AARMS Volume 14, Issue 2 (2015)

Contents

ISSN 2064-0021

Volume 14, Issue 2 (2015) 133–264

AARMS

AARMS

Academic and Applied Research in Military and Public Management Science

Volume 14
Issue 2
2015

ZELE Balázs:
Distribution of Fire Cases and the Role of Human Factors in Coal-Firing Power Plants in Fuel-Supply Fields and Distribution Systems

BARÁTH Artur:
Selection of Authentication Systems for Hungarian Health Care, Based on Physiological Study

JUHÁSZ Attila, NEUBERGER Hajnalka:
Detecting Military Historical Objects by LiDAR Data

KRAJNC Zoltán:
The Near and More Distant Future Environment of Air Defence Missions

SCIENTIFIC JOURNAL OF THE NATIONAL UNIVERSITY OF PUBLIC SERVICE, HUNGARY

AARMS is a peer-reviewed international scientific journal devoted to reporting original research articles and comprehensive reviews within its scope that encompasses the military, political, economic, environmental and social dimensions of security and public management.

AARMS is published in one volume of four issues per year by the National University of Public Service, Budapest, Hungary, under the auspices of the Rector of the University: Prof. Dr. PATYI András.

All correspondence should be addressed to
Prof. Dr. KOVÁCS László, Managing Editor
National University of Public Service
P. O. Box 15, H-1581 Budapest 146
Hungary

Phone: +36-1-432-9062
Fax: +36-1-432-9040
aarms@uni-nke.hu
www.uni-nke.hu/aarms/index.htm

AARMS Vol. 14, No. 2 (2015) 133–147. **10.32565/aarms.2015.2.1**

The Application and Tasks of the Anti-tank Artillery in Combined Arms Combat

FURJÁN Attila[1]

Anti-tank is defined as combat against the enemy's tanks and other armoured equipment and vehicles. Its associated tasks include the closing of endangered directions, flanks, and junctions. Anti-tank is the part of combined arms combat which is directed against the tanks grouped for attack or already directly attacking, and the armoured equipment detached to their battle formation. For successful combat, the organization of a solid anti-armour system is required, which includes anti-tank units as a basic element.

Keywords: *anti-tank artillery, anti-armour, fire capacity, tactical assembly area, deployment area, manoeuvre, battle formation*

Introduction

"Fire support is the collective and coordinated use of target intelligence, indirect laying weapons, armed aircraft, other lethal and nonlethal means and support modes of the battle plan." [1: 33] This definition shows that, according to NATO principles, direct laying anti-tank artillery weapons are not part of fire support. In the armed forces of older NATO member states, anti-tank units belong directly to manoeuvre forces, and their operation is planned and tasks are assigned by the commander and staff of the manoeuvre force.

The activities of anti-tank units forms a part of the anti-armour system; therefore, they do not operate independently but always cooperate with other combined arms units. One of the most important tasks of the Combined Arms Commanders is to organize the anti-armour system, including the planning of the combat operations of anti-tank units.

In the Hungarian Defence Forces, anti-tank units traditionally belong to the artillery. Here some contradiction appears, as direct laying weapons – according to the NATO concept – do not belong to fire support. How shall we interpret that, and how shall we plan and use anti-tank artillery units in practice? After a consultation at the NATO Artillery Working Group session, we were reassured, that in the context of national planning and application, anti-tank artillery forces can be regarded as part of fire support. In case of an operation planned and executed within a NATO formation, the anti-tank artillery forces will be regarded as manoeuvre forces and their application will be planned by the Joint Force Commander and his staff as an anti-tank manoeuvre group. This is a peculiarity we should not be afraid of, as there are other arms peculiarities and differences in several NATO member countries.

1 National University of Public Service, Budapest, Hungary; e-mail: furjan.attila@uni-nke.hu

Characteristics, Combat Capabilities of Anti-tank Artillery

Characteristics of Anti-tank Artillery Range

The range of guided armour-piercing missiles is determined by their maximum distance of launch and the distance of the line of fire – influenced by the terrain in front of their firing positions. Ideally, the range will be the same as the maximum distance of launch.

Example: The maximum distance of launch of the KONKURSZ guided armour-piercing missile combat machine is 4 km.

There is a hill (or a tall forest, row of trees, etc.) in front of the firing position at a distance of 3 km. In that case, the range of the KONKURSZ combat machine will be 3 km.

Fire Capacity

The fire capacity of the anti-armour artillery is determined by the amount, types and combat effectiveness of its anti-tank assets.

The evaluation of the combat effectiveness of the anti-tank assets is based on the anti-tank equipment type's capability of how many attacking, defending, mobile, fixed, armoured targets – located in a shelter or open area – they are expected to destroy before they would be destroyed, in specified circumstances (attacking, defending, from under shelter or from an open position, etc.).

The combat effectiveness of an anti-tank asset against a given type of armoured target is determined by the following factors, taking into account the mutual counter-actions:

- range (maximum distance of launch, blank range);
- hit probability at a target of a given size (tank, anti-tank asset);
- effectiveness of the projectile (armour penetrating power, destruction zone);
- the probability of destroying the target in case of a hit;
- firing rate;
- protection (in shelter, open area, armoured or non-armoured);
- the type of combat operation (attacking, defending);
- situation (fixed, moving), manoeuvre capability;
- concealment or exposed to reconnoitring;
- preceding in opening fire;
- other factors (e.g. operators' skills, weather-, visibility-, terrain circumstances, etc.). As

the potential to destroy different types of armoured targets is different case by case, artillery professionals have defined a unit of armoured targets, a so-called *Tank Unit (TU)*, for the determination of combat effectiveness.

The TU is a tank having pre-defined characteristics and combat capabilities, the vulnerability of which is used as a baseline to determine the vulnerability of tanks, armoured vehicles, etc. having different combat capabilities.

In the estimations, the tanks that were standardized before 1980 (T–55, T–62, T–72, Leopard–1, etc.) are regarded as equivalent to TU; tanks produced after 1980 (T–80, Leopard–2, Abrams M–1, etc.) are worth 2 TUs; while armoured carriers, infantry combat vehicles, guided armour-piercing missile machines are regarded as 0.7×TU.

The indicator of the combat effectiveness of anti-tank assets is the combat-effectiveness factor.

The combat-effectiveness factor (Table 1.) is the measure of how many TUs of armoured targets the anti-tank asset is expected to destroy before it would be destroyed. The values of the combat effectiveness factors differ depending on the type of combat operation, situation of the anti-tank assets and the enemy's armoured equipment, and how these are protected.

Table 1. Combat-effectiveness factor of anti-tank assets. [1: 24]

Anti-tank assets	Values of tactical calculations (in a concrete case)					Medium values of tactical calculations	
	Defending		Attacking		Meeting engagement Open	Defending	Attacking
	Own anti-tank asset, situation		Enemy tanks situation			Own anti-tank assets situation Covered 2/3 Open/1/3	Enemy tanks situation Covered 2/3 Open/1/3
	Covered	Open	Covered	Open	Open		
SPG–9	1.5	1.2	0.7	1.0	1.2	1.4	0.8
FAGOT, METISZ	2.0	1.5	1.0	1.3	1.5	1.8	1.1
MALJUTKA	2.5	2.0	1.0	1.5	2.0	2.3	1.2
KONKURSZ	2.8	2.3	1.2	1.7	2.3	2.6	1.4
100 mm anti-tank gun	2.0	1.5	1.0	1.3	1.5	1.8	1.1
RPG–7	0.3	0.2	–	0.2	0.2	0.25	0.1
BMP (Guided AP M.)	2.0	1.5	1.0	1.3	1.5	1.8	1.1
T–55, T–62	2.0	1.5	1.0	1.3	1.3	1.8	1.1
T–64, T–72	2.8	2.0	1.5	1.6	1.7	2.5	1.5

The fire capacity of the antitank *artillery* (anti-armour capability) can be expressed as the amount of those armoured targets which a given anti-tank group – under normal conditions – is expected to be able to destroy during the battle, or the assault of which the group is able to reject in a specific combat situation.

