm reconnect with R_1 with R_3 as $D_{1\text{-}3}$, and this procedure is repeated until $D_{1\text{-}i+1}$ crossing any obstacle. Then the local optimal path up to this point is denoted by $D_{1\text{-}i}$. If there are no further obstacles in the terrain, then the similar procedure as mentioned above is repeated between the inter start point R_i and the target R_n to obtain the path $D_{i\text{-}n}$. Now the robot has to move along the optimal path ($D_{1\text{-}i}$ *to* $D_{1\text{-}n}$), which consumes less energy when compared with the path obtained by FMM algorithm.

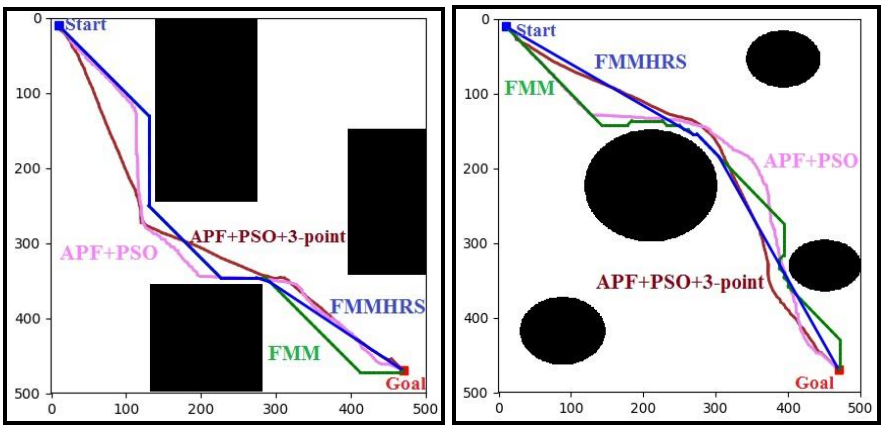
3 Results and Discussions Related to Simulation Studies

In this section, the simulation results related to the FMM and FMMHRS in two dimensional work spaces under static and dynamic environments are presented. Once the path planning algorithms have been developed, the effectiveness in generating the collision free paths on various scenarios is studied in computer simulations. These computer simulations are conducted in Python programming environment with the help of a PC that consists of Intel i5 processor running on 2.2 GHz. The 2D simulation space considered in the programming environment is fixed at 500 × 500 pixel.

3.1 Simulation Results in the Static Environment

In the present section FMM and FMMHRS algorithms are compared with an artificial potential field method (APF) combined with particle swarm optimization (PSO) and three point smoothing method available in the literature [30] are shown in Fig. 3. From Fig. 3, it can be observed that both the proposed approaches and the approach available in [30] are found to generate the collision-free paths in both the scenarios shown above. Further, Table 1 gives the path lengths and time taken to generate the path during the said simulation study. It is seen that the length of the path generated by FMM and FMMHRS algorithms are seen to be less when compared with the APF with PSO and three point smoothing method.

Moreover, it has also been observed that the hybrid algorithm (that is, FMMHRS) proposed in the present research has performed better than the other algorithms considered in this study in terms of both the path length and time taken to generate the path. Further, the authors have tested the proposed algorithms on new scenarios/maps in both static and dynamic environments.



(a) (b)

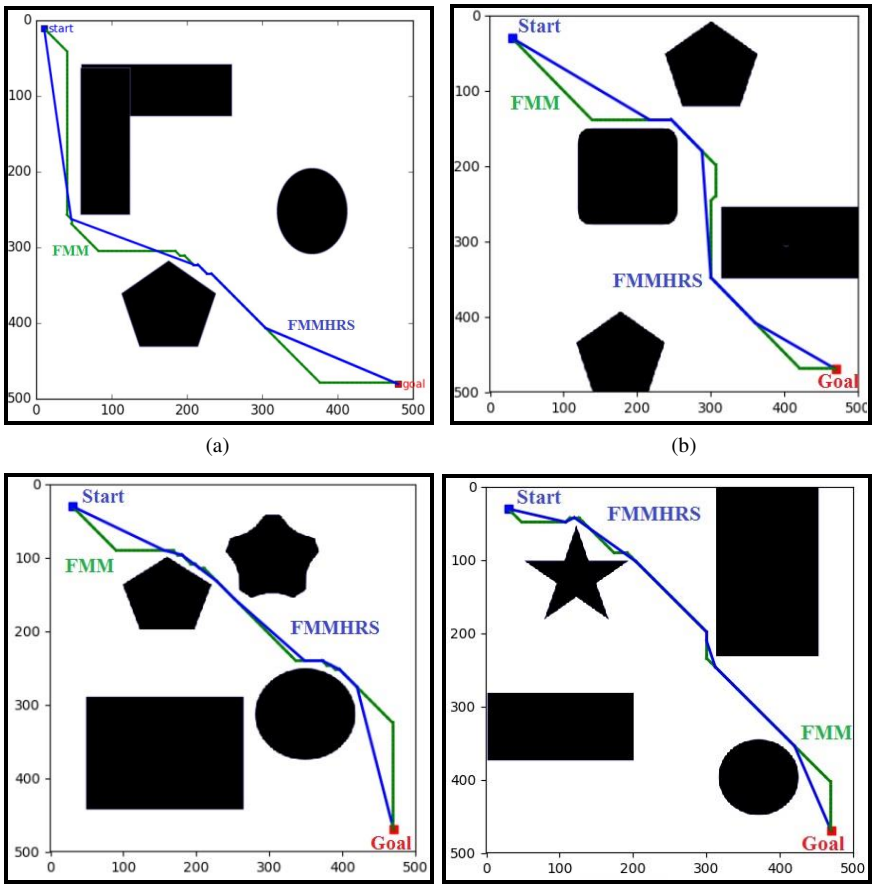
Figure 3
Simulation results of various approaches on different scenarios (a) map 1 and (b) map 2

Table 1
Comparison of path length and time needed to generate paths during simulation

Maps	Path length in pixels				Time taken to generate the path in 'sec'			
	FMM	FMM HRS	APF+ PSO	APF+ PSO+3-point	FMM	FMM HRS	APF+ PSO	APF+ PSO+3-point
map 1	676.08	664.26	819. 87	753.26	30	28	46	40
map 2	708.74	663.34	822. 49	715.35	33	27	47	35

The results related to the generation of path on new terrains after employing FMM and FMMHRS are shown in Fig. 4. It can be observed that in every case, the path developed by the FMMHRS is shortest and optimal in nature when compared with the path obtained by standard FMM algorithm. It is to be noted that in all the maps the obstacles are marked with black color and certain amount of clearance is provided around the obstacles. The path developed by the FMM and FMMHRS are indicated with green and blue color lines, respectively.

Further, Table 2 gives the path length and time required to generate the path for various terrains. From the results of Fig. 4 and Table 2, it has been observed that the FMMHRS approach is seen to provide a shorter path when compared with the standard FMM approach. This may be due to the fact that in FMM approach initially the collision-free path is obtained basically by not considering the shortest route. Later, regression search is employed in which it is always trying to draw a straight line between interstate point and goal point. Then the algorithm will try to determine the location and providing certain clearance around the obstacle to safely navigate the robot without any collision. This fact led to the generation of shortest path.



(a) (b)

(c) (d)

Figure 4

Simulation results related to the static environment at different scenarios (a) map-1, (b) map-2, (c) map-3 and (d) map-4

Table 2

Simulation results related to the static obstacles

Maps	Path length in pixels		Time taken to generate the path in	
	FMM	FMMHRS	FMM	FMMHRS
map 1	791.8965	749.0306	40.23	35.56
map 2	789.7787	684.0287	39.52	32.21
map 3	697.7787	627.8427	34.42	28.25
map 4	678.1463	622.2539	30.25	27.56

3.2 Simulation Results in the Dynamic Environment

The developed FMM and FMMHRS algorithms are also used to generate collision-free optimal path in some dynamic environments. The results of simulation for the scenarios involving one, two and three dynamic obstacles are shown in Figs. (5) and (6), respectively. In all the cases, the scenarios are created in such a way, that the straight line path of the robot will be disturbed, so that once again the robot will plan its future course of its action without deviating from the goal.

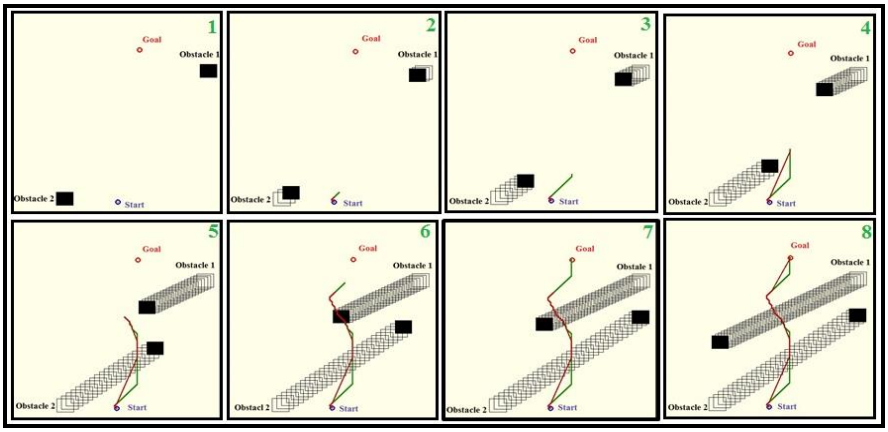


Figure 5

Simulation results related to the dynamic environment with two obstacles

The obstacles are painted with black color and the start and goal points are indicated by blue and green color circles, respectively. The paths generated by the FMM and FMMHRS algorithms are indicated with green and brown color lines, respectively. It can be observed that, in all the scenarios the robot is trying to avoid the collision with the obstacles. The path length and time of travel for the robot to reach the goal are given in Table 3. In this case also FMMHRS algorithm is seen to provide optimal path when compared with FMM algorithm. The reason for this is also same as the one explained above for the static obstacle case.

Table 3

Simulation results related to the dynamic obstacles

Obstacles	Path length in pixels		Time taken to generate the path in	
	FMM	FMMHRS	FMM	FMMHRS
two	441.5878	418.8691	23.21	21.14
three	590.8082	569.0802	38.23	36.12

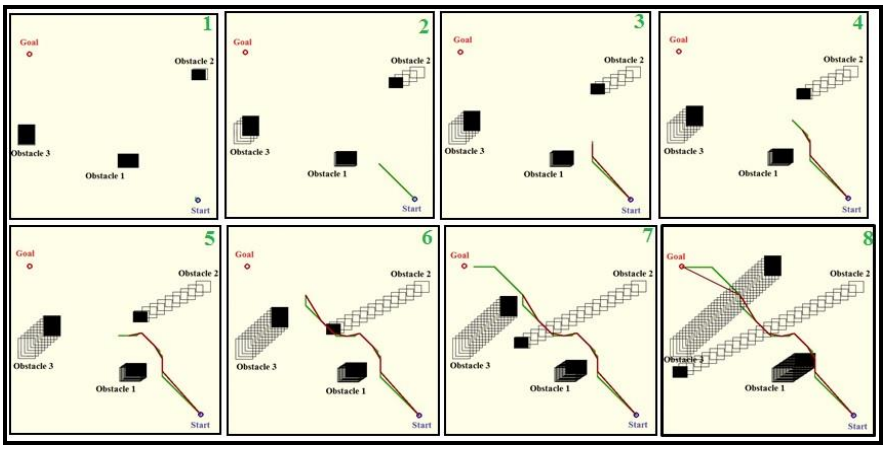


Figure 6
Simulation results related to the dynamic environment with three obstacles

4 Experiments in the Real Environment

In the present work, the effectiveness of the developed motion planning algorithms is verified by conducting real time experiments. To execute the paths developed by FMM and FMMHRS algorithms, the authors have chosen a biped robot [36].

4.1 Experimental Results in the Static Environment

To find the effectiveness of the developed motion planning algorithms, in the present study two different scenarios shown in Figs. (7) and (8) are considered. For ease in identification during image processing, the terrain and the obstacles are painted in white and red color, respectively. The obstacles (that is, static) are located on the terrain in a particular fashion to reflect different scenarios. Further, the start and goal points are marked using marker pen on the terrain. An overhead camera is mounted at the top of the terrain to capture the video of the scene. The two shoulders of the biped robot are marked with green color to indicate the location in the terrain through image processing. The algorithms are implemented using Python software and the image processing technique is used to detect the locations of the obstacles and robot in the scene. While conducting the real time experiments, a wired communication has been employed between the robot and computer terminal to transmit the data related to the on-line path developed by the algorithms, and the required gait angles that are generated to track the path

decided by using the said algorithms. The paths developed by the FMM and FMMHRS algorithms are marked by thick yellow and green lines which are shown in Figs. (7) and (8), respectively. The path length, distance travelled by the robot and time taken by the robot to reach the goal position are given in Table 5.

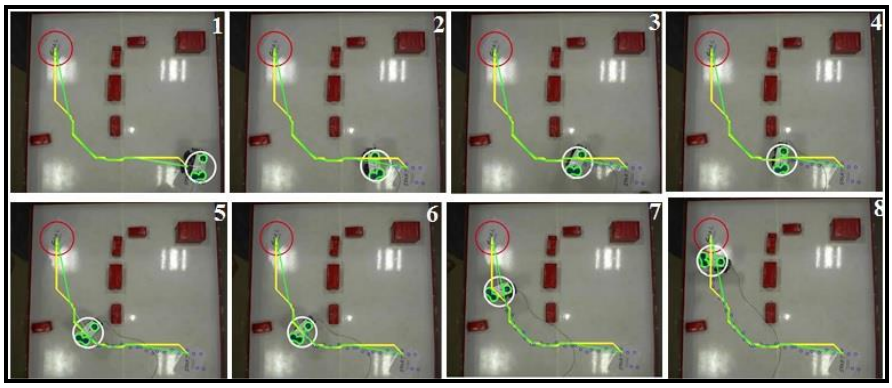


Figure 7
Experimental results for navigation of the biped robot in real time static environment (Scenario 1)

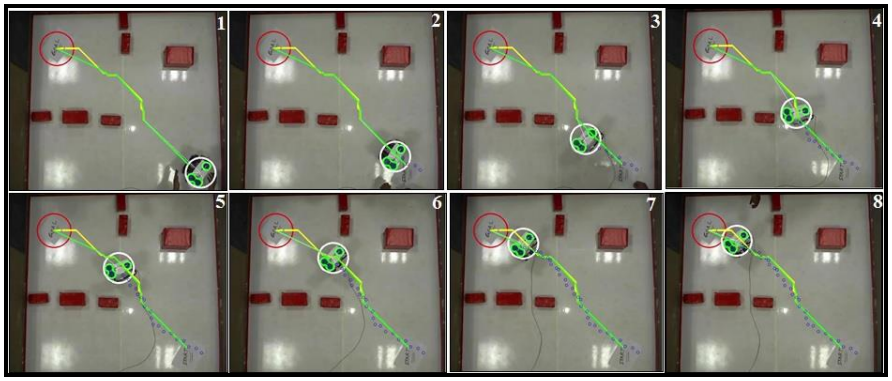


Figure 8
Experimental results for navigation of the biped robot in real time static environment (Scenario 2)

Table 5
Results related to the path length and time travel for various scenarios

Scenarios	Path length in pixels			Robot travel time in sec
	FMM	FMMHRS	Robot	
Scenario 1	586.6761	564.5822	630.0924	215.45
Scenario 2	504.3817	494.9290	657.3502	221.16

It can be observed that the path decided by FMMHRS is seen to be the most optimal when compared with the path developed by the FMM algorithm. The reason for this is same as the one explained earlier. Further, the distance travelled by the robot is seen to be more when compared with FMMHRS. This may be due to the fact that the biped robot is a mechanism that is supported on discrete foot holds and the balancing is a serious problem. Therefore, when the path generated by the algorithm is curved in nature, it will be difficult for the robot to track the exact path due to the following reasons.