The fire capacity of anti-tank units can be expressed with the following formula:

$FCTU = \sum^{i}n = 1\ ni \times CEFi$ [1: 25]

where:

FCTU – the fire capacity of the anti-tank unit in TUs; ni – is the number of (i.) type anti-tank assets; CEFi – the combat effectiveness factor of the (i.) type anti-tank asset.

Example: The fire capacity of a KONKURSZ guided armour-piercing missile battery on the defence, from a prepared deployment area (from shelter).

FCTU = 8 pcs. KONKURSZ Guided AP Missiles×2.8 = 22.4 TU

The result of this example *on the one hand* means that a KONKURSZ battery is – theoretically, until the last KONKURSZ combat machine is destroyed – able to destroy 22 TUs of armoured targets from shelter, in case of defence operations, *on the other hand* that the KONKURSZ battery is able to repulse the assault of an armoured grouping equivalent to 22 TUs with high (90%) probability and cause a minimum 50% loss to that group in a specific situation, at a time.

Manoeuvre Capability

The manoeuvre capability of guided anti-tank missile units is determined by their mobility (movements to firing positions, during manoeuvres to deployment areas), and the time required for the occupation and leaving of deployment areas.

Basic Issues of Anti-Armour Operations

The attacking party will not distribute their tanks evenly, so it is recommended to group anti-tank units in the main tank-threatened directions. The anti-armour system shall be built up in such a way as to establish coherent and depth-echeloned fire zones in the main tank-threatened directions. Thereby, the continuous destruction of attacking armoured assets will be ensured both in front of the foremost line and in the full depth of defence.

The basis of the anti-armour system is a well-organized anti-tank fire system. To ensure maximum efficiency, a larger part of the anti-tank assets shall be positioned in the vicinity of the front edge so that they inflict as high losses on tanks as possible. In positioning the assets, the commanders shall focus on maintaining continuous fire on tank- and infantry obstacles, as well as ensuring connection of fires between units and the destruction of armoured assets breaking in into the depth.

The Functions and the Basics of Combat Application of Anti-tank Units

The functions of anti-tank *units:* to destroy the enemy's tanks, fire weapons, and other armoured assets and combat vehicles, as well as protective installations and the personnel located in those.

Anti-tank units are capable of carrying out fast manoeuvres in endangered directions and deployment areas, with the aim to destroy tanks with medium armour protection and any other types of armoured combat vehicles located there. The fire capacity of anti-tank units depends on the degree of armour protection of the targets, the amount of the involved assets, the skills of operator personnel, the organization of the fire system, the effects of missiles (projectiles) on targets, and the circumstances of the mission.

Let us examine the functions of anti-tank artillery units in the light of the draft anti-armour doctrine: "the Anti-tank Missile Battalion has been integrated into the establishment of the Infantry Brigade for the function of destroying the enemy's tanks, other armoured combat vehicles, and protective installations; as well as close endangered directions, open flanks and

junctions. Its further tasks include the destruction of armoured assets breaking into the defence and preventing them from gaining ground and providing support on the flanks for own forces' counter-attack. Furthermore, it is applied to destroy low-flying and hovering helicopters. The anti-tank missile battalion is the basic anti-armour fire control and tactical unit of the Infantry Brigade. It is capable of carrying out fast manoeuvres in endangered directions and deployment areas, and performing its basic functions there." [2: 1]

The establishment of the battalion have recently been under continuous change; and we also expect modifications to continue, considering the force development aims of all organizational elements of the Hungarian Defence Forces.

In accordance with operational requirements, the establishment of the Anti-tank Missile Battalion includes headquarters and staff – these composing the leadership – and three anti-tank missile batteries and a staff support battery. An anti-tank missile battery is made up of a battery staff and two platoons. Each platoon is equipped by 4–4 pcs. of 9P148 combat vehicles.

Picture 1. 9 P148 combat machine. [3]

The anti-tank battery is the fire control and tactical unit of the anti-armour artillery. It carries out its tasks either independently or under the command of a higher-level organization.

The artillery platoon is a unit of the anti-armour artillery, which usually carries out its tasks under the command of a Battery, sometimes independently.

The guided armour-piercing missile combat machine and launcher are the fire weapons of the anti-tank unit, which are usually used in combat under the command of a platoon. The personnel directly ensuring the operation of the combat machines and launchers are called the squad or the operating personnel.

The Battle Formation of Anti-tank Units

The battle formation is the purposeful grouping of forces and assets for the execution of the assigned mission. The battle formation shall ensure the execution of the units' tasks, the full utilization of the units' combat potential, continuous and reliable cooperation and

connectivity with the superior and other units, the possibility to manoeuvre the unit and the fires, maximum utilization of the opportunities provided by the terrain, and the continuous, uninterrupted command of subordinates.

The battle formation of the anti-tank battery is composed of the battle formations of artillery platoons, the Command (Observation) Post of the Battery Commander, the Armour Observer Post, and the Loading Point.

Upon order by the Brigade Commander, the anti-tank battery can reinforce the anti-tank activities of an infantry battalion, or can operate as an anti-tank manoeuvre unit (anti-tank reserve) under the command of a battery, in case there are more than one tank-threatened directions in the combat area of the brigade.

The anti-tank squads of the support platoons of the infantry company are situated in the battle formation of the company in such a way as defined by the Company Commander.

The Areas and Lines of Anti-tank Units

Assembly areas, one or two tank-threatened directions, and 2–3 deployment areas per direction are designated for anti-tank units.

The assembly area is the part of the terrain which the anti-tank unit holds or has prepared for occupying. Units are situated in an extended order in the assembly area, in accordance with the expected engagement. "A Battery's assembly area shall be 500 m wide and 500 m long." [2: 4–12]

The assembly area shall be designated in the direction of the enemy's expected principal effort (or in the direction of the main attack of own forces) in relation to the deployment area, in such a way that available manoeuvre routes ensure the deployment also in the furthermost areas. The battery must maintain a 15 minute starting notice. On the defence, 1–2 reserve assembly areas can be designated and prepared at a distance of 1–2 km from the assembly area to the side and back. During the attack, the unit takes temporary assembly areas, keeping with the rhythm of the attack. [2: 4–12] The assembly areas shall be named after flowers.

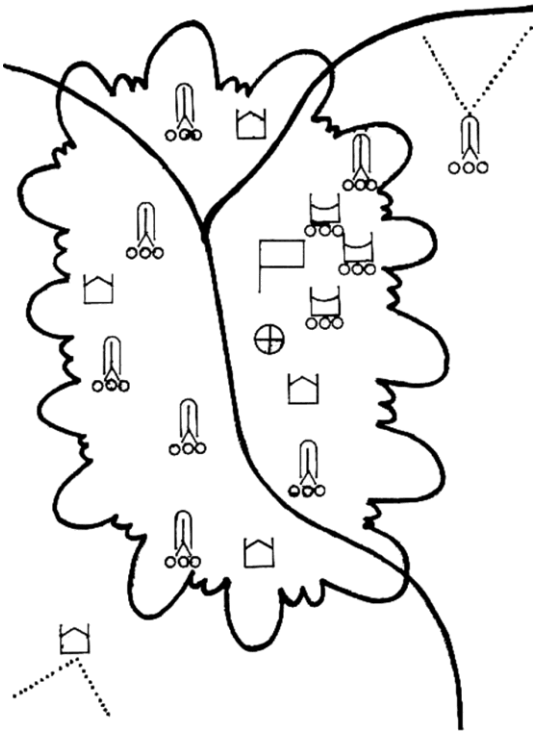


Figure 1. Guided anti-tank missile battery at the assembly area. [2: 4–13]

The anti-tank squads of the support platoons of the infantry company are situated in the readiness area (sheltered area) of the company in such way as defined by the company commander.

The *deployment area* is the part of the terrain in the tank-threatened direction which is pre-assigned or prepared to be occupied, or which the unit has taken during combat operations in order to destroy the enemy's tanks and armoured combat vehicles. The deployment areas shall be named after mountains and with numbers.

E.g.: "Bakony–2"

This area is the 2. deployment area in the "Bakony" tank-threatened direction.

During the deployment of units into the battle formation on the deployment area, their disposition shall ensure the successful execution of the tasks, mutual connection of fires, transferring of fires from one direction to the other, and all-round defence.