1. The robot cannot make sharp turns due to balancing problems raised by the changes in inertia of the robot.
2. The play exists in the transmission mechanism between the motor and the joint also allows for certain misalignment of the path while tracking.

Further, it has been observed that the time taken by the robot to travel from start to finish is seen to be very high when compared with the time required to generate the path by the algorithm. It might have happened because the mobility of the biped robot is very slow due to the balancing problems. The other reason could be the wired transmission of data between the computer terminal and the robot. However, the biped robot has successfully navigated the path among the static obstacles.

4.2 Experimental Results in the Dynamic Environment

The real experiments related to the execution of motion planning in dynamic environments consists of two and three obstacles as shown in Figs. 9 and 10, respectively. As it is a dynamic environment, the obstacles used in this study are allowed to move slowly to meet the requirements of slow walking of the biped robot. During dynamic walking, the path will be updated at regular intervals of time after considering the new location of the robot and moving obstacles. The path update rate can be varied based on the velocity of the biped robot and obstacles. Figures 9 and 10 show the path generated by the FMM and FMMHRS algorithms and tracked by the robot on the terrain while moving among two and three obstacles, respectively. Table 6 shows the result related to the distance covered by the robot from start to the goal and the time taken by the robot to reach the goal position. The results shows that both the algorithms that is, FMM and FMMHRS are capable of generating the path in real time for the environments that contain dynamic obstacles. Further, the biped robot is seen to follow the optimal path generated by FMMHRS with little deviation. The reasons for this are explained in the experiments related to the cases involving static obstacles.

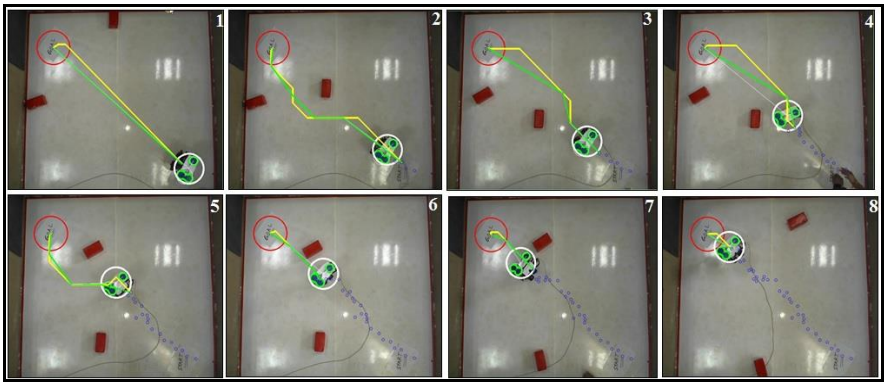


Figure 9
Experimental results for navigation of the biped robot in real time dynamic environment (Scenario 1)

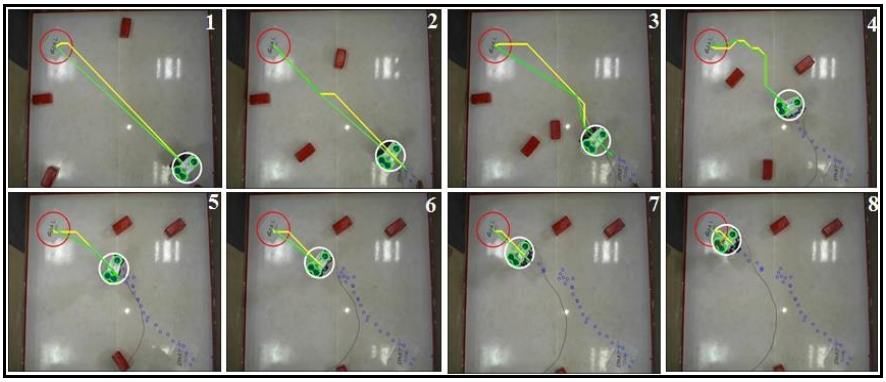


Figure 10
Experimental results for navigation of the biped robot in real time dynamic environment (Scenario 2)

Table 6
Results related to the distance covered and time of travel for various scenarios

Scenarios	Path covered by the robot in pixels	Time travel in 'sec'
Scenario	607.554	210.52
Scenario	647.112	218.23

4.3 Comparison with Other Work

Based on the literature, the authors performed some qualitative comparisons with the approaches reported in [24, 31, and 37-40]. Till date, some of the researchers had used a FMM algorithm to generate the path in computer simulations. In this work, the authors not only used this algorithm to generate the path in a cluttered

environment, but also developed an optimal path after combining FMM with regression search (FMMHRS). Moreover, some of the researchers [29, 30 and 41] had worked on the generation of collision-free path in both the static and dynamic environments in computer simulations only. In the present research, the authors have implemented the said algorithms in both computer simulations and in real time environments. Further, a majority of research in motion planning involves the usage of wheeled robots for validation, which has having better mobility and stability. The only drawback is that, it only can navigate on a continuous terrain, whereas the biped robots are planned to use in the environments that are non-continuous. It is important to note that the mobility and stability of the biped robot is poor while in motion and it can navigate on a discontinuous terrain. In the present study, the proposed FMM and FMMHRS algorithms are successfully implemented on the biped robot in both the static and dynamic environments.

5 Conclusions

This paper explains the features of the FMM and FMMHRS algorithms used to generate a path in both the static and dynamic obstacles environments. Initially, both the algorithms are used to solve the motion planning problem in simulations and in various scenarios. Based on the results of simulation, it can be observed that the developed algorithms are capable of generating collision free paths from start to the goal point. It has been observed that the FMMHRS algorithm is seen to perform better than the FMM approach, for both the static, as well as dynamic scenarios. It may be due to the fact that FMMHRS always tries to provide a straight-line path between the start point and the goal point, when there is no obstacle in the line of path. Further, the real-biped robot is seen to track the path with little deviation and reach the goal point safely.

References

[1] J.-C. Latombe, Robot motion planning, Dordrecht, Netherlands: Kluwer Academic Publishers (1991)

[2] Istvan N, Behaviour study of a multi-agent mobile robot system during potential field building, Acta Polytechnica Hungarica, 6 (4) (2009) 111-136

[3] Masehian, E. and Sedighizadeh. D, Classic and heuristic approaches in robot motion planning – a chronological review, World Academy of Science Engineering and Technology, 29 (5) (2007) 101-106

[4] Oh, J. S., Choi, Y. H. and Park, J. B, Complete coverage navigation of cleaning robots using triangular-cell-based map, IEEE Trans on Industrial Electronics, 51 (3) (2004) 718-726

[5] Voronoi, G.F, Nouvelles applications des paramètres continus à la théorie de formes quadratiques, Journal für die reine und angewandte Mathematik, (1908) 134-198

[6] Tarjan, R. E, A unified approach to path problem', Journal of the Association for Computing Machinery, 28 (3) (1981) 577-593

[7] Takahashi, O. and Schilling, R. J., Motion planning in a plane using generalized Voronoi diagrams, IEEE Trans on Robotics and Automation, 5 (2) (1989) 143-150

[8] Avneesh, S., Erik, A. and Sean, C., Real-time path planning in dynamic virtual environment using multi-agent navigation graphs, IEEE Trans on Visualization and Computer Graphics, 14 (3) (2008) 526-538

[9] Cai, C. H. and Ferrai, S., Information-driven sensor path planning by approximate cell decomposition, IEEE Trans on Systems, Man, and Cybernetics, Part B: Cybernetics, 39 (3) (2009) 672-689

[10] Nilsson, N. J., Problem-solving methods in artificial intelligence, Artificial Intelligence: A New Synthesis, Morgan Kaufmann Publishers (2000)

[11] Stentz, A., The focused D* algorithm for real-time re-planning, Proceeding of the International Joint Conference on Artificial Intelligence (1995) 1995-2000

[12] Sedighi, K. H., Ashenayi, K. and Manikas, T. W., Autonomous local path planning for a mobile robot using a genetic algorithm, Proceedings of Congress on Evolutionary Computation, 2 (2004) 1338-1345

[13] Blackowiak, A. D. and Rajan, S. D., Multipath arrival estimates using simulated annealing: application to crosshole tomography experiment, IEEE Journal of Oceanic Engineering, 20(3) (1995) 157-165

[14] Garcia, M. A. P., Montiel, O. and Castillo, O., Path planning for autonomous mobile robot navigation with ant colony optimization and fuzzy cost function evaluation, Applied Soft Computing, 9 (3) (2009) 1102-1110

[15] A. Ayari and S.Bouamama, A new multiple robot path planning algorithm: dynamic distributed particle swarm optimization, Robotics and Biomimetic, 4 (8) (2017)

[16] Prases K. Mohanty & Dayal R. Parhi, Optimal path planning for a mobile robot using cuckoo search algorithm, Journal of Experimental & Theoretical Artificial Intelligence, 28 (1-2) (2016) 35-52

[17] Panda, M. R., et al., Hybridization of IWO and IPSO for mobile robots navigation in a dynamic environment. Journal of King Saud University – Computer and Information Sciences (2017) https://doi.org/10.1016/j.jksuci.2017.12.009

[18] Md. Arafat Hossain and Israt Ferdous, Autonomous Robot Path Planning in Dynamic Environment Using a New Optimization Technique Inspired by Bacterial Foraging Technique, 2013 International Conference on Electrical Information and Communication Technology (2013) 1-6

[19] Liu, C., Gao, Z. and Zhao, W., A new path planning method based on firefly algorithm. In Fifth international joint conference on computational sciences and optimization (CSO) Harbin, China (2012) 775-778 doi:10.1109/CSO.2012.174

[20] D. K. Pratihar, K. Deb & A. A Ghosh, Fuzzy-Genetic Algorithms and Time-Optimal Obstacle-Free Path Generation for Mobile Robots, Engineering Optimization, 32 (1) (1999) 117-142

[21] I. Engedy and G. Horváth, Artificial Neural Network based Mobile Robot Navigation, 6th IEEE International Symposium on Intelligent Signal Processing (2009) 241-246

[22] S. Garrido, L. Moreno, J. V. Gómez and P. U. Lima, General Path Planning Methodology for Leader-Follower Robot Formations, International Journal of Advanced Robotic Systems, 10 (2013) 1-10

[23] L. Janson and M. Pavone, Fast marching trees: a fast marching sampling-based method for optimal motion planning in many dimensions, 16th International Symposium on Robotics Research (2013)

[24] Q. H. Do, S. Mita and K. Yoneda, Narrow Passage Path Planning Using Fast Marching Method and Support Vector Machine, 2014 IEEE Intelligent Vehicles Symposium (2014) 630-635

[25] Alberto Valero-Gomez, Javier V. Gomez, Santiago Garrido and Luis Moreno, Fast Marching Methods in Path Planning, IEEE Robotics & Automation Magazine, 20 (2013) 111-120

[26] Garrido, S., Moreno, L., Blanco, D., Exploration of a cluttered environment using voronoi transform and fast marching. Robotics Autonomous Systems 56 (12) (2008) 1069-1081

[27] Petres, C., Pailhas, Y., Petillot, Y., Lane, D., Underwater path planning using fast marching algorithms. In: Proceedings of the Oceans – Europe (2005) 814-819

[28] R. Song, Y. Liu , R. Bucknall, A multi-layered fast marching method for unmanned surface vehicle path planning in a time-variant maritime environment, Ocean Engineering 129 (2017) 301-317

[29] G. Li, Y. Tamura, A. Yamashita and H. Asama, Effective improved artificial potential field-based regression search method for autonomous mobile robot path planning, Int. J. Mechatronics and Automation, 3 (3) (2013) 141-170

[30] R. K. Mandava, S. Bondada, P. R. Vundavilli, An Optimized Path Planning for the Mobile Robot using Potential Field Method and PSO algorithm, 7th International Conference on soft computing and problem solving (socpros-2017) Bhubaneswar, India, 2017

[31] J. A. Sethian, A fast marching level set method for monotonically advancing fronts, Proc. Nat. Acad Sci. U.S.A. 93 (4) (1996) 1591-1595

[32] S. Jbabdi, P. Bellec, R. Toro, J. Daunizeau, M. Pélégrini-Issac, and H. Benali, Accurate anisotropic fast marching for diffusion-based geodesic tractography, Int. J. Biomedical Imaging, 2008 (2) (2008)

[33] H. Li, Z. Xue, K. Cui, and S. T. C. Wong, Diffusion tensor-based fast marching for modeling human brain connectivity network, Comp. Med. Imag. and Graph. 35(3) (2011) 167-178

[34] K. Yang, M. Li, Y. Liu, and C. Jiang, Multi-points fast marching: A novel method for road extraction, The 18th International Conference on Geo-informatics: GI Science in Change, Geo-informatics, (2010) 1-5

[35] S. Osher and J. A. Sethian, Fronts Propagating with Curvature Dependent Speed: Algorithms based on Hamilton-Jacobi Formulations, Journal of Computational Physics, 79 (1) (1988) 12-49

[36] R. K. Mandava and P. R. Vundavilli, Whole body motion generation of 18-DOF biped robot on flat surface during SSP and DSP, International Journal of Modeling Identification and Control, In Press (2018)

[37] Santiago Garrido, Luis Moreno, Dolores Blanco and Piotr Jurewicz, Path Planning for Mobile Robot Navigation using Voronoi Diagram and Fast Marching, International Journal of Robotics and Automation (IJRA) 2 (1) (2011) 42-64

[38] 1P. Melchior, B. Orsoni, O. Lavialle, A. Poty and A. Oustaloup, Consideration of obstacle danger level in path planning using A* and fast-marching optimization: comparative study, Signal Processing, 11 (2003) 2387-2396

[39] C. H. Chiang, P. J. Chiang, A comparative study of implementing Fast Marching method and A* search for mobile robot path planning in grid environment: effect of map resolution, in Proc. IEEE Advanced Robotics and Its Social Impacts, (2007) 1-6

[40] Q. H. Do, S. Mita and K. Yoneda, A practicle and optimal path planning for autonomous parking using fast marching algorithm and support vector machine, IEICE Trans. Inf. & Syst. 96 (12) (2013) 2795-2804

[41] J. Vascak and M. Rutrich, Path planning in dynamic environment using Fuzzy Cognitive Maps, in 2008 6th International Symposium on Applied Machine Intelligence and Informatics (2008) 5-9

On Distributivity and Conditional Distributivity of S-uninorms over Uninorms

Dragan Jočić

Novi Sad School of Business, Vladimira Perića-Valtera 4, 21000 Novi Sad, Serbia, dragan.jocic@vps.ns.ac.rs

Ivana Štajner-Papuga

Department of Mathematics and Informatics, University of Novi Sad, Trg D. Obradovića 4, 21000 Novi Sad, Serbia, ivana.stajner-papuga@dmi.uns.ac.rs

Abstract: Aggregation operators with an annihilator are in the focus of a significant number of research papers due to their applicability in both theoretical and practical areas of mathematics. Therefore, the main topic of this paper is distributivity and conditional distributivity for some classes of aggregation operators with this property. The characterization of all pairs (F, G) of aggregation operators that are satisfying distributivity law, on both whole and restricted domain, where F is a S-uninorm from U_{min}, and G is a t-norm or a uninorm from U_{min} or U_{max} is given.