Types of deployment areas:

- planned and prepared deployment area;
- planned, unprepared deployment area;
- unplanned deployment area.

The unit usually takes a *planned, prepared deployment area* when it is on the defence, and has enough time to inspect the deployment area; thoroughly organize the fire system; practice occupation and leaving of the area, and the fire tasks; and perform the engineering works. In case of a planned, unprepared deployment area, they carry out the inspection of the deploy-

ment area and the organization of the fire system, but do not have enough time to practice occupation and leaving of the area, and the fire tasks; and perform the engineering works.

The *unplanned deployment areas* are occupied during combat, when quick changes in the situation compel the commander to use the anti-tank unit quickly, at any place. The occupation of unplanned deployment areas require the unit commander to exercise a high degree of self-reliance and creativity. In this case, the squad commanders' skills and abilities for the tasks are decisive.

"In order to ensure the effectiveness of fire control and fire coordination, the distance in length is 100–200 m between Guided Armour-piercing Missile combat machines, and 300–400 m between the platoons. The dimensions of deployment areas can extend to 2.5 km in width and 1 km in length for batteries, 1 km in width and 500 m in length for platoons; depending on the combat situation, terrain, and the number of assets involved." [4: 6]

The usual dispositions of units on the deployment area: wedge front (back), echelon right (left), line, or horseshoe formation. These formations usually do not take place clearly by themselves, they are usually applied in combinations, depending on the number of the involved anti-tank assets, the width of the deployment area, and the type of terrain. In today's modern combat, it may happen that only 1–2 assets are actually tasked. The anti-tank squads of the support platoons of the infantry companies operate as ordered by the Company Commander, they have platoon- and squad-level tasks.

Wedge formation:

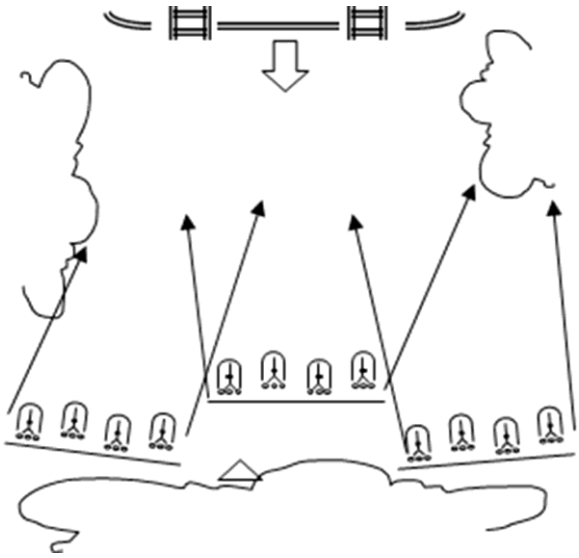


Figure 2. Wedge formation of an anti-tank missile battery (a variant). [4: 6]

This battle formation is recommended if a narrow area must be closed, or if the enemy's battle formation is deeply echeloned. It is advantageous due to the relatively great depth, ensuring the utilization of the maximum fire distance. The disadvantage is that a relatively narrow area can be closed and the units cannot begin firing at the same time.

Echelon formation:

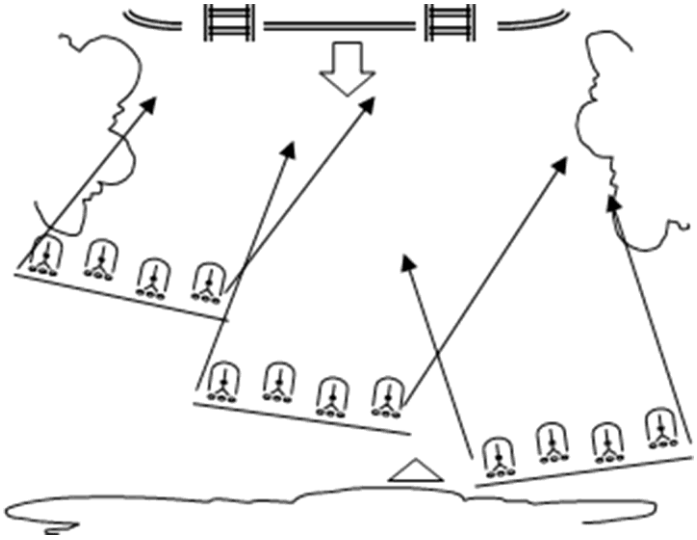


Figure 3. Echelon right formation of an anti-tank missile battery (a variant). [4: 7]

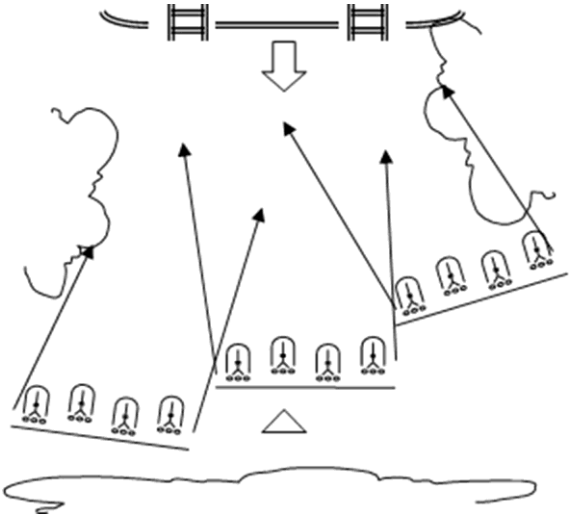


Figure 4. Echelon left formation of an anti-tank missile battery (a variant). [4: 7]

This battle formation is recommended on hilly, mountainous terrain, and on a hill which diagonally crosses the deployment area, because the distance of the hill allows a uniform distance of line of firing for all units.

Its advantage is that opening fire is possible from a large distance, and slant firing can also be conducted. Its disadvantage is that fire can be opened only when the targets have reached the effective range and that is why opening fire simultaneously is not possible, and why the units can close only narrow areas.

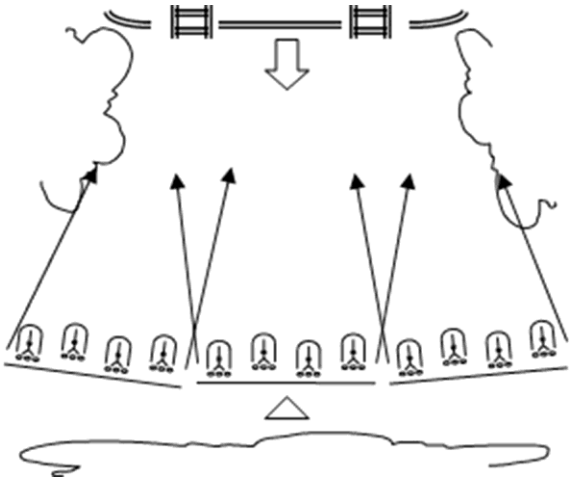


Figure 5. Line formation of an anti-tank missile battery (a variant). [4: 7]

Horseshoe formation:

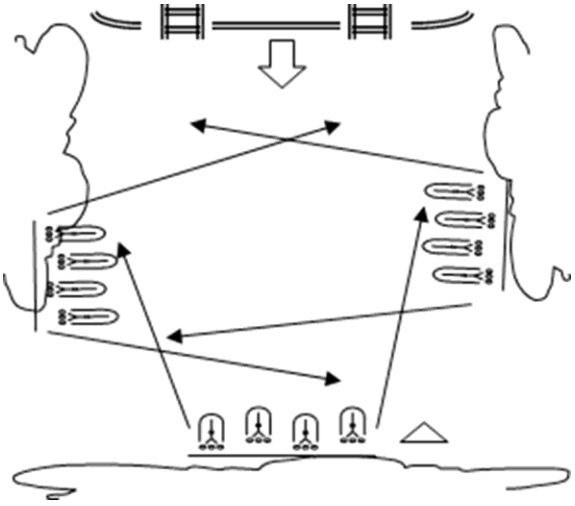


Figure 6. Horseshoe formation of a guided anti-tank missile battery (a variant). [4: 7]

This is the most frequently used battle formation. It provides the possibility for units to deploy to a maximum width and length. It ensures firing at the largest distance, allows the elimination of dead spaces for the weapons, and creates good opportunities for delivering unexpected and high-impact fire.

This battle formation is usually applied for the closing of narrow passages, valleys. The horseshoe formation is in fact an anti-tank pocket of fire. Its advantage is that it allows opening fire simultaneously, and enables fire weapons to fire at the most vulnerable points of tanks. Its disadvantage is the relatively narrow width; and that missed missiles and those which become uncontrollable can cause damage to own forces.

The Armour Observer Post must be provided with a radar which can reliably detect advancing, deploying tank formations both during day and night, and in fog and under poor visibility conditions. Today, modern defence forces use radars with various performance abilities for the reliable detection of enemy tanks.

The RATAC–S mobile, multi-purpose ground target detection and acquisition radar station is developed and produced by the German ALCATEL SEL AG. The RATAC–S is a coherent-pulse, twin bundle, monopulse Doppler radar, which filters out stationary targets. Its operation is facilitated by a user-friendly menu system, familiar to anyone who uses computers. Its most important functions are:

- radar map can be depicted (mapping indicator mode);
- the radar picture can be stored and any sectors can be zoomed;
- operators can be alarmed in optical and acoustic ways;
- the definition of target coordinates can be refined by freezing the screen;
- GPS can be easily adapted;
- the detection and tracking of the target can be easily documented by the plotter;
- it can independently operate with its own battery unit.

Picture 2. The RATAC reconnaissance radar station, used for the detection and display of enemy tanks and armoured combat vehicles. [5: 75]

The operator can set the reconnaissance, target tracking, and fire control area from 0° to 360° and 0 km to 38 km in multiple grades, depending on the task. Sector search can also be set. In that case, the antenna system automatically searches for moving targets in a sector of maximum 110° width and 20 km length. Thanks to the sophisticated Doppler technology, the radar can determine even the type of target. Thereby, a soldier, a wheeled or tracked combat vehicle, a train, or rotary wing aircraft can be differentiated.

The RATAC–S is able to maintain detection even in such natural and artificial weather- and atmospheric conditions which restrict visual reconnaissance (e.g. fog, heavy rain, snow, night, smoke, dust).

A controllable, automatic system ensures the filtering of natural background interference – caused by the movement of plants blown by the wind, the waves of the water surface, and changes in the refractive index of the atmosphere. The control unit includes a foldable control panel providing simple, quick and efficient use, a multi-function display unit, a control computer, and a processor for digital signal processing. The control unit is also equipped with a speaker, which emits a sound corresponding to the speed and type of the target being tracked. The control unit can be augmented with further accessories: navigation-, plotter- and video camera device, a headphone instead of the speaker, and elements of connection and the information system.

RATAC–S characteristics: its operating frequency is 9.5 GHz, its output of the transmission pulse is 7 kW, its total weight is 125 kg, and its operating temperature range is 32°C to 55°C.

The detection capability of the radar in case of a 90° detection sector: [5: 76]

- moving people: from 18 km;
- light, wheeled vehicles: from 24 km;
- heavy tracked vehicles: from 30 km;
- helicopters: from 28 km;
- march column: from 38 km.

In case of artillery fire control, the points of burst can be calculated by graphic intersection with an accuracy of ±10m, which ensures appropriate support for fire control.

The Command of Anti-tank Units

The command of anti-tank units is the Commanders' activity through which they maintain the combat capabilities of the units, prepare their combat operations, and control the units during the execution of the assigned tasks. The most important element of command is fire control, which covers the reconnaissance of targets, the clarification of firing tasks, taking decisions regarding the firing tasks, assigning firing tasks, the fire manoeuvring activities of units, and the supervision thereof.

The quick and harsh changes of the situation in combined arms combat require the commanders of anti-tank units to exercise continuous, solid, operational and hidden command.

Continuity, permanent influence on the activities of the battery and platoons, which becomes decisive in the critical phases of combat. Its preconditions shall be established during the organization of command. Its main basic factor is the maintaining of uninterrupted communication, on multiple channels. The handover-takeover and if necessary, the immediate takeover of command functions shall be prepared, in case the commander is eliminated.

Solidness, the pursuit to take decisions in due time and to carry out the defined tasks firmly and tenaciously. To this end, the commander must be familiar with the superior's concept, places and roles of his units in the superior's battle formation and tactics. The commander shall continuously impose high requirements on Subordinates to ensure solidness.

Operationality, the collective enforcement of efficiency, speed, and practicality in the activities of the unit. It is realized through quick reaction to situations and the introduction of measures in a timely manner.

Stealthiness, by keeping all measures taken for the preparation and execution of tasks secret, using encrypted coordinates; and strict compliance with and enforcement of information security regulations.

Unit commanders command their units through oral orders, combat orders, and commands. Squad leaders command their squads by commands and signals. Orders, instructions, commands shall be short and clear.

The unit commander usually organizes the combat in the field. Due to lack of time, tasks must sometimes be clarified during or immediately after establishing the battle formation and its elements. This is the situation in every such case when an unplanned deployment area or a temporary assembly area is to be occupied.

The commander's schedule for organizing the combat depends on the concrete situation, on the assigned task, and on the time available. The preparation of the combat activity starts when the superior's combat order (preliminary combat order) is received, and lasts until the beginning of advancing to battle formation.

Depending on the superior's working method, the commander's working method can be *gradual* or *parallel*.

Gradual working method is usually applied at the beginning of organizational work. It is primarily characterized by the sequential work of different command levels, and that tasks are built upon each other.

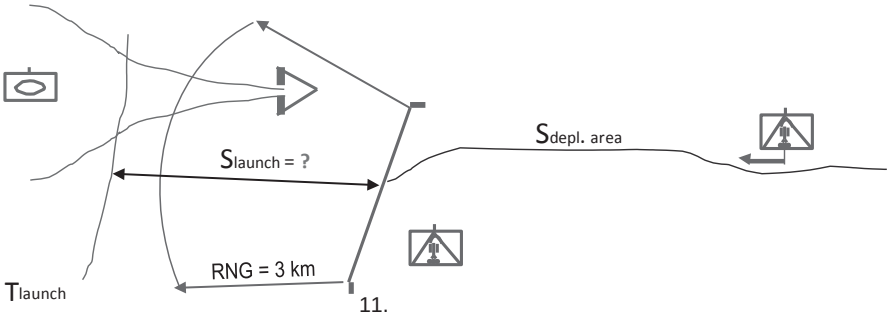
In case of the *parallel* method, the command functions are performed upon receiving the superior's order (command), nearly at the same time. The command levels work on the task simultaneously. This method is usually used when the commander has a limited time to organize the activities.

The unit gets prepared for movement in the assembly area on a 10 minute starting notice, positioned in an order corresponding to the battle formation. The commanders prepare their combat control documents related to the given area. Upon order by the superior, the unit starts the march and takes battle formation positions, in accordance with practice training. If there has been no practice training, the squad leader, using terrain coverage, disembarks in the vicinity of the firing position (about 50 m distance) and controls the combat vehicle into the firing position.

Having taken the firing position, they make combat machines ready for fire, and report to the platoon leader. When the platoon has reached fire readiness, the platoon leader reports that to the battery commander. The battery commander can report about fire readiness to the superior when the battery has reached at least 50% readiness. After having reached fire readiness, radio silence takes effect. Squad leaders continuously watch their fire sectors until the beginning of fire, and refine their sector sketches if necessary.

Planning of the manoeuvre of the Guided Anti-tank Battery

Example: (own source)



$$t_{readyn.} = \frac{S_{depl.\ area}}{V_{mnvr.}} \times 60 + t_{occup.} \qquad S_{launch} = \frac{t_{readyn.}}{60} \times V_{tank} + RNG$$

Figure 7. Calculation of a manoeuvre of an anti-tank unit. [6: 57]

Data:
$S_{depl.\ area}$ = 3 km
$V_{mnvr.}$ = 27 *km*/h
$t_{occup.}$ = 4 min
RNG = 3 km
V_{tank} = 10 km/h

$$t_{readyn.} = \frac{3\ km}{27\ km/h} \times 60 + 4\ min. = 10.67 \sim \underline{11\ min}$$

$$S_{launch} = \frac{11\ min}{60} \times 10\ km/h + 3\ km = \underline{4.83\ km}$$

The readiness of the anti-tank reserve (anti-tank manoeuvre group) on the designated deployment area means that the units have taken their battle formation, the Commander's Observation Post has been established, the fire system has been organized, that is the anti-tank reserve is able to begin the destruction of the enemy's armoured assets at the maximum distance of opening fire.