Keywords: aggregation operator; annihilator; t-norm; uninorm; S-uninorm; distributivity; conditional distributivity

1 Introduction

Lately, aggregation operators have been intensively investigated due to their valuable role in many applications, from mathematics and natural sciences to economics and social sciences (see [9, 11, 15]). Of the special interest is their role in the integration theory [22] and in the utility theory [6, 11, 13]. Regarding this, the main problem that is being studied is the characterizations of the pairs of aggregation operators that are distributive. This issue appeared first in [1]. The more recent results concern t-norms and t-conorms [9], quasi-arithmetic means [2], uninorms and nullnorms [5, 8, 18, 19, 23], semi-t-operators and uninorms [24, 25], etc. The issue of the simultaneous distributivity of t-norms and t-conorms over uninorms was investigated in [4]. Also, the problem of distributivity that is directed towards the restricted domain, i.e., the conditional (restricted)

distributivity, is highly important since this approach can provide more solutions ([12, 15, 16, 17, 21, 22]).

The next step is to direct this type of research towards the general commutative aggregation operators with an annihilator, namely towards T-uninorms and S-uninorms. The characterization of this type of operators was done in [20]. Therefore, the aim of this paper is to continue the research from [14] where the problem of T-uninorms and uninorms was considered. Now the S-uninorms, which are a generalization of conjunctive uninorms and nullnorms (t-operators), are observed. The first part of paper considers distributivity of S-uninorms over t-norms and t-conorms and uninorms from the class $U_{\min} \cup U_{\max}$. The second part deals with distributivity equations on the restricted domain. Since the conditional distributivity of nullnorms over uninorms was considered in [12], the results given here upgrades the previous results.

2 Basic Notions

The core of this research are aggregation operators with an annihilator. As stated in [11], an aggregation operator in $[0,1]^{\mathrm{n}}$ is a function $A^{(\mathrm{n})}\colon [0,1]^{\mathrm{n}} \to [0,1]$ that is non-decreasing in each variable and that fulfills the boundary conditions

$$A^{(\mathrm{n})}(0,\dots,0) = 0 \quad \text{and} \quad A^{(\mathrm{n})}(1,\dots,1) = 1.$$

The integer n is the number of input values of the observed aggregation. Further on the binary aggregation operators are being investigated, therefore, the notation A will be used for $A^{(2)}$. Of course, depending on the intended application, some other properties can be required, e.g. associativity, commutativity, idempotency, decomposability, neutral and annihilator elements, etc., (see [11]). Also, if required, the previous can be extended to an arbitrary real interval $[\mathrm{a}, \mathrm{b}]$.

Therefore, the firs part of this section consists of an overview of aggregation operators that are essential for the presented research. Necessary notions concerning distributivity are given in the second part of this section.

2.1 Uninorms

The first type of aggregation operators that is needed for the presented research is an aggregation operator with a neutral element, namely the uninorm.

Definition 1 *([27]) A uninorm* $U\colon [0,1]^2 \to [0,1]$ *is binary aggregation operator that is commutative, associative, and for which there exists a neutral element* $e \in [0,1]$, *i.e.,* $U(x,e) = x$ *for all* $x \in [0,1]$.

If $e = 1$, the uninorm U becomes a t-norm (triangular norm) and it is denoted by T. If $e = 0$, the uninorm U is a t-conorm (triangular conorm) denoted by S.

A uninorm is called conjunctive if $U(0,1) = 0$, and disjunctive if $U(0,1) = 1$. Uninorms for which both functions $U(x,0)$ and $U(x,1)$ are continuous, except perhaps at the point e, are characterized based on the value $U(0,1)$ by the following theorem from [10].

Theorem 2 *([10]) Let U be a uninorm with a neutral element $e \in (0,1)$ such that both functions $U(x,1)$ and $U(x,0)$ are continuous except at the point $x = e$.*

If $U(0,1) = 0$, then

$$U(x,y) = \begin{cases} eT\left(\frac{x}{e}, \frac{y}{e}\right) & on\ [0,e]^2, \\ e + (1-e)S\left(\frac{x-e}{1-e}, \frac{y-e}{1-e}\right) & on\ [e,1]^2, \\ \min(x,y) & otherwise, \end{cases} \quad (1)$$

where T is a t-norm, and S is a t-conorm.

If $U(0,1) = 1$, then

$$U(x,y) = \begin{cases} eT\left(\frac{x}{e}, \frac{y}{e}\right) & on\ [0,e]^2, \\ e + (1-e)S\left(\frac{x-e}{1-e}, \frac{y-e}{1-e}\right) & on\ [e,1]^2, \\ \max(x,y) & otherwise, \end{cases} \quad (2)$$

where T is a t-norm, and S is a t-conorm.

T from (1) (and (2)) is the underlying t-norm of U and S is the underlying t-conorm of U. The family of all uninorms of the form (1) is denoted by $U_{\min}$, while the family of all uninorms of the form (2) is denoted by $U_{\max}$. More on t-norms, t-conorms and uninorms can be found in [10,11,26,27] and the relation between mentioned classes is given by the Figure 1.

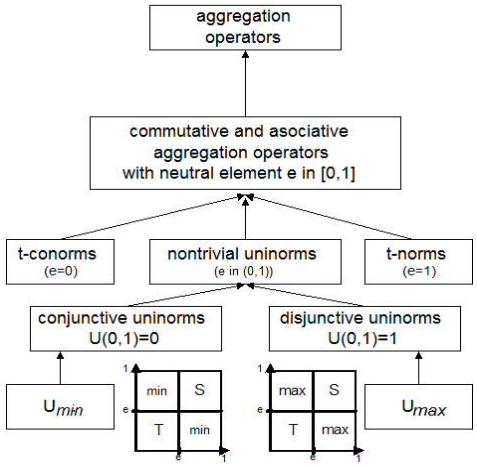


Figure 1
Uninorms

Example 3 The first uninorms in the terms of Definition 1 were considered by Yager and Rybalov (see [27]) and they are idempotent uninorms U_e^{min} and U_e^{max} from classes $U_{\min}$ and $U_{\max}$, respectively, of the following form

$$U_e^{min} = \begin{cases} \max & on \ \ [e,1]^2, \\ \min & otherwise, \end{cases} \tag{3}$$

and

$$U_e^{max} = \begin{cases} \min & on \ \ [0,e]^2, \\ \max & otherwise. \end{cases} \tag{4}$$

Uninorms (3) and (4) are the only idempotent uninorms from classes $U_{\min}$ and $U_{\max}$. On the other hand, the only idempotent t-norm (t-conorm) is minimum (maximum). The idea of a uninorm appeared for the first time in [3] in the form of the aggregative operator, which now can be considered as a generated uninorm.

2.2 Commutative Aggregation Operators with an Annihilator

Another type of aggregation operators that is needed for the presented research consists of aggregation operators with an annihilator (absorbing element). An element $a \in [0,1]$ is an annihilator for an aggregation operator A if

$$A(a,x) = A(x,a) = a$$

for all $x \in [0,1]$. Further on, the general commutative aggregation operators with an annihilator $a$are denoted with a-CAOA (see [20]).

For any binary operator $A\colon [0,1]^2 \to [0,1]$ and any element $c \in [0,1]$, the section $A_c\colon [0,1] \to [0,1]$ is given by

$$A_c(x) = A(c,x).$$

Now, the continuity (discontinuity) of sections A_0 and A_1 plays a crucial role in classification and characterization of associative a-CAOA operators as given by Figure 2 (see [20]).

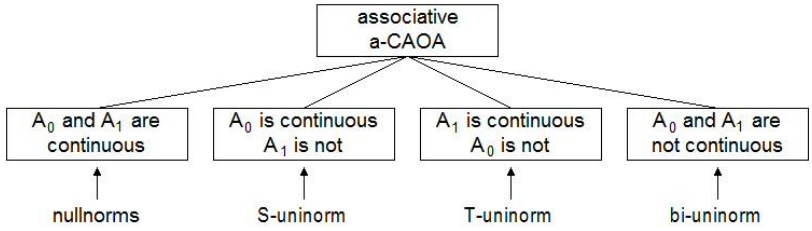


Figure 2
Classification of a-CAOA

2.2.1 S-uninorms

Definition 5 *([20]) A binary operator $A\colon [0,1]^2 \to [0,1]$ is called a S-uninorm if it is an associative a-CAOA satisfying the following properties:*

- *Section A_0 is continuous and section A_1 is not.*
- *There is $e \in (0,1)$ such that e is an idempotent element, the section A_e is continuous and $A_e(1) = 1$.*

Theorem 6 *([20]) Let $A\colon [0,1]^2 \to [0,1]$ be a binary operator. The following statements are equivalent:*

- *A is a S-uninorm.*
- *There exists $a \in [0,1)$, a t-conorm S' and a conjunctive uninorm U' with neutral element $e' \in (0,1)$ such that A is given by*

$$A(x,y) = \begin{cases} aS'\left(\frac{x}{a},\frac{y}{a}\right), & on\ [0,a]^2, \\ a + (1-a)U'\left(\frac{x-a}{1-a},\frac{y-a}{1-a}\right), & on\ [a,1]^2, \\ a, & on\ [0,a]\times[a,1] \cup [a,1]\times[0,a]. \end{cases} \tag{5}$$

- *There exists $a \in [0,1)$, a t-conorm S and a conjunctive uninorm U with neutral element $e \in (0,1)$ such that $U(x,a) \le a$ for all $x \in [0,1]$, $U \le S$ and $A = med(a,U,S)$.*

Remark 7 Let $A\colon [0,1]^2 \to [0,1]$ be a S-uninorm.

- For a $= 0$,operator A becomes a conjunctive uninorm, i.e., $A = U'$.
- If a $\neq 1$, in order to ensure the discontinuity of A_1 and since $A_e(1) = 1$, a $<$ e.
- If $U' \in U_{min}$, then A is a S-uninorm from U_{min}.

Example 8 Binary operator $A\colon [0,1]^2 \to [0,1]$ of the form

$$A(x,y) = \begin{cases} a, & on\ [0,a]\times[a,1] \cup [a,1]\times[0,a], \\ \max(x,y), & on\ [0,a]^2 \cup [e,1]^2, \\ \min(x,y), & otherwise, \end{cases} \tag{6}$$

is an idempotent S-uninorm from $U_{\min}$ with annihilator a. It is obtained from (5) when for the t-conorm S' the operator max is taken and the uninorm U' is $U_e^{\min}$.

2.3 Distributivity on the Whole Domain

Let $A, B\colon [0,1]^2 \to [0,1]$ be two arbitrary operators.

- A is left distributive over B if

$$A\big(x, B(y,z)\big) = B\big(A(x,y), A(x,z)\big), \quad \text{for all } x,y,z \in [0,1].$$

- A is right distributive over B if

$$A\big(B(y,z), x\big) = B\big(A(y,x), A(z,x)\big), \quad \text{for all } x,y,z \in [0,1].$$

The previous two functional equations are called the left and the right distributivity laws (see [1], p. 318), and are denoted with (LD) and (RD). Of course, for a commutative A, (LD) and (RD) coincide. Now, A is distributive over B if it is both left and right distributive over B.

The following two lemmas answer some starting questions regarding distributivity.

Lemma 9 *([5]) Let $X \neq \emptyset, A: X^2 \to X$ and let $e \in Y$, where $Y \subset X$, be a neutral element for the operator A on Y ($\forall_{x \in Y} A(e,x) = A(x,e) = x$). If the operator A is left or right distributive over some operator $B: X^2 \to X$ that fulfils $B(e,e) = e$, then B is idempotent on Y.*

Lemma 10 *([5]) All increasing functions are distributive over max and min.*

2.4 Distributivity on the Restricted Domain

As seen in [9], the problem of distributivity of a t-norm Tover a t-conorm S on the whole domain has only the trivial solution, i.e., t-conorm in question has to be $S_M = \max$. In order to obtain more solutions, the domain had to be restrict. That is, for the classical functional equation $T(x, S(y,z)) = S(T(x,y), T(x,z))$, the additional condition $S(y,z) < 1$ is necessary (see [15], p. 138). This distributivity under the given restriction is called the conditional (restricted) distributivity and it is denoted with (CD).

The similar restriction holds for conditional distributivity of a t-conorm S over a t-norm T:

$$(CD) \qquad S\big(x, T(y,z)\big) = T\big(S(x,y), S(x,z)\big), \ \ whenever\ T(y,z) > 0,$$

for all $x, y, z \in [0,1]$.

The previous concept can be extended to some more general aggregation operators, as given by the following definition.

Definition 11 *Let F be a S-uninorm with annihilator $a \in (0,1)$ and let G be a t-norm or $G \in U_{min} \cup U_{max}$. F is conditionally distributive (CD) over G if for all $x, y, z \in [0,1]$ the following holds*

$$(CD) \qquad F\big(x, G(y,z)\big) = G\big(F(x,y), F(x,z)\big), \ \ whenever\ G(y,z) > 0.$$

Lemma 12 *([5]) All increasing functions are conditionally distributive over max and min.*

3 Distributivity Laws for S-uninorms over Uninorms

This section considers distributivity of a S-uninorm from $U_{\min}$ with an annihilator a over a t-conorm, a t-norm or a uninorm from $U_{\min} \cup U_{\max}$.

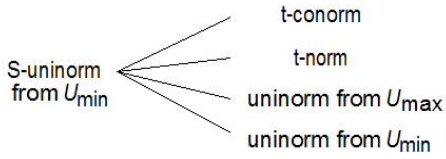


Figure 3
Topics of the Section 3

Since a S-uninorm is a commutative operator, there is no need to discus (LD) and (RD) separately. The distributivity of F over G for $a = 0$, that is for F being a uninorm from $U_{\min}$, was investigated in [18, 19]. Therefore, the further assumption is that $a \in (0,1)$. Also, further on the neutral element of the underlying uninorm of F will be denote by e. The similar issues were simultaneously investigated in [7] and some of the following results are independently confirmed.

3.1 S-uninorm vs. t-norm and t-conorm

The results concerning t-norms and t-conorms are not very surprising since the idempotence still plays an important role. Additionally, some aspects of the proofs are analogous to ones from [14], therefore, they can be omitted. Also, see [7].

Theorem 13 *Let F be a S-uninorm from U_{min} and let T be a t-norm. F is distributive over T if and only if $T = min$.*

Theorem 14 *Let F be a S-uninorm in U_{min} and let S be a t-conorm.*

(i) If F is distributive over S then $S(x,x) = x$ for all $x < e$.
(ii) Let the function $s(x) = S(x,x)$ be left-continuous at the point $x = e$. Then, F is distributive over S if and only if $S = max$.

3.2 S-uninorm vs. uninorm from U_{min}

Now the second operator is a conjunctive uninorm U with continuous functions (except perhaps at the point e) $U(x,0)$ and $U(x,1)$. The first two lemmas are necessary for the proof of the main theorem of this subsection.

The main idea behind the proofs that follow is analogous to one from [14] for F being a T-uninorm from U_{max} and U be a uninorm from the class U_{max}. However, it is very interesting to see how the duality of operators influences the process of proving and, therefore, the proofs in this section are not omitted.

Lemma 15 *Let F be a S-uninorm from U_{min} and let U be a uninorm from the class U_{min} with a neutral element $e_1 \in (0,1)$. If F is distributive over U then $e_1 > a$.*

Proof. Let suppose the opposite, i.e., $e_1 < a$. For $x = e_1,\ y = 0,\ z = 1$ assumed distributivity gives the following contradiction

$$e_1 = F(e_1, 0) = F(e_1, U(0,1)) = U(F(e_1, 0), F(e_1, 1)) = U(e_1, a) = a.$$

Therefore, $e_1 \geq a$.

If the assumption is now $e_1 = a$, then for $e_1 = a < x < e$ and $y = 0,\ z = 1$, from the distributivity law follows

$$a = F(x, 0) = F\big(x, U(0,1)\big) = U\big(F(x, 0), F(x, 1)\big) = U(a, x) = U(e_1, x) = x.$$

That is again a contradiction and, hence, $e_1 > a$. ■

The previous lemma shows that $e_1 > a$. The following one will explain relation between neutral elements e and e_1. Element e, as stated at the beginning of this section, is the neutral element of of the underlying uninorm of the observed S-uninorm, while element e_1 is the neutral element of the considered uninorm U from U_{min}.

Lemma 16 *Let F be a S-uninorm from U_{min} and let U be a uninorm from the class U_{min} with a neutral element $e_1 \in (0,1)$. If F is distributive over U then $e_1 = e$ or $e_1 < e$.*

Proof. Let suppose the opposite, that is that $e_1 > e$. For $e_1 < x < 1,\ y = e,\ z = 1$, the assumed distributivity leads to the following contradiction

$$x = F(x, e) = F(x, U(e, 1)) = U(F(x, e), F(x, 1)) = U(x, 1) = 1.$$

Therefore, either $e_1 = e$ or $e_1 < e$ holds. ■

The following theorem is the main result of this subsection.

Theorem 17 *Let F be a S-uninorm from U_{min} and U be a uninorm from the class U_{min} with a neutral element $e_1 \in (0,1)$ and underlying t-conorm S such that $S(x,x)$ is left-continuous at the point $x = e$. F is distributive over U if and only if $e_1 > a$ and exactly one of the following cases is fulfilled:*

(i) $e_1 = e$, *and U is an idempotent uninorm, i.e., $U = U_{e_1}^{min}$,*
(ii) $e_1 < e, U = U_{e_1}^{min}$, *and F is given by*

$$F(\mathrm{x}, \mathrm{y}) =$$

$$\begin{cases} aS'\left(\frac{x}{a},\frac{y}{a}\right) & on\ [0,a]^2, \\ a+(e_1-a)T_1'\left(\frac{x-a}{e_1-a},\frac{y-a}{e_1-a}\right) & on\ [a,e_1]^2, \\ e_1+(e-e_1)T_1''\left(\frac{x-e_1}{e-e_1},\frac{y-e_1}{e-e_1}\right) & on\ [e_1,e]^2, \\ e+(1-e)S_1\left(\frac{x-e}{1-e},\frac{y-e}{1-e}\right) & on\ [e,1]^2, \\ a & on\ [0,a]\times[a,1]\cup[a,1]\times[0,a], \\ \min(x,y) & otherwise, \end{cases} \tag{7}$$

where S_1 and S' are t-conorms, and T_1' and T_1'' are t-norms.

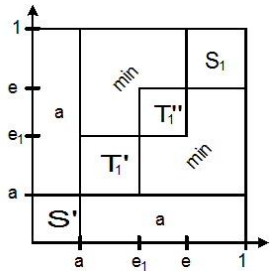


Figure 4
The form of the S-uninorm from Theorem 17

Proof. (⇒) Let F be a S-uninorm from the class $U_{\min}$ and let U be a uninorm from $U_{\min}$ that satisfy distributivity law. From Lemma 15 and Lemma 16 follows that $e_1 > a$, and either $e = e_1$ or $e > e_1$. The next step is to prove that U is an idempotent uninorm and it can be done analogously to the corresponding proof from [14]. Consequently, U is an idempotent uninorm, i.e., $U = U_{e_1}^{\min}$, and the claim (i) holds.

The next issue is the structure of F for $e > e_1$. The first step is to show that $F(e_1, e_1) = e_1$. Let $x = y = e_1, z = e$. From the assumed distributivity follows $e_1 = F(e_1, e) = F\big(e_1, U(e_1, e)\big) = U\big(F(e_1, e_1), F(e_1, e)\big) = U(F(e_1, e_1), e_1) = F(e_1, e_1)$.

For $a \le x \le e_1$,due to the distributivity law, holds

$$x = F(x, e) = F(x, U(e_1, e)) = U(F(x, e_1), F(x, e)) = U(F(x, e_1), x).$$

Since $F(x, e_1) \le F(x, e) = x \le e_1$ and $U = U_{e_1}^{\min}$, the following can be obtained

$$x = U(F(x, e_1), x) = \min(F(x, e_1), x) = F(x, e_1).$$

Also, for $e_1 \le x \le e$ holds $e_1 = F(e, e_1) \ge F(x, e_1) \ge F(e_1, e_1) = e_1$. Therefore,

$$F(x, e_1) = \begin{cases} x & for\ a \le x \le e_1, \\ e_1 & for\ e_1 \le x \le e. \end{cases} \tag{8}$$

Now, from (5) and (8) follows that F has to be of the form (7).

($\Leftarrow$) It is enough to prove the claim (ii), since the proof for the claim (i) is analogous. Therefore, let F be a S-uninorm given by (7) and $U = U_{e_1}^{\min}$. To prove the distributivity law, we have to consider $4^3 = 64$ cases. However, directly from the Lemma 10, distributivity for $x \in [0,1]$ and $(y,z) \in [0,e_1]^2 \cup [e_1,1]^2$ holds. Otherwise, for $y < e_1 < z, U(y,z) = y$ and $F(x,y) \leq F(x,z)$. Now, L will be used to denote the left side of distributivity law, i.e., $L = F(x,U(y,z)) = F(x,y)$. Also, the right side is denoted with R, i.e., $R = U\big(F(x,y),F(x,z)\big)$. As in [14], there are four cases for evaluation of the R: $x \geq e, e_1 \leq x \leq e, a \leq x \leq e_1$ and $x \leq a$. In all cases $R = \min(F(x,y),F(x,z)) = F(x,y)$ is obtained.

As seen above, in all considered cases $L = R$ is obtained, which proves that the distributivity law holds. ■

Remark 18 a) If the assumption of the of left-continuity for the function $S(x,x)$ at $x = e$ is omitted, the claim (i) from the previous theorem still holds,while the claim (ii), according to Theorem 14, is of the following form: *If F is distributive over U then $U(x,x) =$ for $x < e$ and F is given by (7).*

b) The restriction of the previous theorem to $a = 0$, i.e., to S-uninorm being just a uninorm from the class $U_{\min}$, has been shown in [18, 19]. The case (i) generalizes the Proposition 6.6 from [18, 19], and the case (ii) generalizes the Proposition 6.7 from [18].

3.3 S-uninorm vs. Uninorm from U_{max}

The second operator in this subsection is a disjunctive uninorm U with continuous functions (except perhaps at the point e) $U(x,0)$ and $U(x,1)$, i.e., a uninorm from the class $U_{\max}$. Now, as in [14], the following can be shown. Also, see [7].

Lemma 19 *Let F be a S-uninorm in U_{min} and let$U$ be a uninorm from the class U_{max} with a neutral element $e_1 \in (0,1)$. If F is distributive over U then $e_1 < a$.*

Theorem 20 *Let F be a S-uninorm in U_{min} and let U be a uninorm from the class U_{max} with a neutral element $e_1 \in (0,1)$ and underlying t-conorm S such that $S(x,x)$ is left-continuous at the point $x = e$. F is distributive over U if and only if $e_1 < a, U = U_{e_1}^{max}$ and F is given by*

$$F(\mathrm{x},\mathrm{y}) =$$

$$\begin{cases} e_1 S_1'\left(\frac{x}{e_1}, \frac{y}{e_1}\right) & on\ [0, e_1]^2, \\ e_1 + (a - e_1) S_2'\left(\frac{x-e_1}{a-e_1}, \frac{y-e_1}{a-e_1}\right) & on\ [e_1, a]^2, \\ a + (1-a) U'\left(\frac{x-a}{1-a}, \frac{y-a}{1-a}\right) & on\ [a, 1]^2, \\ a & on\ [0, a] \times [a, 1] \cup [a, 1] \times [0, a], \\ \max(x, y) & otherwise, \end{cases} \tag{9}$$

where S_1', S_2' are t-conorms, and U' is a uninorm from the class U_{min}.

Remark 21 According to [18] (see Lemma 6.5), if F is a uninorm from the class U_{min}, i.e., a S-uninorm in U_{min} with annihilator $a = 0$, then there is no uninorm U from the class U_{max} such that F is distributive over U. Theorem 20 shows that, when S-uninorm in U_{min} has annihilator $a \in (0,1)$, there is a uninorm $U = U_{e_1}^{max} \in U_{max}$ with neutral element $e_1 < a$ such that F is distributive over U.

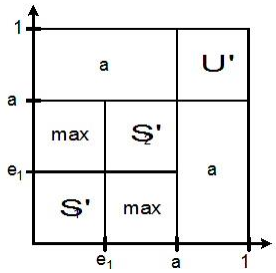


Figure 5
The form of the S-uninorm from Theorem 20

4 Distributivity Laws on Restricted Domain for S-Uninorms over Uninorms

Theorems from the previous section illustrate that the distributivity law (on whole domain) is a very strong condition since it considerably simplifies the structure of the inner operator. In this case, the inner operator is reduced to an idempotent operator. The research so farhas shown that restriction of the domain of the distributivity law can provide some new solutions that are non-idempotent. Therefore, this section contains the counterparts of theorems 13, 17, 20 from the previous section, now done for the restricted domain. Now, in order to characterize all pairs (F, G) satisfying (CD) condition, some kind of continuity for F and G has to hold (see [15]). The following results are counterparts to results from [14] and, for the sake of rounding up this topic, the proofs are not omitted.

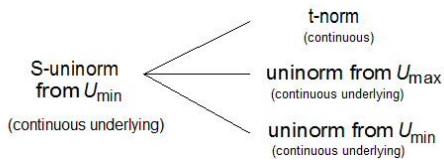


Figure 6
Topics of the Section 4

4.1 $\boldsymbol{S}$-uninorm vs. t-norm

Theorem 22 *Let F be a S-uninorm in U_{min} with continuous underlying t-conorm S', and let T be a continuous t-norm. F is conditionally distributive over T if and only if exactly one of the following cases is fulfilled:*

(i) $T = T_M$;
(ii) there is $c \in (0, a]$ *such that* T *is given by*

$$T(x,y) = \begin{cases} cT_L\left(\frac{x}{c}, \frac{y}{c}\right) & on\ [0,c]^2, \\ \min(x,y) & otherwise, \end{cases} \tag{11}$$

and F is given by

$$F(x,y) = \begin{cases} cS_P\left(\frac{x}{c}, \frac{y}{c}\right) & on\ [0,c]^2, \\ c + (a-c)S_1\left(\frac{x-c}{a-c}, \frac{y-c}{a-c}\right) & on\ [c,a]^2, \\ a + (1-a)U'\left(\frac{x-a}{1-a}, \frac{y-a}{1-a}\right) & on\ [a,1]^2, \\ a & on\ [0,a] \times [a,1] \cup [a,1] \times [0,a], \\ \max & otherwise, \end{cases} \tag{12}$$

where S_1 is a continuous t-conorm, and U' is a uninorm from the class U_{min}.

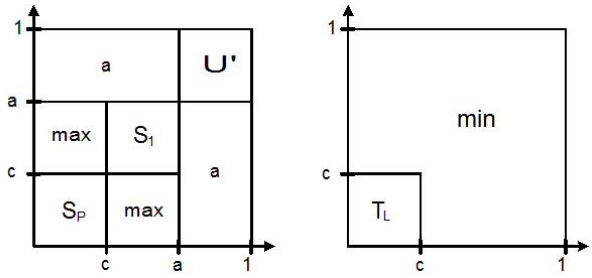


Figure 7
Conditionally distributive pair from Theorem 22

Proof. (⇒) Let F be conditionally distributive over T.

For $x \geq a$, as in Theorem 13, it can be shown that $T(x,x) = x$.

Let $x \leq a$. If $c \in (0,a]$ is an idempotent element of T, then, as in [14], there can be shown that all elements from $[c,a]$ are idempotents of T. Hence, either all elements from [0,1] are idempotent elements for t-norm T and, therefore $T = T_M = \min$, or there is the smallest nontrivial idempotent element $c \in (0,a]$ of T, i.e.,

$$T(x,y) = \begin{cases} cT^*\left(\frac{x}{c},\frac{c}{c}\right) & on\ [0,c]^2, \\ \min(x,y) & otherwise, \end{cases}$$

where T^* is a continuous Archimedean t-norm. Now, as in Theorem 5.21 from [15] (pp. 138-140), it can be proved that c is also an idempotent element of F, i.e., S-uninorm F on the square $[0,a]^2$ is of the following form

$$F(x,y) = \begin{cases} cS_2\left(\frac{x}{c},\frac{y}{c}\right) & if \quad (x,y) \in [0,c]^2, \\ c + (a-c)S_1\left(\frac{x-c}{a-c},\frac{y-c}{a-c}\right) & if \quad (x,y) \in [c,a]^2, \\ \max(x,y) & otherwise\ , \end{cases}$$

where S_1 and S_2 are continuous t-conorms. Also, in the same manner as in Theorem 5.21 from [15] (pp. 138-140), it can be obtained that T^* is a nilpotent t-norm, i.e., T is of the form (11), and that S_2 is a strict t-conorm such that F is of the form (12).

(⇐) Now, if the starting assumption is that T is a t-norm of the form (11) and F a S-uninorm of the form (12), it can be easily shown that condition (CD) holds. For input values from $[0,c]^2$ the problem is reduced to the pair (S_P, T_L) which satisfies (CD), and in all other cases it follows from Lemma 12.■

Example 23 Operator F given by

$$F(x,y) =$$

$$\begin{cases} \max & if \quad (x,y) \in [\frac{3}{5},1]^2 \cup [\frac{1}{4},\frac{1}{2}] \times [0,\frac{1}{2}] \cup [0,\frac{1}{2}] \times [\frac{1}{4},\frac{1}{2}], \\ x + y - 4xy & if \quad (x,y) \in [0,\frac{1}{4}]^2, \\ \frac{1}{2} & if \quad (x,y) \in [0,\frac{1}{2}] \times [\frac{1}{2},1] \cup [\frac{1}{2},1] \times [0,\frac{1}{2}], \\ \min & otherwise, \end{cases}$$

is a S-uninorm in $U_{\min}$ with annihilator $a = \frac{1}{2}$, obtained by (12) where $U' = U_{\frac{3}{5}}^{\min}$, $S_1 = \max$ and $c = \frac{1}{4}$. The corresponding t-norm is of the form (11).