The anti-tank reserve of the brigade usually executes its tasks in cooperation with the *mobile closing detachment of the brigade*. The area of the mobile closing detachment is situated near the assembly area of the anti-tank reserve, possibly in front of that (in the direction of the deployment areas).

The mobile closing detachment (three mine-layer platoons with 3 pcs. of PMZ–4 mine- layers per platoon) is able to lay a minefield of 2.4–3.3 km width (one platoon: 0.8–1.0 km) and 30–60 m length, with one standard load of mines.

The mobile closing detachment lays the minefield – either continuously or fractionally, depending on the terrain and the battle formation – in the front of the area of the anti-tank reserve, at a distance of 0.5–1.0 of the blank range of anti-tank guns, or 0.5 of the maximum distance of launch of the guided armour-piercing missiles. The mobile closing detachment has to begin the manoeuvre in due time to have 5–10 minutes time advantage over the an- ti-tank reserve on the concrete area.

Conclusions

The application of anti-tank units shall be planned for the sake of achieving combined arms objectives. As the main striking power of land forces consists in tank troops, the destruction of tank- and other armoured targets is of key importance for the success of combined arms combat. Force superiority applied at the right time and place can settle the outcome of the battle.

Anti-tank units are capable of carrying out fast manoeuvres in the endangered directions and deployment areas, and successfully performing their basic functions there.

References

[1] FURJÁN A.: *The Basics of Fire Support, as well as Combat Use and Command of the Artillery*. Budapest: National University of Public Service, 2009.

[2] *The Principles of Combat Application of the Anti-tank Missile Batteries, Draft Doctrine.* Székesfehérvár: Hungarian Defence Forces Land Force Command, 2005. [3] www.1999.co.jp/eng/10278975 (downloaded: 14 01 2015)

[4] ŐSZ Cs.: *The Tasks of Anti-Tank Unit Commanders during the Planning and Commanding of Combat Operations.* Székesfehérvár: Hungarian Defence Forces Land Force Command, 2005.

[5] FURJÁN A: *Theoretical and Practical Issues of Processing Artillery Intelligence and Intelligence Data in an Integrated Intelligence System in the Hungarian Defence Forces.* Budapest: National University of Public Service, 1999. (Ph.D. dissertation)

[6] FURJÁN A.: *The Military Use of the Artillery Subunits in Combat Activities of the Infantry Battalion (Company), as well as in Crisis Management and in Peace Support Operations.* Budapest: National University of Public Service, 2015.

AARMS Vol. 14, No. 2 (2015) 149–159. **10.32565/aarms.2015.2.2**

Long-Term Storage of Digitally Signed Documents

GYURÁK Gábor[1]

Digital documents play an increasingly important role in our lives. Reliable digital storage of these documents is complicated and we have to deal with other problems if we would like to store these documents for a long time. Some documents, especially the most important documents are electronically signed. The long-term storage of electronically signed documents is more difficult, because we have to ensure the long-term validity as well. Electronic invoices (e-invoices) are also electronically signed documents and their role is becoming more important. The proposal of the European Committee on e-invoicing aims to facilitate the use of e-invoices. By 2020 e-invoicing will be general usage in the EU. This paper describes the problems in connection with long-term storage of digitally signed documents. Possible solutions are also presented. In connection to this, the regulation of preserving electronically signed documents is also examined from the point of view of Hungarian legislation. Finally it is shown how ETSI's (European Telecommunication Standards Institute) PAdES (PDF Advanced Electronic Signature) might support the long-term validity of e-documents, using the widely used portable document format (PDF).

Keywords: *long-term storage, digital signature, PAdES, pdf*

Introduction

Recording information has always been important in the history of humanity and also to save it for posterity. The oldest relics we found were cave paintings and were made about 13,000 BC, of which the most famous is located in Lascaux, France. In 3,000 BC a new age began with the formation of writing. After 5,000 years we are also able to recognize the ancient symbols of Uruk, and after 4,000 years we can easily read Hammurabi's laws and papyrus scrolls from the second century BC. We keep these several thousand year old relics in our libraries. [1]

The technological development in the last couple of decades has basically changed these thousand years old ancient traditions. Instead of clay tablets, papyrus, and paper we use magnetic tapes, optical disks, and other electronic devices.

New technologies have a great advantage over the old ones but we have to mention two non-typical properties. While the information was readable with human senses on conventional containers, like paper, now we need special devices to recognize the content of the new data storage (like a Blu-Ray Disc or a Pendrive). Another important difference is that while obtaining information from traditionally stored data does not require any special knowledge, interpreting and displaying data stored in binary format raises some difficulties. To better

1 University of Pécs, Faculty of Engineering and Information Technology, Pécs, Hungary, e-mail: gabor@gyurak.hu

understand these new challenges, let's think about a rightly famous Botticelli painting, made in 1486. Anyone who visits the Uffizi Gallery in Florence can enjoy this artwork. But to view a picture which is saved on a floppy disc and written in Dr. Halo's CUT[2] format may face some complications in opening it. The first problem is that, no one uses this old hardware nowadays, which reads the disc such as a floppy driver. Even if we can manage to read the disc, it will be hard to find software that can open such an obsolete file format and can display the picture.

Most information nowadays is published in digital formats. It makes creating, modifying and forwarding data much more convenient. This goes to the extent that even paper based documents are created electronically and then printed out. We can easily convert our older documents to digital format (e.g. by using a scanner) for easier access.

Digital formats are not only significant because of the more practical management of information, but for its preservation. Certain information such as pictures about unrepeatable events can be crucial for a person. In addition there is lots of information which we have to keep and protect for posterity. Examples range from scientific information to cultural heritage information, [2] but we can also classify the results of nuclear experiments in this group.

Regardless whether we speak about social or personal interests, there are data for which storage has to be guaranteed for decades or centuries. It is hard to explain what is meant by "long-term" data storage, due to the fact there is no clear-cut margin. Depending on the appliances, a few years can be classified as "long-term" but decade long storage can definitely be considered "long-term".

The aim of this paper is to give an overview about challenges with long-term digital signatures and also to describe one possible solution.

Long-Term Data Storage

Based on the previous discussion, the main problems of the long-term data storage of digital documents can be defined. The result of the advance of technology is that the hardware devices and software tools rapidly become out of date. We have to mention that nowadays data storage devices can only store data for a limited time. A commonly used optical data storage disk (CD, DVD, Blu-ray) can only store data for a few years in a trustworthy way. Sadly even the special coated, top of the line disks cannot be expected to work for more than 10–15 years. [3] By the way, more than 15 years storage with one disk is unnecessary because the technology becomes obsolete and there will not be any drives around to obtain information from the disks.[3]

There are only two ways to carry out the practical usage of the long-term data storage. The first one is the migration technique and the other is the emulation one. [4]

The main point of migration is that we transform our data to apply the new technology into a physical and logical frame. While using logical transformation, we convert from an old, obsolete format to a new, standardized one (e.g. converting a Word '97 document to a Word 2013 format). On a physical level while using migration, we have to switch over in certain periods to the new data storage technology (like when we are copying data from

2 Obsolete file format which was supported by the famous picture editor Paintshop in 1997.

3 There are newly developed disks, which according to the manufacturer's claims can store data for 1,000 years. One example for this is the M-DISC by Milleniata (www.mdisc.com)

floppy disc to CD), thus bypassing the technology becoming obsolete. Aside from switching to new technologies, we still have to consider the possibility of storing data in a traditional (non-digital) way (e.g. microfilm,[4] paper).

The other long-term storage solution is emulation, which saves our data in the original format. We eliminate these formats' obsolescence by eliminating the old system's hardware and software environment. There are several types of virtualization techniques to use for this purpose.