4.2 S-uninorm vs. Uninorm from U_{min}

Theorem 24 *Let F be a S-uninorm in U_{min} with a continuous underlying t-conorm S', and let U be a uninorm from the class U_{min} with a neutral element $e_1 \in (0,1)$ and continuous underlying t-norm and t-conorm. F is conditionally distributive over U if and only if $e_1 > a$ and exactly one of the following cases is fulfilled:*

(i) $e_1 = e$, *and U is an idempotent uninorm, i.e., $U = U_{e_1}^{min}$,*

(ii) $e_1 = e$, *and there is a $c \in (0, a]$ such that F and U are given by*

$$U(x,y) = \begin{cases} \max(x,y) & on\ [e_1,1]^2, \\ cT_L\left(\frac{x}{c},\frac{y}{c}\right) & on\ [0,c]^2, \\ \min(x,y) & otherwise \end{cases} \tag{13}$$

and

$$F(x,y) = \begin{cases} cS_P\left(\frac{x}{c},\frac{y}{c}\right) & on\ [0,c]^2, \\ c + (a-c)S_2\left(\frac{x-c}{a-c},\frac{y-c}{a-c}\right) & on\ [c,a]^2, \\ a + (1-a)U'\left(\frac{x-a}{1-a},\frac{y-a}{1-a}\right) & on\ [a,1]^2, \\ a & on\ [0,a]\times[a,1] \cup [a,1]\times[0,a], \\ \max & otherwise, \end{cases} \tag{14}$$

(iii) $e_1 < e, U = U_{e_1}^{min}$, *and F is given by (7),*

(iv) $e_1 < e$, *and there is a $c \in (0, a]$ such that U is given by (13) and F is given by*

$$F(x,y) = \begin{cases} cS_P\left(\frac{x}{c},\frac{y}{c}\right) & on\ [0,c]^2, \\ c + (a-c)S_3\left(\frac{x-c}{a-c},\frac{y-c}{a-c}\right) & on\ [c,a]^2, \\ a + (e_1-a)T_1'\left(\frac{x-a}{e_1-a},\frac{y-a}{e_1-a}\right) & on\ [a,e_1]^2, \\ e_1 + (e-e_1)T_1''\left(\frac{x-e_1}{e-e_1},\frac{y-e_1}{e-e_1}\right) & on\ [e_1,e]^2, \\ e + (1-e)S_1\left(\frac{x-e}{1-e},\frac{y-e}{1-e}\right) & on\ [e,1]^2, \\ \max & on\ [c,a]\times[0,c] \cup [0,c]\times[c,a], \\ a & on\ [0,a]\times[a,1] \cup [a,1]\times[0,a], \\ \min & otherwise, \end{cases} \tag{15}$$

where U' is a uninorm from the class U_{min}, T_1', T_1'' are t-norms, S_1 is a t-conorm, and S_2, S_3 are continuous t-conorms.

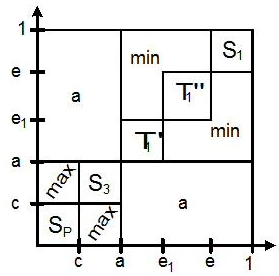


Figure 8
Operator *F* from Theorem 24 (iv)

Proof. (⇒) Let F be conditionally distributive over U. The first step is to prove that $e_1 > a$, which can be done by supposing the opposite (see [14]).

As in Lemma 16, it can be proved that either $e = e_1$ or $e > e_1$ holds. In the sequel it is supposed that $e > e_1$, since the case$e = e_1$ is similar.

For $x \geq a$, as in Theorem 17, holds $U(x, x) = x$ and the structure of F on the square $[a, 1]^2$ is given as in (15).

For $x \leq a$, as in Theorem 22, it can be proved that either U is an idempotent uninorm and F is given by (7), or there is a $c \in (0, a]$ such that U and F are given by (13) and (15), respectively.

(⇐) On the other hand, if the observed S-uninorm F and uninorm U are of forms (15) and (13), the (CD) condition can be proved as in Theorem 17.■

4.3 S-uninorm vs. Uninorm from U_{max}

Theorem 25 *Let F be a S-uninorm in U_{min} with a continuous underlying t-conorm S', and let U be a uninorm from the class U_{max} with a neutral element $e_1 \in (0,1)$ and continuous underlying t-norm and t-conorm. F is conditionally distributive over U if and only if $e_1 < a$ and exactly one of the following cases is fulfilled:*

(i) $U = U_{e_1}^{max}$, *and F is given by (9);*
(ii) there is a $c \in (0, e_1]$ *such that F and U are given by*

$$U(x,y) = \begin{cases} \max(x,y) & on\ (e_1,1] \times [0,1] \cup [0,1] \times (e_1,1], \\ cT_L\left(\frac{x}{c}, \frac{y}{c}\right) & on\ [0,c]^2, \\ \min(x,y) & otherwise, \end{cases} \tag{17}$$

and

$$F(x,y) =$$

$$\begin{cases} cS_P\left(\frac{x}{c},\frac{y}{c}\right) & on\ [0,c]^2, \\ c+(e_1-c)S_2''\left(\frac{x-c}{e_1-c},\frac{y-c}{e_1-c}\right) & on\ [c,e_1]^2, \\ e_1+(a-e_1)S_2'\left(\frac{x-e_1}{a-e_1},\frac{y-e_1}{a-e_1}\right) & on\ [e_1,a]^2, \\ a+(1-a)U'\left(\frac{x-a}{1-a},\frac{y-a}{1-a}\right) & on\ [a,1]^2, \\ a & on\ [0,a]\times[a,1]\cup[a,1]\times[0,a], \\ \max & otherwise, \end{cases} \tag{18}$$

where U' is a uninorm from the class U_{min}, and S_2', S_2'' are continuous t-conorms.

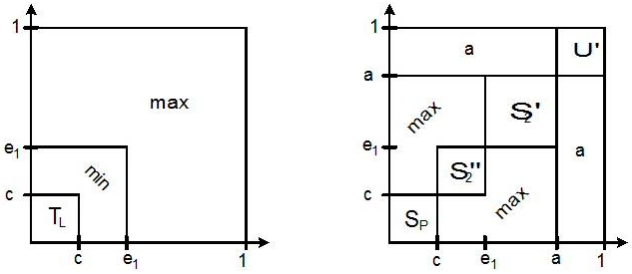


Figure 9
Conditionally distributive pair from Theorem 25 (ii)

Proof. (⇒) Let F be conditionally distributive over U.

As in Theorem 20, it can be shown that $e_1 < a$ and that, for $x \geq a$, holds $U(x,x) = x$.

Now, analogously to [14], it can be shown that $U = \max$ on the square $[e_1, 1]^2$.

The nex step is to show that e_1 is an idempotent element of F. For $x = e_1$, $z = e_1$, and an arbitrary $y \in (0, e_1)$ from equation (CD) follows

$$F(e_1, y) = F(e_1, U(y, e_1)) = U(F(e_1, y), F(e_1, e_1)).$$

Due to the assumption of continuity, the previous equality can be extended to $y = 0$ and $e_1 = F(e_1, 0) = U(e_1, F(e_1, e_1)) = F(e_1, e_1)$. Now, since $S' = F|_{[0,a]^2}$ is a continuous t-conorm immediately follows that S' is ordinal sum S_1' and S_2', i.e., F is given by (9). Therefore, $U = \max$ on the square $[e_1, 1]^2$ and F is given by (9).

For $x \leq e_1$,analogous to Theorem 22 for $x \leq a$, it can be proved that either U is an idempotent uninorm and F is given by (9), or there is a $c \in (0, e_1]$ such that U, F are given by (17) and (18) respectively.

(⇐) On the other hand, if the observed S-uninorm F and uninorm U are of forms (18) and (17), respectively, the (CD) condition can be shown as in Theorem 17. ■

Conclusion

Investigation of distributivity and conditional distributivity of a S-uninorm from $U_{\min}$ with an annihilator $a \in (0,1)$ is presented in this paper. The first set of results given in the third section concerns distributivity law on the whole domain and they extend and upgrade the corresponding ones from [18, 19]. Section 4 illustrates that the conditional distributivity produces a larger variety of solutions and the represented research is the continuation of investigations of conditional distributivity for aggregation operators with annihilator from [12, 17]. The further research will be directed towards possible application of the obtained structures to utility theory.

Acknowledgement

This work was supported by the Ministry of Education, Science and Technological Development of Republic of Serbia 174009.

References

[1] J. Aczél, Lectures on Functional Equations and their Applications, Academic Press, New York 1966

[2] T. Calvo, On some solutions of the distributivity equations, Fuzzy Sets and Systems 104 (1999) 85-96

[3] J. Dombi, Basic concepts for a theory of evaluation: The aggregative operator, European Journal of Operational Research 10 (1982) 282-293

[4] J. Dombi, On a certain class of aggregative operators, Information Sciences 245 (2013) 313-328

[5] J. Drewniak, P. Drygas, E. Rak, Distributivity between uninorms and nullnorms, Fuzzy Sets and Systems 159 (2008) 1646-1657

[6] D. Dubois, E. Pap, H. Prade, Hybrid probabilictic-possibilistic mixtures and utility functions, Preferences and Decisions under Incomplete Knowledge, Studies in Fuzziness and Soft Computing, Vol. 51, Springer-Verlag 2000, 51-73

[7] B. W. Fang, B. Q, Hu, Distributivity and conditional distributivity for S-uninorms, Fuzzy Sets and Systems, corrected proof

[8] Q. Feng, Z. Bin, The distributive equations for idempotent uninorms and nullnorms, Fuzzy Sets and Systems 155 (2005) 446-458

[9] J. C. Fodor, M. Roubens, Fuzzy Preference Modelling and Multicriteria Decision Support, Kluwer Academic Publishers, Dordrecht, 1994

[10] J. C. Fodor, R. R. Yager, A. Rybalov, Structure of uninorms, Internat. J. Uncertainty, Fuzziness and Knowledge-Based Systems 5 (1997) 411-427

[11] M. Grabisch, J. Marichal, R. Mesiar, E. Pap, Aggregations Functions, Cambridge University Press, 2009

[12] D. Jočić, I. Štajner-Papuga, Restricted distributivity for aggregation operators with absorbing element, Fuzzy Sets and Systems 224 (2013) 23-35

[13] D. Jočić, I. Štajner-Papuga, Some implications of the restricted distributivity of aggregation operators with absorbing elements for utility theory, Fuzzy Sets and Systems 291 (2016) 54-65

[14] D. Jočić, I. Štajner-Papuga, Distributivity and conditional distributivity for T-uninorms, Information Sciences 424 (2018) 91-103

[15] E. P. Klement, R. Mesiar, E. Pap, Triangular Norms, Kluwer Academic Publishers, Dordrecht, 2000

[16] G. Li, H.-W. Liu, Distributivity and conditional distributivity of uninorm with continuous underlying operators over a continuous t-conorm, Fuzzy Sets and Systems 287 (2016) 154-171

[17] G. Li, H.-W. Liu, Y. Su, On the conditional distributivity of nullnorms over uninorms, Information Sciences 317 (2015) 157-169

[18] M. Mas, G. Mayor, J. Torrens, The distributivity condition for uninorms and t-operators, Fuzzy Sets and Systems 128 (2002) 209-225

[19] M. Mas, G. Mayor, J. Torrens, Corrigendum to "The distributivity condition for uninorms and t-operators" [Fuzzy Sets and Systems 128 (2002), 209-225], Fuzzy Sets and Systems 153 (2005) 297-299

[20] M. Mas, R. Mesiar, M. Monserat, J. Torrens, Aggregation operations with annihilator, Internat. J. Gen. System 34 (2005) 1-22

[21] D. Ruiz, J. Torrens, Distributivity and conditional distributivity of uninorm and a continuous t-conorm. IEEE Transactions on Fuzzy Systems 14 (2) (2006) 180-190

[22] W. Sander, J. Siedekum, Multiplication, distributivity and fuzzy-integral I, II, III, Kybernetika 41 (2005) 397-422; 469-496; 497-518

[23] Y. Su, H.-W. Liu, D. Ruiz-Aguilera, J. Vicente Riera, J. Torrens, On the distributivity property for uninorms, Fuzzy Sets and Systems 287 (2016) 184-202

[24] Y. Su, W. Zong, H.-W. Liu, On distributivity equations for uninorms over semi-t-operators, Fuzzy Sets and Systems 299 (2016) 41-65

[25] Y. Su, W. Zong, H.-W. Liu, P. Xue, The distributivity equations for semi-t-operators over uninorms, Fuzzy Sets and Systems 287 (2016) 172-183

[26] M. Takács, Approximate Reasoning in Fuzzy Systems Based on Pseudo-analysis and Uninorm Residuum, Acta Polytechnica Hungarica 1 (2004) 49-62

[27] R. R. Yager, A. Rybalov, Uninorm aggregation operators, Fuzzy Sets and Systems 80 (1996) 111-120

Cover Processing-based Steganographic Model with Improved Security

Daniela Stănescu[1], Mircea Stratulat[1], Romeo Negrea[2], Ioana Ghergulescu[3]

[1]Computer Department, Politehnica University of Timisoara, 2 Vasile Parvan Boulevard, 300223 Timisoara, Romania (e-mail: daniela.stanescu@cs.upt.ro, mircea.stratulat@cs.upt.ro)

[2]Department of Mathematics, Politehnica University of Timisoara, 300006 Timisoara, Romania (e-mail: romeo.negrea@mat.upt.ro)

[3]Adaptemy, 27 Mount Street Lower, Dublin 2, Ireland (e-mail: ioana.ghergulescu@adaptemy.com)

Abstract: Steganography uses specialised techniques to conceal messages in different cover objects such as image or video so that only the sender and receiver know of the message's existence and are able to decipher it. Previous research conducted in the area has mainly focused on steganography and steganalysis techniques. This paper proposes a new model for steganography called Cover Processing-based Steganographic Model (CPSM) that processes the cover objects and transmits them in a way to improve the security of steganographic objects. A comprehensive demonstration based on information theory proves that CPSM provides improved security in terms of lower relative entropy as compared to previous models from the literature. Moreover, experimental tests show a decrease of the relative error between the cover and steganographic objects of up to 14%.

Keywords: steganographic model; cover processing; secret communication; security improvement; entropy

1 Introduction

The exchange of information plays a central role in many applications, with the Internet being the most representative example. As there is growing number of cyberattacks that affect businesses and end-users [1], the security of information storage and communication has become increasingly important. A recent study by IBM and Ponemon Institute showed that the average cost of a data breach was $3.79 million in 2015, while another study by Juniper Research forecasted that cybercrime will be a $2.1 trillion problem by 2019 [2].

In this context, there has been increased research interest on methods for securing information transmission such as steganography, watermarking and cryptography [3], [4]. While historical evidence suggests that steganography and cryptography methods have been applied since ancient times [5], their popularity and applicability were especially accelerated by the digital revolution of the past few decades [6]. Despite the fact that steganography, cryptography and watermarking are all methods for securing information, there are notable differences between them. Cryptography focuses on securing the information by making it illegible without having the proper key [7]. As opposed, steganography focuses on hiding the important information within another carrier, making it invisible to an observer. Based on the carrier type steganography can be divided into text or linguistic steganography [8], digital media steganography based on video, audio or images [9]–[11], as well as network steganography that exploits communication protocols [12]. While watermarking is also a method for embedding information, it differs from steganography in the sense that it is focused on protecting to carrier, and not the secret information [13].

Significant research effort was also dedicated to steganalysis methods[14]–[16]. Steganalysis represents the art of detecting the presence of hidden information, and depending on what the end goal is, to further determine the type of steganography, to extract the secret message or to tamper it so that the receiver can no longer extract it [14]. Therefore, steganographic systems must be both secure and robust to tampering by an active attacker or to artifacts that could result in the loss of the secret message such as network transmission errors.

Security represents the most important criteria of steganographic systems, with a system being considered secure if the existence of the message cannot be determined with higher probability than a random guessing. Existing approaches to quantify the security of steganographic systems include: information theory-based approach that considers the relative entropy or the difference between two probability distributions; ROC-based approach that considers the difference between true positive and false positive classification rates; and statistics-based approach that considers the maximum mean discrepancy to test if two samples are generated from the same distribution [14].

This paper proposes a new steganographic model called Cover Processing-based Steganographic Model (CPSM) that improves the security of steganographic objects by processing the carrier. The main advantage of the CPSM model is that the cover processing makes more difficult to detect and extract the message for an attacker. In extreme cases, it could reach a point where the detection would be too costly for an attacker. A comprehensive mathematical demonstration proves that CPSM provides improved security in terms of lower relative entropy as compared to previous models proposed in the literature. Furthermore, experimental testing shows that applying simple processing such as shifting the binary information of the cover image can lead to a decrease of the relative error between the cover and steganographic objects of up to 14%.

The rest of the paper is structured as follows. Section 2 presents related works in the area of steganography. Section 3 describes the proposed CPSM, while Section 4 presents the theoretical demonstration of the model using information theory. Section 5 presents the results of the experimental tests.

2 Related Work

Steganography has been the focus of much research interest over the past few decades, as well as increasing applicability into the real world [6]. A multitude of papers (e.g., [3], [4], [12], [14]–[22]), have reviewed the various techniques proposed in the literature for different types of steganography such as text, image, audio, video, or network steganography. Analysing those papers, one can note that past research works have mainly focused on specialised steganography and steganalysis techniques, with few generic models having been proposed.

Steganography as a method of hiding information was initially best described by Simmons in the prisoners' problem [23]. In this problem there are two prisoners that want to communicate. The only way of communication is via messages exchanged through an open channel, a warden. The warden will allow the message exchange as long as the information is open for inspection and there is no suspicion of hidden information. Furthermore, the warden will try to detect and intercept any suspicious messages. In order to communicate the prisoners will have to find a way of hiding information into innocent messages.

Zöllner et al. [24] proposed a basic embedding model that aimed to represent a steganographic system in an abstract and generic form. Figure 1 illustrates the basic embedding model for the case of image steganography. The model highlights that the sender wants to transmit a secret message m to a receiver. As the communication channel is not secure, the sender will use an innocent cover object C, in which it will hide the message using an embedding steganographic function f_E. The embedding process will result in the steganographic object S. For improved security, the system makes use of a steganographic key k that is passed as a parameter to the embedding function. The receiver will use an extraction function f_E^{-1} that will output the message m^* and the cover object C^*. If the extraction process is correct the message m^* will be the same as m. The authors also make use of information theory to model the security of a steganographic system. For a system to be considered secure, the embedding function should create a steganographic object S that has the same entropy as the cover object C (where the entropy $H(S)$ describes the uncertainty about S). However, the authors concluded that this cannot be achieved in practice assuming that the attacker can access and compare the cover and steganographic objects. Moreover, the authors concluded that only indeterministic steganography can be secure, by introducing a level of uncertainty about the cover that is higher than the entropy of the secret message.

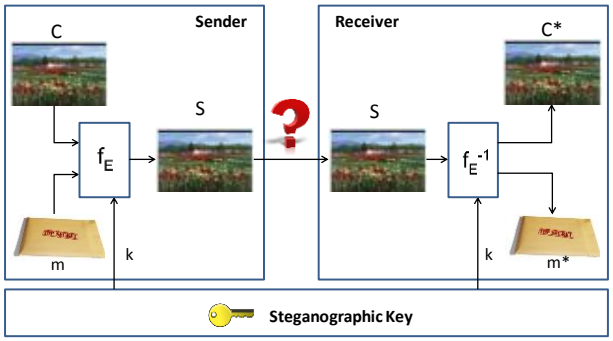


Figure 1
Basic embedding model of a steganographic system

Cachin [25] proposed an information-theoretic model for steganography that quantifies the security of the system in terms of the relative entropy between the probability distributions of the cover object C and steganographic object S. The author assumes that the sender avails of a set of innocent cover objects, in which it randomly embeds the secret message. The sender transmits steganographic objects or simply cover objects which have the purpose to confuse the attacker. The author also assumes that a passive attacker has complete access to the communication channel and has knowledge of the embedding function and of the cover object (i.e., knows the probability distribution of C). For a steganographic system to be secure the attacker should not be able to distinguish computationally between the cover and steganographic objects (i.e., the relative entropy to be ideally 0, or smaller thanεin case of anε-secure system). As the receiver would also not be able to detect the steganographic objects, the author proposes to use an oracle where the receiver has knowledge of when the sender is active. While this model presents much value from a theoretical point of view, the many assumptions limit its applicability in real-world steganographic systems.

Sallee [26] also proposed an information-theoretic model that uses statistical information of the cover object. The author also proposed a generic method to determine the maximum embedding capacity of the cover object while being resistant to first order statistical attacks, and further demonstrated the applicability of the model to JPEG images.

Raphael and Sundaram [27] have proposed a model that combines cryptography with steganography in order to increase the security of data communication. First, the secret message is encrypted using either secret or public cryptography key, and then embedded in the cover object using the steganographic key. In [28] the authors added another layer of protection to the model and proposed to transform the encrypted text into Unicode before hiding it into the cover image. However, while the authors have implemented a prototype and explained its functionality, they did not conduct a comprehensive evaluation of the proposed model.

Schöttle and Böhme [29] have proposed a universal game-theoretic framework to model adaptive embedding steganography systems which are considered to provide additional security as compared to systems based on random embedding. The model identifies the optimal adaptive embedding strategy that will maximise the security against attackers who would anticipate the adaptivity. The authors demonstrate that for real-world imperfect steganography systems the optimal embedding strategy is between naive adaptive and random uniform embedding.

Fakhredanesh et al. [30] have proposed a solution to overcome the perceptual detectability limitation of steganography systems based on cover image statistic models. By using Watson's human visual system model to compute the maximum acceptable changes in each DCT coefficients, the authors showed that steganographic objects with improved security and visually imperceptible changes can be obtained.

Song et al. [31] have proposed a digital steganography model based on additive noise and an embedding optimisation strategy aimed at providing guidance for the design of steganographic algorithms. The optimisation is done in terms of embedding modification position and direction. The authors have also validated experimentally that the proposed embedding optimisation technique can improve the security of steganographic algorithms such as LSBM and MG.

Denemark and Fridrich [32] have proposed a model-based embedding steganography method that makes use of multivariate Gaussian model to better estimate the acquisition noise, an important random aspect that makes digital images and videos suitable for steganography.

3 CPSM Overview

This section describes the proposed Cover Processing-based Steganographic Model (CPSM), that processes the cover objects and transmits them in a way to improve the security of steganographic objects from both a mathematical and practical point of view.

Figure 2 presents the functional block-level diagram of the CPSM model. The model pre-requisite is that the sender avails of a set of original cover objects C_R, which can be processed to create cover objects C that will be used in the steganographic process. The cover objects C are obtained by processing each original object C_R with the help of a processing function f_p. To confuse a possible attacker, the sender selects and incorporates the secret message only in some of the cover objects, which become steganographic objects S. However, the entire set of cover objects including those without hidden information are sent to receiver.

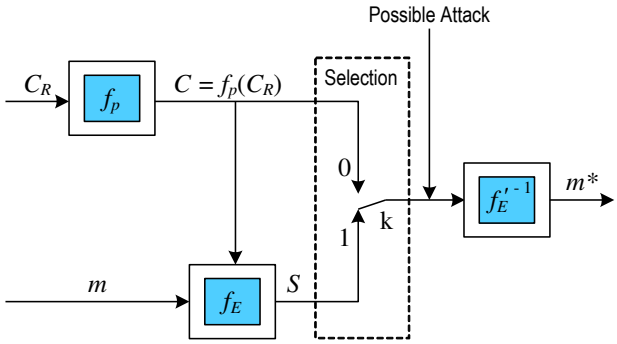


Figure 2
Functional diagram of the CPSM model

The selection of the cover objects is done with the help of a switch k. If the switch is on "0", then the cover object $C$is sent to the receiver. This does not contain any secret information but will help confuse the attacker. If the switch is on "1", then the steganographic object $S$is sent to the receiver. The operation of the switch is controlled according to a function known by both the sender and the receiver.

The processing function f_p can be based on an algorithm known by the sender, but depending on the available ways to improve the entropy of the resulting object this may not necessarily be known by the receiver. Embedding additional information into the cover object will increase its entropy, and a possible attacker could notice this increase if the cover object is not carefully selected.

Therefore, the changes made using the processing function will be done in such a way that the original cover object will not differ too much from the processed object. Transformations that could be applied through the processing function f_p include: applying noise, shifting the binary information towards higher or lower values, etc. As these transformations are applied in the same way to all of the original cover objects, the entropy increases for all objects not only for those that will be later transformed into steganographic objects. As such, it will be more difficult for a possible attacker to identify the transmitted objects containing the secret message. The critical condition is for the attacker not to have access to the original cover objects C_R.

The secret message m is embedded in some of the processed cover objects by applying the steganographic function f_E. Following this step, the complete set of objects, including steganographic objects S as well as processed cover objects $C$are sent to the receiver.

On another side, to extract the secret message the receiver will apply the inverse decoding function $f_E'^{-1}$, which consists of the inverse processing function f_p^{-1} composed with the inverse steganographic function f_E^{-1}. The composition of the two functions is done in such way that the output message m^*would be obtained

in a format as similar as possible to that of the original message m. Moreover, steganographic keys could be used to make it more difficult for an attacker to extract the secret message. However, in case of pure steganography it is not mandatory to use keys, as long as the steganographic algorithms are carefully selected [33], [34].

The next section demonstrates from a mathematical point of view how the security of steganographic systems is improved by the proposed CPSM model.

4 Theoretical Demonstration of CPSM

The aim of this section is to demonstrate that the proposed CPSM model provides an improved security as compared to the information theoretic model proposed by Cachin [25]. Cachin's approach is the most suitable for demonstrating the efficiency of steganographic systems from a probabilistic point of view. Other approaches from the literature review have only used simulations or empirical experiments to demonstrate the improved performance of their steganographic methods. According to Cachin, a steganographic object is perfectly secure if it meets the condition:

$$D(P_C \parallel P_S) = 0 \tag{1}$$

where, P_C is the probability distribution of the cover object C, while P_S is the probability distribution of the steganographic object S.

Moreover, $D(P_C \parallel P_S)$ represents the *relative entropy*, a measure of the difference between the two probability distributions P_C and P_S that characterise the steganographic process. The relative entropy is defined based on the Kullback–Leibler divergence [35]as in equation (2), where the units of entropy are bits and the log is logarithm to the base 2.

$$D(P_C \parallel P_S) = \sum_{c \in C} P_C(c) \cdot log \frac{P_C(c)}{P_S(c)} \tag{2}$$

If the condition from equation (1) is met there is no difference between the two probability distributions, and thus an attacker cannot distinguish between the cover object C and the steganographic object S. In this case, the attacker needs to analyse all the objects sent (C and S) and will not be able to extract in real time the hidden message from S using a polynomial algorithm. If there are differences between P_S and P_C, the attacker can focus only on the steganographic objects and will be able to extract the hidden message from S using a polynomial algorithm.

As perfect steganography is difficult to achieve in practice, it is desired to have a probability distribution P_S as close as possible to P_C. In this context, Cachin [25]defines a steganographic system to be ε-secure if:

$$D(P_C \parallel P_S) \leq \varepsilon \tag{3}$$

The smaller ε is, the harder will be for the attacker to distinguish between C and S, thus the harder to extract the hidden message from S.

Let k represent the switch from Figure 2 that can take two values:

$$k = \begin{cases} 0, & \text{if } c \in C_0 \\ 1, & \text{if } c \in C_1 \end{cases} \tag{4}$$

where the C alphabet is defined as:

$$C = C_0 \oplus C_1 \tag{5}$$

which means that C_0 and $C_1$are partitions of C, with C_0representing the subset when cover objects are transmitted to the receiver and C_1 representing the subset when steganographic objects are transmitted to the receiver, as such:

$$C_0 \cup C_1 = C, \text{ respectively } C_0 \cap C_1 = \emptyset \tag{6}$$

According to [25], in the above case a steganographic system is ε-secure for:

$$\varepsilon = \delta^2 / \ln 2 \tag{7}$$

where:

$$\delta = \Pr[c \in C_0] - \Pr[c \in C_1] \tag{8}$$

In equation (8), Pr denotes probability, while$\delta > 0$ because $\Pr[c \in C_0] > \Pr[c \in C_1]$, otherwise there would be big differences between P_S and P_C.

All of these are demonstrated starting from the following relationship:

$$P_S(c) = \begin{cases} \dfrac{P_C(c)}{1+\delta}, & \text{if } c \in C_0 \\ \dfrac{P_C(c)}{1-\delta}, & \text{if } c \in C_1 \end{cases} \tag{9}$$

which results by partitioning $C = C_0 \oplus C_1$ based on the total probability expressed as in equation (10), and conditional probability expressed as in equation (11) [36].

$$P(A) = \sum_{i \in I} P(A_i) \cdot P(A|A_i) \tag{10}$$

where, $i \in I$ indexes $A_i$mutually exclusive and exhaustive partitions of A.