Authentic Documents in Electronic Format

In the previous section we talked about documents in general, however we have not mentioned their content. From the beginning of writing there have been "documents" containing vital information. Several techniques evolved in history that were supposed to protect these documents, mainly their confidentiality, integrity and authenticity.

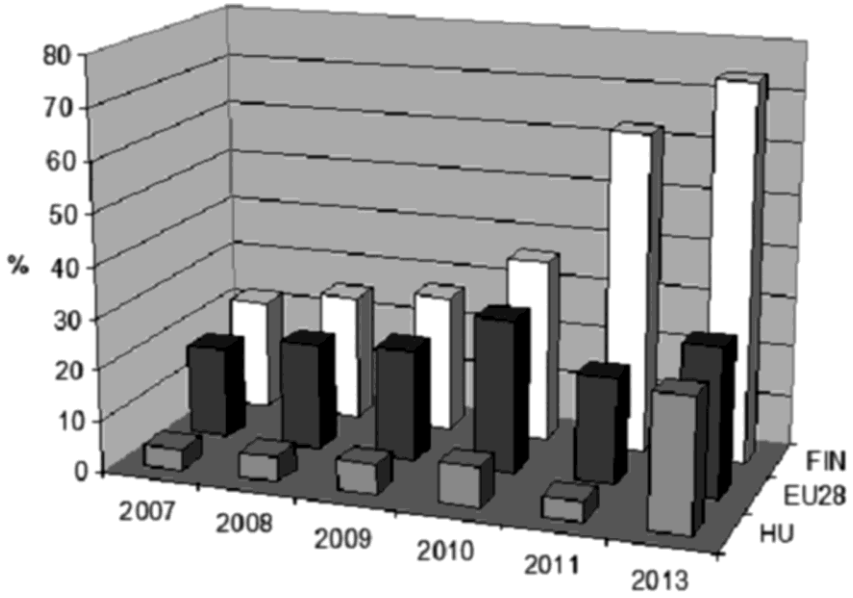
The purpose of confidentiality is to prevent non-authorized people accessing the content of the document. To achieve this, people have been using encryption ever since Sun-Ce's time.[5]

The authenticity of a document means that there is proof about who the creator of the document was. Traditional documents were marked by the handwritten signature of the creator. This is how a document became authentic. Most of the information nowadays is only available electronically, and it would be convenient to store authentic documents that way, as well. Nevertheless, in the field of authenticity, paper based documents are still more dominant, as the advantages of electronic documents (easy to create, copy and modify) in regards of authenticity quickly become their biggest flaws.

One of the important elements of the information society is the formation of a "paperless" government (e-government) system, and also the opportunity for electronic administration. Most of the procedures require the use of authentic documents, mainly electronic invoicing and electronic contracting. One of the core parts of "Europe 2020", the digital agenda classifies e-government as an essential part of a competitive union economy. The EU puts significant effort into spreading electronic administration, mainly towards making electronic invoicing a standard. In the European Commission's communication, titled "Reaping the benefits of e-invoicing for Europe" they called member states to make e-invoicing the standard way of invoicing by 2020. [5]

4 The oldest microfilm is more than 70 years old.

5 Sun-Ce was a military theoretician and mathematician in the 5th century BC.



Graph 1. The rate of eligible concerns of electronic billing. [6]

A statistic provided by Eurostat (Graph 1) shows how companies in different categories can provide the service of electronic invoicing. Hungary reached a huge breakthrough in the year of 2013 by catching up to the EU average. [6]

Electronic invoicing is showing a tendency of growth, and both parties are taking steps towards achieving the goal by 2020. Naturally, we would like to store authentic documents for a long time and sometimes our legal obligation is to store these documents long-term. Over the problems which we had met in the second paragraph, a lot of new challenges arise when we have to guarantee the long-term authenticity of documents. We are going to discuss this in later chapters.

Authentication in Electronic Documents

Technical Background [7] [8: 59–60] [9: 120–124]

The basic principle of authenticating electronic documents is essentially the same as with conventional documents: we sign the document and that signature identifies the creator of the document. The difference between authenticating electronic documents is that we use electronic signatures, generated with cryptographic algorithms.

Just as with handwritten signatures, digital signing should be done in a way that is verifiable and non-forgeable. That is, it must be possible to prove that a document signed by an individual was indeed signed by that individual and that only that individual could have signed the document. Let us consider, Alice and Bob[6] want to communicate via an electronic way. When Bob signs a message, Bob must put something on the message that is unique to

6 Alice and Bob are two commonly used placeholder names in cryptography.

him. Bob could consider attaching a MAC (Message Authentication Code) as the signature, where the MAC is created by appending his key (unique to him) to the message, and then taking the hash. But for Alice to verify the signature, she must also have a copy of the key, in which case the key would not be unique to Bob. Public-key cryptography is an excellent candidate for providing digital signatures.

The gist of the public-key cryptography system is that both parties get a pair of keys. One of the keys is a secret (private) key that the owner cannot give to anyone. The other one is a public key which can be accessed by anyone.

Suppose that Bob wants to digitally sign a document, *m*. We can think of the document as a file or a message that Bob is going to sign and send. To sign this document, Bob simply uses his private key, K_{Bpriv} to compute $E_{KBpriv}(m)$, where E is the encryption algorithm. This value is called the digital signature of the document. If Alice wants to verify the signature she has to take Bob's public key (K_{Bpub}) and she computes $D_{KBpub}[E_{KBpriv}(m)]$, where D is the decryption algorithm. It produces *m* which exactly matches the original document. Encryption and decryption are mathematical operations (exponentiation to the power of *e* or *d* in RSA). After this procedure Alice can be sure about the integrity and author of the message, because of the following reasons:

- Whoever signed the message must have used the private key, K_{Bpriv}, in computing the signature $E_{KBpriv}(m)$, such that $D_{KBpub}[E_{KBpriv}(m)] = m$.
- According to the main principle of the public key cryptography, the only person who could have known the private key, K_{Bpriv}, is Bob.

It is also important to note that if the original document, *m*, is ever modified to some alternate form, *m'*, the signature that Bob created for *m* will not be valid for *m'*, since $D_{KBpub}[E_{KBpriv}(m)]$ does not equal *m'*. Thus we can see that digital signatures also provide message integrity, allowing the receiver to verify that the message was unaltered as well as the source of the message.

One concern with signing data by encryption is that encryption and decryption are computationally expansive. Given the overheads of encryption and decryption, signing data via complete encryption/decryption can be overkill. A more efficient approach is to introduce hash functions into the digital signature. Hash algorithms take a message, m, of arbitrary length and compute a fixed-length fingerprint of the message, donated by H(m). Using a hash function, Bob signs the hash of the message rather than the message itself. Bob calculates $E_{KBpriv}[H(m)]$. Since H(m) is generally much smaller than the original message, the computational effort required to create the digital signature is substantially reduced.

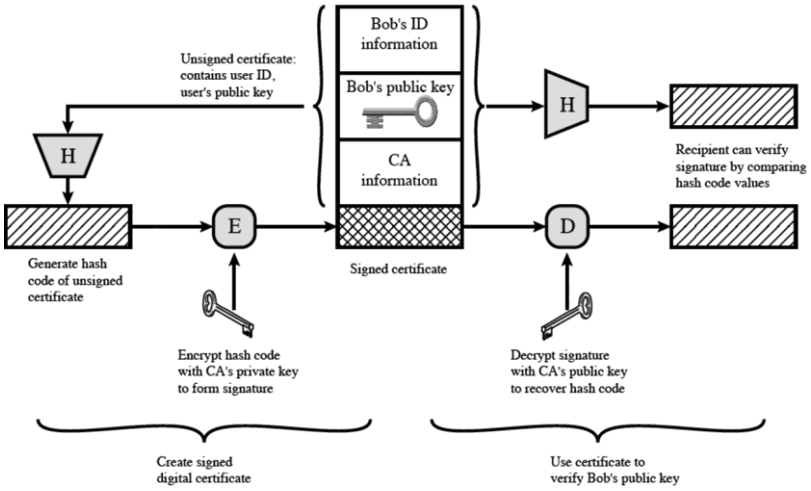


Figure 1. Public-key certificate use. [9: 60]

The digital signature system requires an underlying Public Key Infrastructure (PKI) with certification authorities. Public key certification is certifying that public key belongs to a specific entity. For example, when Alice wants to communicate with Bob using public key cryptography, she needs to verify that the public key that is supposed to be Bob's is indeed Bob's. Binding a public key to a particular entity is typically done by a Certification authority (CA), whose job is to validate identities and issue certificates. A CA has the following roles:

- A CA verifies that an entity (a person, a device, and so on) is who it says it is.
- Once the CA verifies the identity of the entity, the CA creates a certificate that binds the public key and globally unique identifying information about the owner of the public key. The certificate is digitally signed by the CA.