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \text{ or } P(A \cap B) = P(A|B) \cdot P(B) \tag{11}$$

Indeed, we have:

$$\Pr[S = c] \overset{def}{=} \Pr[S = c | c \in C_0] \cdot \Pr[c \in C_0] + \Pr[S = c | c \in C_1] \cdot \Pr[c \in C_1] \tag{12}$$

However,

$$\begin{aligned}\Pr[S = c | c \in C_0] &= \Pr[C = c | c \in C_0 \text{ or } c \notin C_1] \\ &= \frac{\Pr[C = c \cap (c \in C_0 \text{ or } c \notin C_1)]}{\Pr[c \in C_0 \text{ or } c \notin C_1]} \\ &= \frac{\Pr(C = c)}{1 + \delta}\end{aligned} \tag{13}$$

if $c \in C_0$, because:

$$\begin{aligned}1 + \delta &= 1 + \Pr[c \in C_0] - \Pr[c \in C_1] \\ &= \Pr[c \in C_0] + 1 - \Pr[c \in C_1] \\ &= \Pr[c \in C_0] + \Pr[c \notin C_1] \\ &= \Pr[c \in C_0 \text{ or } c \notin C_1]\end{aligned} \tag{14}$$

Similarly:

$$\begin{aligned}\Pr[S = c | c \in C_1] &= \Pr[C = c | c \notin C_0 \text{ or } c \in C_1] \\ &= \frac{\Pr[C = c \cap (c \notin C_0 \text{ or } c \in C_1)]}{\Pr[c \notin C_0 \text{ or } c \in C_1]} \\ &= \frac{\Pr(C = c)}{1 - \delta}\end{aligned} \tag{15}$$

if $c \in C_1$, because:

$$\begin{aligned}1 - \delta &= 1 - \Pr[c \in C_0] + \Pr[c \in C_1] \\ &= \Pr[c \notin C_0] + \Pr[c \in C_1] \\ &= \Pr[c \notin C_0 \text{ or } c \in C_1]\end{aligned} \tag{16}$$

Moreover,

$$\Pr[c \in C_0] = \begin{cases} 1, & \text{if } c \in C_0 \\ 0, & \text{if } c \in C_1 \end{cases} \tag{17}$$

$$\Pr[c \in C_1] = \begin{cases} 0, & \text{if } c \in C_0 \\ 1, & \text{if } c \in C_1 \end{cases} \tag{18}$$

According to [24], steganographic systems cannot be secure if an attacker knows C and S, thus being able to compare two objects that are similar but still contain different information. To address this issue, the authors introduce a degree of uncertainty to the cover object C. This will confuse the attacker, as the comparison will be done between the steganographic object S containing hidden information, and a cover object C that the attacker does not know and only estimates how it looks like. We will investigate the behaviour of entropy, which characterises both elements considered in the comparison by the attacker, namely: information and uncertainty.

An example in this sense would be capturing a photo and using it as a medium for transmitting some secret information. The sender will choose the scene and will capture it using a photo camera. The original photo representing the cover object C is processed to incorporate the secret message becoming the steganographic

object S, which is sent over an unsecured channel. When intercepted by an attacker, this recognises the scene represented in the photo but does not have access to the original photo C, hence the uncertainty. As the original photo C is not available, the attacker cannot compare it with the intercepted steganographic object S, and will not be able to extract the secret message from S.

In order to obtain the original cover object, one approach would be for the attacker to identify the scene captured in the photo, make similar photos and compare them with S. As digital camera sensors are sensitive to factors that cannot be accurately controlled such as temperature, the attacker would notice small differences even between photos of the scene captured consecutively. Therefore, if differences are noticed due to uncontrolled factors no matter how many attempts are made to obtain the cover object C, the attacker might conclude that it is normal for the steganographic object S to also present differences. As opposed, if all captured photos are identical and only S presents differences, the attacker might think that S contains a hidden message.

As such, the steganographic model proposed by Zöllner et al. [24] involves choosing a cover object that is unknown to a possible attacker, and pre-processing it using different equipment or digital techniques before being used to create the steganographic object. However, the authors do not demonstrate that the model provides improved security. The steganographic model proposed by Cachin [25] involves choosing a set of cover objects, with only some of them being used to create the steganographic object. In case of this model, the sender transmits both the cover and the steganographic objects to the receiver.

The CPSM steganographic model proposed in this paper involves choosing a set of cover objects that are individually processed, and only some of them are used to create steganographic objects. Next, we will prove mathematically that applying a processing function on the cover objects can improve the security of steganographic systems.

As illustrated in Figure 2 the CPSM model applies a processing function f_p on each cover object. By applying this function, the relative error ε will decrease to be lower than the value obtained by Cachin.

Suppose that the chosen processing function takes the form:

$$f_p(x) = a \cdot x, \text{ where } a > 1 \tag{19}$$

Using this function, the set of cover objects C can be obtained based on the initial set C_R, as follows:

$$C = f_p(C_R) \tag{20}$$

Following the processing we will prove that:

$$\varepsilon = \frac{1}{a^2} \cdot \delta^2 / \ln 2 \tag{21}$$

thus, the ε measure obtained is lower than the one obtained by Cachin and presented in equation (7).

To prove equation (21), we start from Theorem 1 and Theorem 2 proposed in [25], according to which:

$$D(P_C \parallel P_S) \leq D(P_{C_R} \parallel P_S) \leq \delta^2 / \ln 2 \tag{22}$$

By performing a change of variable on equation (2) the relative entropy can be expressed in terms of the new variable d, as:

$$D(P_C \parallel P_S) = \sum_{d \in C} P_C(d) \cdot \log \frac{P_C(d)}{P_S(d)} \tag{23}$$

However,

$$P_S(d) = \begin{cases} \dfrac{P_C(d)}{1+\delta}, & \text{if } d \in C_0 \\ \dfrac{P_C(d)}{1-\delta}, & \text{if } d \in C_1 \end{cases} \tag{24}$$

where:

$$\delta = \Pr[d \in C_0] - \Pr[d \in C_1] \tag{25}$$

Therefore,

$$D(P_C \parallel P_S) = \sum_{d \in C_0} P_C(d) \cdot \log(1+\delta) + \sum_{d \in C_1} P_C(d) \cdot \log(1-\delta) \tag{26}$$

For $C = f_p(C_R)$ we have:

$$P_C(d) = \left| \frac{1}{f_p{}'(f_p^{-1}(d))} \right| \cdot P_{C_R}(f_p^{-1}(d)) \tag{27}$$

if:

$$f_p(x) = a \cdot x \Rightarrow \begin{cases} f_p^{-1}(x) = \dfrac{x}{a} \\ f_p{}'(x) = a \end{cases} \tag{28}$$

The following notation is adopted:

$$f_p^{-1}(d) = \frac{d}{a} = c \tag{29}$$

where $d \in C$, which implies that:

$$c \in C_a = \frac{1}{a} C \tag{30}$$

In the context of the paper, C represents the set of pixels from an image, thus C_a represents the set of pixels scaled with the a constant. Therefore, based on equations (26) and (27) results that:

$$\begin{aligned} D(P_C \parallel P_S) &= \frac{1}{a}\sum_{d\in C_0} P_{C_R}\left(\frac{d}{a}\right)\cdot\log(1+\delta) + \frac{1}{a}\sum_{d\in C_1} P_{C_R}\left(\frac{d}{a}\right)\cdot\log(1-\delta) \\ &= \frac{1}{a}\sum_{c\in C_0}\frac{1}{a}P_{C_R}(c)\cdot\log(1+\delta) + \frac{1}{a}\sum_{c\in C_1}\frac{1}{a}P_{C_R}(c)\cdot\log(1-\delta) \\ &= \frac{1}{a^2}\sum_{c\in C_0} P_{C_R}(c)\cdot\log(1+\delta) + \frac{1}{a^2}\sum_{c\in C_1} P_{C_R}(c)\cdot\log(1-\delta) \\ &= \frac{1}{a^2}\cdot\left[\frac{1+\delta}{2}\cdot\log(1+\delta) + \frac{1-\delta}{2}\cdot\log(1-\delta)\right] \end{aligned} \tag{31}$$

because,

$$\sum_{c\in C_0} P_{C_R}(c) = \frac{1+\delta}{2} \tag{32}$$

$$\sum_{c\in C_1} P_{C_R}(c) = \frac{1-\delta}{2} \tag{33}$$

Moreover, using the fact that

$$\log(1+x) \le \frac{x}{\ln 2} \tag{34}$$

results:

$$\begin{aligned} D(P_C \parallel P_S) &\le \frac{1}{a^2}\left(\frac{1+\delta}{2}\cdot\frac{\delta}{\ln 2} + \frac{1-\delta}{2}\cdot\frac{-\delta}{\ln 2}\right) \\ &\le \frac{1}{a^2}\cdot\frac{\delta^2}{\ln 2} \end{aligned} \tag{35}$$

On another side:

$$\begin{aligned} D(P_{C_R} \parallel P_S) &= \sum_{c\in C_0} P_{C_R}(c)\cdot\log(1+\delta) + \sum_{c\in C_1} P_{C_R}(c)\cdot\log(1-\delta) \\ &= \frac{1+\delta}{2}\cdot\log(1+\delta) + \frac{1-\delta}{2}\cdot\log(1-\delta) \\ &\le \frac{1+\delta}{2}\cdot\frac{\delta}{\ln 2} + \frac{1-\delta}{2}\cdot\frac{-\delta}{\ln 2} \\ &\le \frac{\delta^2}{\ln 2} \end{aligned} \tag{36}$$

where,

$$\delta = P_{C_R}[c|c \in C_0] - P_{C_R}[c|c \in C_1] \tag{37}$$

and:

$$\begin{aligned} 1 + \delta \quad &= P_{C_R}[c|c \in C_0] + 1 - P_{C_R}[c|c \in C_1] \\ &= P_{C_R}[c|c \in C_0] + P_{C_R}[c|c \notin C_1] \end{aligned} \tag{38}$$

If $|C_0| = |C_1|$, then:

$$\sum_{c \in C_0} P_{C_R}(c) = \sum_{c \in C_1} P_{C_R}(c) = \frac{1}{2} \tag{39}$$

because,

$$\sum_{c \in C} P_{C_R}(c) = 1 \tag{40}$$

Finally, based on equations (31) and (36) results that:

$$\frac{D(P_C \parallel P_S)}{D(P_{C_R} \parallel P_S)} = \frac{1}{a^2} \tag{41}$$

where $a > 1$.

The conclusion that can be drawn from the theoretical demonstration is that processing the cover object with a coefficient where $a > 1$, leads to a decrease by $1/a^2$in the relative entropy between the probability distribution of the steganographic object obtained and the probability distribution of the cover object. Therefore, the proposed CPSM model enables improved security of steganographic systems thought three different aspects: (i) the generation of a set of cover objects that are known by the sender but not necessarily known by the receiver, (ii) the individual processing of the cover objects, and (iii) the random selection of one or multiple cover objects in which to embed the secret messages.

In order to support the receivers, it is desired to inform them about the procedure used for selecting the cover objects that will be used as steganographic objects. If this information is missing, the receiver will have to apply the inverse decoding function $f_E'^{-1}$ for the full set of received objects, thus requiring a longer time to retrieve the hidden message. It is also possible that multiple messages that are retrieved to have significance, thus confusing the receiver. To avoid such situations, one solution would be to inform the receiver about the function used in order to select the cover objects used in the steganographic process.

5 Experimental Validation of CPSM

A number of experimental tests were conducted in order to validate the proposed CPSM model from an empirical perspective. The Segment Compression Steganographic Algorithm (SCSA) proposed in [37] was used for the experimental testing. SCSA is based on the Karhunen-Loève Transform (KLT) that is widely considered to achieve optimal signal processing for data representation, compression and analysis. A detailed description of the algorithm can be found in [37].

A multitude of colour images with different size and content characteristics were used as cover objects. The secret messages were also represented by colour images that were incorporated within the cover objects on the least important bits. In line with the principle of the CPSM model, the cover objects were first processed by applying a number of transformations. In particular, the binary information of each pixel was shifted with a number of steps towards black, and respectively white.

Tables 1 to 3 present the experimental results for the three scenarios considered for hiding the secret messages (i.e., on the least important 1, 2 and 4 bits of the cover objects). Columns 2 to 5 present the name and size in pixels of the cover objects and secret messages. Columns 6 to 10 present the computed relative errors between the cover objects and the steganographic objects for five different cases: the pixels binary information was shifted towards black with a value of 10 and respectively 6, the cover object was not processed, and the pixels binary information was shifted towards white with a value of 6 and 10. The last column presents the improvement (as percentage) of the relative error that was achieved through the processing of the cover object.

The results analysis shows that processing the cover objects can decrease the relative error between the cover and the steganographic objects. In particular, shifting the pixels binary information towards black leads to a decreased relative error, for all three test scenarios using the SCSA algorithm to hide the message on 1, 2 and 4 bits. The results show that the maximum improvement of the relative error was 13.68% in case of the 'sphinx' cover object and 'Hawk' secret message using SCSA on 4 bits. While for some cases the improvement of the relative error is not significant, one observation made was that in such cases the cover objects usually presented large areas with the same information (e.g., background). Therefore, one can safely conclude that such cover objects are not recommended for steganography.

Figure 3 illustrates the processing for one example considered in the experimental testing (i.e., 'Wildflowers' cover object and 'watch' secret message using SCSA on 4 bits). The images show that the difference between the cover and steganographic objects is unnoticeable, for all three scenarios: unprocessed cover object, processing towards black and towards white respectively. In terms of the

relative error between the cover and steganographic object, the improvement achieved was 2.09% (see Table 3, line 2).

The experimental results validate that the CPSM model can lead to better steganographic objects and thus improved security, as compared to not processing the cover objects.

Table 1

Cover object processing experimental results using the SCSA algorithm on 1 bit

	Cover Object		Secret Message		Cover Object Processing					
Seq.	**Name**	**Size [px]**	**Name**	**Size [px]**	ε^{-10}	ε^{-6}	ε^{0}	ε^{6}	ε^{10}	**%**
1	lena	256x256	firefox	128x128	0.19735	0.19744	0.19745	0.19760	0.19778	2.17
2	Aquaria	256x256	firefox	128x128	0.19558	0.19611	0.19657	0.19652	0.19664	5.42
3	dogs	640x480	wildflowers	200x135	0.19510	0.19530	0.19592	0.19595	0.19605	4.86
4	dogs	640x480	watch	200x135	0.19506	0.19545	0.19617	0.19637	0.19689	9.38
5	fruit	512x512	lena	256x256	0.19428	0.19460	0.19557	0.19635	0.19664	1.21
6	fruit	512x512	Aquaria	256x256	0.19413	0.19452	0.19554	0.19635	0.19670	1.32
7	Lena512	512x512	Aquaria	256x256	0.19630	0.19631	0.19631	0.19638	0.19646	0.08
8	Lena512	512x512	lena	256x256	0.19619	0.19620	0.19620	0.19625	0.19628	0.04
9	building	640x480	wildflowers	200x135	0.18922	0.19062	0.19504	0.19772	0.19850	4.10
10	building	640x480	watch	200x130	0.18731	0.18930	0.19586	0.20201	0.20422	9.02
11	Alicia	1024x1024	Lena512	512x512	0.19419	0.19505	0.19576	0.19576	0.19576	0.8
12	Alicia	1024x1024	fruit	512x512	0.19102	0.19402	0.19566	0.19566	0.19566	2.4
13	Alicia	1024x1024	dogs	640x480	0.19081	0.19384	0.19565	0.19565	0.19565	2.5
14	car	1024x1036	dogs	640x480	0.19457	0.19459	0.19459	0.19461	0.19461	0.02
15	car	1024x1036	fruit	512x512	0.19400	0.19402	0.19402	0.19402	0.19403	0.015
16	car	1024x1036	Leno512	512x512	0.19638	0.19639	0.19639	0.19639	0.19640	0.01
17	football	1600x1200	building	640x480	0.19305	0.19416	0.19574	0.19629	0.19641	1.74
18	football	1600x1200	hawk	800x600	0.19301	0.19426	0.19590	0.19646	0.19660	1.86
19	football	1600x1200	sphinx	800x600	0.19435	0.19493	0.19583	0.19609	0.19617	0.9
20	fish	1600x1200	building	640x480	0.19373	0.19454	0.19559	0.19575	0.19586	1.1
21	fish	1600x1200	hawk	800x600	0.19353	0.19437	0.19552	0.19568	0.19580	1.17
22	fish	1600x1200	sphinx	800x600	0.19497	0.19538	0.19585	0.19591	0.19596	0.5

Table 2

Cover object processing experimental results using the SCSA algorithm on 2 bits

	Cover Object		Secret message		Cover Object Processing					
Seq.	**Name**	**Size [px]**	**Name**	**Size [px]**	ε^{-10}	ε^{-6}	ε^{0}	ε^{6}	ε^{10}	**%**
1	lena	256x256	merlin	128x128	0.54283	0.54341	0.54516	0.54465	0.54605	0.59
2	Lena	256x256	firefox	128x128	0.54657	0.54734	0.54914	0.54873	0.55023	0.53
3	Aquaria	256x256	firefox	128x128	0.54226	0.54432	0.54822	0.55144	0.55219	1.83
4	Aquaria	256x256	merlin	128x128	0.53958	0.54173	0.54277	0.54904	0.54979	1.89
5	Aquaria	256x256	watch	200x135	0.51725	0.51843	0.52248	0.52335	0.52403	1.31
6	Lena	256x256	watch	200x135	0.52042	0.52071	0.52238	0.52155	0.52341	0.57
7	Lena512	512x512	Aquaria	256x256	0.55789	0.55791	0.55757	0.55833	0.55864	0.13
8	fruit	512x512	lena	256x256	0.56109	0.54215	0.54268	0.56330	0.56588	4.58
9	fruit	512x512	Aquaria	256x256	0.56718	0.54826	0.56248	0.57653	0.57993	5.77
10	sphinx	800x600	fruit	512x512	0.50414	0.50415	0.52086	0.53481	0.53739	8.57
11	hawk	800x600	fruit	512x512	0.51281	0.51281	0.51001	0.51974	0.52670	2.70
12	Alicia	1024x1024	hawk	800x600	0.51574	0.53007	0.53995	0.54134	0.54134	4.96
13	Alicia	1024x1024	sphinx	800x600	0.51144	0.52137	0.52615	0.52680	0.52680	3.00
14	car	1024x1036	hawk	800x600	0.53980	0.53995	0.53938	0.56009	0.54019	0.07
15	car	1024x1036	sphinx	800x600	0.52675	0.52677	0.52362	0.52679	0.52683	0.02
16	fish	1600x1200	Alicia	1024x1024	0.54490	0.54790	0.55266	0.55661	0.55705	2.22
17	football	1600x1200	Alicia	1024x1024	0.54110	0.54497	0.55655	0.56314	0.56442	4.30
18	football	1600x1200	car	1024x1036	0.52558	0.52844	0.53536	0.53898	0.53951	2.65

Table 3

Cover object processing experimental results using the SCSA algorithm on 4 bits

	Cover Object		Secret Message		Cover Object Processing					
Seq.	**Name**	**Size [px]**	**Name**	**Size [px]**	ε^{-10}	ε^{-6}	ε^{0}	ε^{6}	ε^{10}	%
1	merlin	128x128	firefox	128x128	2.20065	2.21202	2.23028	2.21129	2.21925	0.85
2	Wildflowes	200x135	watch	200x135	2.11924	2.10974	2.10306	2.09935	2.16363	2.09
3	Lena	256x256	Aquaria	256x256	2.34364	2.37548	2.43198	2.35479	2.39349	2.12
4	Aquaria	256x256	lena	256x256	2.19141	2.23095	2.23256	2.21132	2.25572	2.93
5	Lena512	512x512	Fruit	512x512	2.22543	2.26987	2.27184	2.22689	2.27292	2.13
6	building	640x480	Dogs	640x480	2.36194	2.36677	2.47886	2.55992	2.60113	10.12
7	sphinx	800x600	Hawk	800x600	2.41660	2.44265	2.65467	2.72818	2.74720	13.68
8	hawk	800x600	Sphinx	800x600	2.34786	2.32531	2.28728	2.41975	2.48425	5.80
9	Alicia	1024x1024	Car	1024x1036	2.36585	2.39247	2.41650	2.53694	2.47603	4.65
10	car	1024x1036	Alicia	1024x1024	2.64180	2.75105	2.55188	2.64529	2.75354	4.22
11	fish	1600x1200	football	1600x1200	2.32144	2.32860	2.35385	2.36267	2.36583	1.91

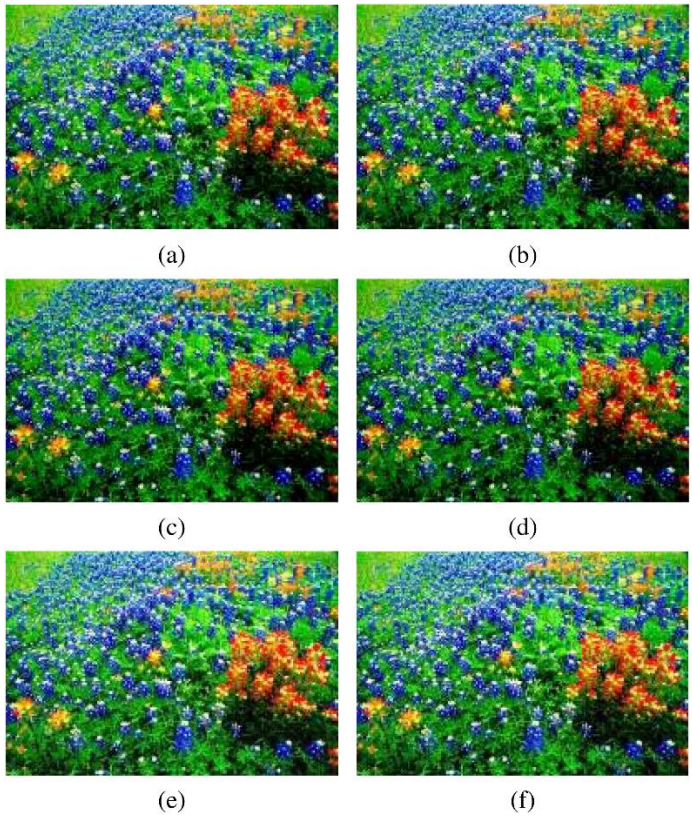


Figure 3

Exemplification of processing for ‘Wildowers’ cover object and `watch' secret message:
(a) Unprocessed cover object and (b) corresponding steganographic object; (c) Cover object with processing towards black and (d) corresponding steganographic object; (e) Cover object with processing towards white and (f) corresponding steganographic object

Conclusions

The increasing need for secure data communication methods, contributed to steganography gradually moving out of the research laboratory and into the real-world applications. To improve the security of steganographic objects, this paper has proposed the Cover Processing-based Steganographic Model (CPSM). CPSM adds a new layer of security to traditional steganographic models by processing the cover objects before embedding the messages. Moreover, to further complicate steganalysis the model makes use of random selection and embedding, where the sender transmits randomly either steganographic objects containing hidden information or processed cover objects aimed at confusing the attacker. A comprehensive demonstration based on information theory, proved that the CPSM model offers an improved security in terms of lower relative entropy as compared to the previous information-theoretic model proposed by Cachin. Experimental tests were conducted in order to further validate the benefits of the proposed model. The results showed that applying simple processing such as shifting the binary information of the cover image can lead to a decrease of the relative error between the cover and steganographic objects of up to 14%. Out future research work will aim to further improve the security of the proposed model by considering additional techniques such as processing the secret message along with the cover objects.

References

[1] Symantec, "Internet Security Threat Report 2017," Symantec Corporation, Mountain View, CA, 22, Apr. 2017 [Online] Available: https://www.symantec.com/security-center/threat-report

[2] L. Kessem, "2016 Cybercrime Reloaded: Our Predictions for the Year Ahead," Jan. 2016 [Online] Available: https://securityintelligence.com/2016-cybercrime-reloaded-our-predictions-for-the-year-ahead/

[3] A. Cheddad, J. Condell, K. Curran, and P. Mc Kevitt, "Digital image steganography: Survey and analysis of current methods," *Signal Processing*, Vol. 90, No. 3, pp. 727-752, Mar. 2010

[4] E. Zielińska, W. Mazurczyk, and K. Szczypiorski, "Trends in Steganography," *Communications of the ACM*, Vol. 57, No. 3, pp. 86-95, Mar. 2014

[5] F. Y. Shih, *Digital Watermarking and Steganography: Fundamentals and Techniques*, 2nd ed. CRC Press, 2017

[6] A. D. Ker, P. Bas, R. Böhme, R. Cogranne, S. Craver, T. Filler, J. Fridrich, and T. Pevný, "Moving Steganography and Steganalysis from the Laboratory into the Real World," in *Proceedings of the First ACM Workshop on Information Hiding and Multimedia Security*, New York, NY, USA, 2013, pp. 45-58

[7] D. Caragata and I. Tutanescu, "On the security of a new image encryption scheme based on a chaotic function," *Signal, Image and Video Processing*, Vol. 8, No. 4, pp. 641-646, 2014

[8] A. Wilson, P. Blunsom, and A. D. Ker, "Linguistic steganography on Twitter: hierarchical language modeling with manual interaction," in *Proc. SPIE 9028, Media Watermarking, Security, and Forensics 2014*, 2014, p. 902803

[9] D. Stanescu, M. Stratulat, V. Groza, J. Ghergulescu, and D. Borca, "Steganography in YUV color space," in *2007 International Workshop on Robotic and Sensors Environments*, 2007, pp. 1-4

[10] Z. Shahid, M. Chaumont, and W. Puech, "Considering the reconstruction loop for data hiding of intra- and inter-frames of H.264/AVC," *SIViP*, Vol. 7, No. 1, pp. 75-93, 2013

[11] B. J. Mohd, S. Abed, B. Na'ami, and T. Hayajneh, "Hierarchical steganography using novel optimum quantization technique," *SIViP*, Vol. 7, No. 6, pp. 1029-1040, 2013

[12] J. Lubacz, W. Mazurczyk, and K. Szczypiorski, "Principles and overview of network steganography," *IEEE Communications Magazine*, Vol. 52, No. 5, pp. 225-229, May 2014

[13] D. Stanescu, V. Groza, M. Stratulat, D. Borca, and I. Ghergulescu, "Robust Watermarking with High Bit Rate," in *Third International Conference on Internet and Web Applications and Services, 2008. ICIW '08*, 2008, pp. 257-260

[14] B. Li, J. He, J. Huang, and Y. Q. Shi, "A Survey on Image Steganography and Steganalysis," *Journal of Information Hiding and Multimedia Signal Processing*, Vol. 2, No. 2, pp. 142-172, 2011

[15] C. Paulin, S. A. Selouani, and É. Hervet, "A comparative study of audio/speech steganalysis techniques," in *2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE)* 2017, pp. 1-4

[16] J. Babu, S. Rangu, and P. Manogna, "A Survey on Different Feature Extraction and Classification Techniques Used in Image Steganalysis," *Journal of Information Security*, Vol. 08, No. 03, pp. 186-202, Jul. 2017

[17] R. Amirtharaj, J. Qin, and J. B. Balaguru R, "Random Image Steganography and Steganalysis: Present Status and Future Directions," *Information Technology Journal*, Vol. 11, No. 5, pp. 566-576, May 2012

[18] N. Hamid, A. Yahya, R. B. Ahmad, and O. M. Al-Qershi, "Image steganography techniques: an overview," *International Journal of Computer Science and Security (IJCSS)*, Vol. 6, No. 3, pp. 168-187, 2012

[19] R. J. Mstafa and K. M. Elleithy, "Compressed and raw video steganography techniques: a comprehensive survey and analysis," *Multimedia Tools and Applications*, Vol. 76, No. 20, pp. 21749-21786, Oct. 2017

[20] M. Douglas, K. Bailey, M. Leeney, and K. Curran, "An overview of steganography techniques applied to the protection of biometric data," *Multimedia Tools and Applications*, Vol. 77, No. 13, pp. 17333-17373, Jul. 2018

[21] M. Hussain, A. W. A. Wahab, Y. I. B. Idris, A. T. S. Ho, and K.-H. Jung, "Image steganography in spatial domain: A survey," *Signal Processing: Image Communication*, Vol. 65, pp. 46-66, Jul. 2018

[22] I. J. Kadhim, P. Premaratne, P. J. Vial, and B. Halloran, "Comprehensive survey of image steganography: Techniques, Evaluations, and trends in future research," *Neurocomputing*, Nov. 2018

[23] G. J. Simmons, "The Prisoners' Problem and the Subliminal Channel," in *Advances in Cryptology*, D. Chaum, Ed. Springer US, 1984, pp. 51-67

[24] J. Zöllner, H. Federrath, H. Klimant, A. Pfitzmann, R. Piotraschke, A. Westfeld, G. Wicke, and G. Wolf, "Modeling the Security of Steganographic Systems," in *Information Hiding*, Springer Berlin Heidelberg, 1998, pp. 344-354

[25] C. Cachin, "An information-theoretic model for steganography," *Information and Computation*, Vol. 192, No. 1, pp. 41-56, Jul. 2004

[26] P. Sallee, "Model-Based Steganography," in *Digital Watermarking*, T. Kalker, I. Cox, and Y. M. Ro, Eds. Springer Berlin Heidelberg, 2003, pp. 154-167

[27] A. J. Raphael and V. Sundaram, "Cryptography and Steganography-A Survey," *International Journal of Computer Technology and Applications*, Vol. 02, No. 03, pp. 626-630, May 2011

[28] A. J. Raphael and V. Sundaram, "Secured Crypto-Stegano Communication through Unicode," *World of Computer Science and Information Technology Journal*, Vol. 1, No. 4, pp. 138-143, 2011

[29] P. Schöttle and R. Böhme, "Game Theory and Adaptive Steganography," *IEEE Transactions on Information Forensics and Security*, Vol. 11, No. 4, pp. 760-773, Apr. 2016

[30] M. Fakhredanesh, R. Safabakhsh, and M. Rahmati, "A Model-Based Image Steganography Method Using Watson's Visual Model," *ETRI Journal*, Vol. 36, No. 3, pp. 479-489, Jun. 2014

[31] H.-T. Song, G.-M. Tang, G. Kou, Y.-F. Sun, and M.-M. Jiang, "Digital steganography model and embedding optimization strategy," *Multimed Tools Appl*, Nov. 2018

[32] T. Denemark and J. Fridrich, "Model based steganography with precover," in *IS&T International Symposium on Electronic Imaging 2017, Media Watermarking, Security, and Forensics 2017*, 2017, pp. 56-66

[33] Z. K. AL-Ani, A. A. Zaidan, B. B. Zaidan, and H. O. Alanazi, "Overview: Main Fundamentals for Steganography," *Journal of Computing*, Vol. 2, No. 3, 2010

[34] S. Katzenbeisser and F. A. P. Petitcolas, *Information Hiding Techniques for Steganography and Digital Watermarking*. Artech House, 2000

[35] D. Dumitrescu, I.-M. Stan, and E. Simion, "Steganography techniques," 341, 2017 [Online] Available: https://eprint.iacr.org/2017/341

[36] J. Devore, *Probability and Statistics for Engineering and the Sciences*, 8th ed. Cengage Learning, 2012

[37] D. Stanescu, I.-G. Bucur, and M. Stratulat, "Segment Compression Steganographic Algorithm," in *2010 International Joint Conference on Computational Cybernetics and Technical Informatics (ICCC-CONTI)* 2010, pp. 349-354