The user can then publish the certificate. Anyone needing this user's public key can obtain the certificate and verify that it is valid by means of the attached trusted signature. The certificate has an expiration date wherein the CA verifies that the particularly public key belongs to a user. The signature on the certificate is made with the CA's private key; therefore this can be verified with its public counterpart. The CA's public key is certified by another CA, thus creating a certificate-chain. The CA is responsible to publish Certificate Revocation Lists (CRL) containing information about invalid certifications and to make certifications verifiable online with Online Certificate Status Protocol (OCSP).

Legal Background

In the previous paragraph we introduced the technical side of the situation, which solves the document authentication in a technical way. However it cannot be used in practice, until it is acknowledged legally. The European Parliament realised the great potential of electronic signatures early, and in 1999 they provided member states with guidelines, with the directive 1999/93/EK. [10] Based on this directive, the law about electronic signatures (Esl – Electronic Signature Law) was passed in Hungary as well, in the form of the year 2001, XXXV. Law (hereinafter: Esl.); [11] it managed to provide sufficient legal background for electronic authentication

The law specifies four services:

- authentication service;
- timestamp service;
- device service;
- electronic archiving service.

The Esl. distinguishes qualified and non-qualified providers. Parallel to that we can talk about qualified electronic signature, increased security electronic signatures and other electronic signatures that do not fit either of those two categories.

Long-Term Certification Affected by Challenges

While signing an electronic document the signer takes responsibility for its content. When authenticating a document, we check whether its signature is valid or not. The Esl. only assigns legal consequences for documents with a valid signature. [11]

Steps of the authentication process:

- we create the hash print of the document H(M);
- we decrypt the signed hash print, using the signer's public key H(M)';
- if the two hashes match [H(M) = H(M)'], we can conclude that the signer of the document possessed the pair of the public key (private key).

With these steps, we can prove that the document has not been modified since it was signed and the signature was made by the private key that belongs to the public key. The next thing we have to check is who the set of keys belong to and whether or not the signer was the only one with access to the private key at the time of the signature. The focus is on the time of commitment, so it is very important that we inspect the circumstances at the time, as well, whether the validation happens right after the signing or decades later. The owner of the public key is verified by the certificate, the authenticity of which CAs are responsible for. The task is to verify if the signer's certificate was valid at the time of signing, as well as if there was a certificate-chain that could be traced back to a root CA's certificate and if all the elements in the chain were valid (the certificates were not suspended or revoked).

As we can see, validity checking is a very complex procedure, which makes inspecting a lot of data necessary. If all of this happens shortly after the signing, the validation process is relatively unproblematic. It is hard to actually say how long this period is exactly, but if we consider the standard expiration time of a certificate, then we talk about a 1–2 year long period (of course if the certificate is not revoked, in that case the time of the revocation is what matters). [12]

If the certificate expires or gets revoked, the signature still remains valid but validity verification becomes necessary. During the verification process, the following problems can arise.

The Signing Date

As far as we cannot prove the signing date, the signature will only be valid, if the certificate is valid too (this means, the validity time has not expired and has not been revoked either). We can increase the validation time, if we are able to prove the signing's date, i.e. putting a timestamp on it. From that point the validity of the timestamp will also be important for verification.

Revocation Information

A certificate validity can be suspended within the validity time, or also can be revoked, typically this happens when we suspect that the private key has been compromised. If there is a timestamp on the signature, in the case of the revocation of the certification, the validation can also been proved. If the signature was made before the revocation, it can be considered valid. The revocation information is published as a CRL by the CA, and it enables OCSP (Online Certificate Status Protocol) queries as well. [13] According to Esl., the service providers are liable to store data after the expiry of the certificate. They should store it for ten years. [14] This also means that if we want to ensure the validity for a longer term, then we have to collect the revocation information, and take care of their long-term storage.

CA Information

To establish the validation of the signature all the data in the certificate chain needs to be checked. The certification authority's certificates can expire. This question affects the validity of the timestamps, because time stamping is usually done by the same organization as the certification management. To achieve long-term validity, it is necessary to collect these data and store it.

Outdated Algorithms

In the background of the electronic signature there are certain cryptographic procedures, encryption algorithms (e.g. RSA [Riverst Sharim Adleman] algorithm), hash algorithms (e.g. MD5, SHA–512) to operate. These have properties that allow the system to work safely, meaning that with the current level of technology, there is no efficient way or sufficient computational power to compromise the system. According to our current knowledge, there is no appropriate way for integer factorization. [7] This is what the RSA encryption is based on, and this is why RSA based electronic signatures are considered safe.

We have arrived at yet another point, where we have to pay attention to the time factor. Those algorithms that we use today might become obsolete in a few years, but decades later will be surely outdated.

The MD5 hash algorithm could be a great example how cryptographical building blocks become obsolete. [15] It was widely used before the millennium. MD5 hashes of documents were provided with an electronic signature. One of the criteria of a hash algorithm's usability is that it has to be collision resistant. It means that it is hard to find two messages that have the same hash print. As it turned out, the MD5 does not meet these criteria so it cannot be used for cryptographic applications.

Another widely used algorithm is the SHA–1, [16] which is not allowed to be used for cryptographic purposes for CA organizations in Hungary since 31 December, 2011.

With the developments of technology a lot more computational power is possible, which makes a brute force attack on one of these solutions really easy. Fortunately, the regulatory side recognized this vulnerability and now there is legislation in place to make companies use safe algorithms and appropriate long keys.

Solutions

We have to find a solution to the problems presented in the previous chapter. A solution that can guarantee the long-term validity of electronically signed documents in technological and legislative aspects.

The 4th paragraph of PadES developed by ETSI, the PadES-LTV (Long-Term Validity) is a development that extends the PDF format with capabilities that allow the long-term validity of an electronically signed document. [17]

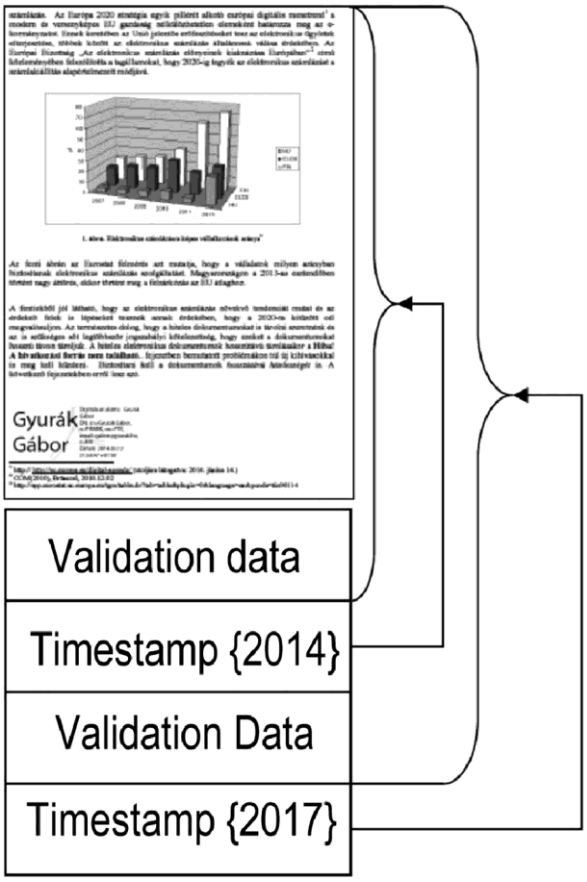


Figure 2. PaDES document. (Created by the author)

The base is a PDF document, provided with an electronic signature and timestamp. Then an extension is added to this, containing additional data that we use for validating the signature (validation data):

- certificates of CAs in the certification chain;
- revocation information (in the form of CRLs and/or OCSP answers);
- certificate of timestamp provider.

According to Figure 2, a timestamp is added to this which provides the document's validity within its expiration date even if the signer's certificate had already expired. The signature can be verified even if the CA's information and revocation information are not available, as they were attached to the document.

If we want to store the document long-term, we can ensure validity through repeating the previously mentioned steps (attaching validation data and timestamp). All we need to pay attention to is that the "update" has to happen before the timestamp's certificate expires and the new algorithm used at the time stamping must be up to the current standards (secure algorithm and long keys). [18] [19: 7–8]

In addition to the technical solution, the system only works if it is supported with the appropriate legislative background. In Hungary, the 114/2007. (XII. 29.) MET (Ministry of Economy and Transport) ministry decree [20] about digital archiving regulates the long-term storage of digitally signed documents.

The decree's para 4 (4) distinguishes, the obligatory period of time to preserve the document, long-term storage, that according to the law in place, means more than 11 years. In this case, it is the job of the one in charge of preservation, to:

- take care of the acquisition and preservation of the information necessary for electronic signatures long-term validation;
- place a timestamp on the certificate-chain, provided by a qualified provider;
- repeat the previous step if the cryptographic algorithms used become obsolete.

We can meet these regulations ourselves, or we can hire an archiving provider to do the job for us. In the latter case, we have to assume the provider does his job well, so in case of a dispute, the conflicting party has to prove the problems with the document's authenticity.

According to the law about accounting, the obligatory time period to preserve electronic invoices is eight years. Based on the rules of archiving this does not classify as long-term so the strict regulations do not apply here. Although, from a technical standpoint, it would still be justified to use the archiving methods mentioned above, even in this "short" period, as the problems outlined in the earlier paragraph can affect our documents in this period, as well.

Summary

From the previous chapters we can clearly see that long-term data storage, especially of authentic documents is a difficult and expensive job. With the above mentioned technologies, we can guarantee long-term authenticity but we should not forget about the usual flaws of long-term storage. Aside from guaranteeing authenticity, we have to still make the interpretation of the original document possible thus reaching a point where the questions discussed in the second chapter come up.

According to the decree of 114/2007. (XII. 29.) para 2 (2):

"The one bound for preservation has to guarantee that the documents stored remain readable – through supplying the appropriate software and hardware environment to open the document – for the time period of said preservation." [20]

According to what we have mentioned above, the migration technique can not be used on a logical level, as the document has to be kept in its original format. The solution to this problem is the emulation technique, or preserving the original hardware and software environment.

References

[1] VÁRKONYI N.: *Az írás és a könyv története.* Budapest: Széphalom Könyvműhely, 2001. [2] ZHOU, M., GENG, G., WU, Z.: *Digital Preservation Technology for Cultural Heritage.* Berlin: Springer, 2012.

[3] PEEK, H., BERGMANS, J. HAAREN, J. van, TOOLENAAR, F., STAN, S.: *Origins and Successors of the Compact Disc*. Berlin: Springer, 2009.

[4] BORGHOFF, U. M., RÖDIG, P., SCHEFFCZKY, J., SCHMITZ, L.: *Long-Term Preservation of Digital Documents*. Berlin: Springer, 2006.

[5] EUROPEAN COMMISSION: *EUROPE 2020.* http://ec.europa.eu/europe2020/ (downloaded: 17 12 2014)

[6] EUROSTAT: *Eurostat Statistics Graph: Enterprises Sending and/or Receiving e-Invoices.* http://epp.eurostat.ec.europa.eu/tgm/table.o?tab=table&plugin=0&language=en&pcode=tin00114 (downloaded: 17 12 2014)

[7] BUTTYÁN L., VAJDA I.: *Kriptográfia és alkalmazásai.* Budapest: Typtotex, 2012.

[8] STALLINGS, W., BROWN, L.: *Computer Security Principles and Practices.* London: Pearson, 2012.

[9] SOLOMON, G., CHAPPLE, M.*: Information Security Illuminated.* Burlington: Jones and Bartlett Learning, 2005.

[10] *The European Parliament and the Commission's directive of 1999/93/EK about electronic signatures social program, 13 December 1999.*

[11] *2001.* évi *XXXV. törvény az elektronikus aláírásról.* http://net.jogtar.hu/jr/gen/hjegy_doc.cgi?docid=A0100035.TV (downloaded: 07 01 2015)

[12] SONG, S., JAJA, J.: *Techniques to Audit and Certify the Long-Term Integrity of Digital Archives.* Berlin: Springer, 2009.

[13] BERTA I.: *Nagy e-szignó könyv.* Budapest: Microsec Kft., 2011. [14] *2001. évi XXXV. törvény az elektronikus aláírásról*.

[15] *RFC 1321: The MD5 Message-Digest Algorithm.* www.ietf.org/rfc/rfc1321.txt (downloaded: 05 01 2015)

[16] *RFC 3174: US Secure Hash Algorithm 1 (SHA1).* https://tools.ietf.org/html/rfc3174 (downloaded: 05 01 2015)

[17] *ETSI TS 102 778 (2009-07) Electronic Signatures and Infrastructures.* www.etsi.org/deliver/etsi_ts/102700_102799/10277801/01.01.01_60/ts_10277801v010101p.pdf (downloaded: 12 01 2015)

[18] POPE, N.: Protecting Long-Term Validity of PDF documents with PAdES-LTV. In. POHLMANN, N., REIMER, H., SCHNEIDER, W. (Eds.), *ISSE 2009 Securing Electronic Business Processes,* Berlin: Springer, 2009. 320–327.

[19] BLANCHETTE, J.: The Digital Signature Dilemma. *Annales des Télécommunications,* Mai/Juin 2006, 1–18.

[20] *114/2007. (XII. 29.) GKM rendelet a digitális archiválás szabályairól.* http://net.jogtar.hu/jr/gen/hjegy_doc.cgi?docid=A0700114.GKM (downloaded: 07 01 2015)

AARMS Vol. 14, No. 2 (2015) 161–171. **10.32565/aarms.2015.2.3**

Distribution of Fire Cases and the Role of Human Factors in Coal-Firing Power Plants in Fuel-Supply Fields and Distribution Systems

ZELE Balázs[1]

Safety issues in power plants and their direct proportional impact on the safety regulations are increasing significantly nowadays. As a result, fire cases on the way from delivery on the conveyor belt to the final usage in the plant, education of the human resources and improvement of safety regulations are basic key issues. Therefore, in this research I examined the distribution of fire cases and the role of human responsibility through the example of a Hungarian lignite firing power plant by using foreign publications and research.

Keywords: *power plant safety regulations, fuel supply fields, human responsibility, distribution of fire cases in power plant*

The status of energy politics in Hungary and all over the world has been changed in a lot of fields. It is not only relevant in case of general technical improvements, but it also plays a great role in the field of decreasing environmental pollution and development of general standards of the public health system.

"One of the main tasks of energy politics is to convince society about the theoretical and functional ambitions of sustainable energetics. Besides that, another important task of energy politics is to emphasize more carbon-free energy generation in order to maintain the safe provision of the regulations (e.g. CO2 quota, choosing of firing resource); and of course to foster the foundations of national power plants which are based on national and import coal in order to reduce growth of the average producer price." [1]

Moreover, the fear of energy addiction does not only effect our present, but is based on the past, we clearly have to increase the safe energy provision and in line with that decrease energy addiction. One of the oldest questions of energy maintenance is the energy transformation process in power plants. In the past few years Europe – and of course Hungary also – is dealing with the global danger of environmental issues, in which the focus problem is environmental protection. Hungarian energy politics is influenced by the usage of the national resources (e.g. coal/lignite resources) in a larger amount besides the utilization of renewable energy resources. This aspect is based on the ambition of energy generation from basically the usage of national coal resources. If we review the Hungarian resources, we can see that there is a larger amount of lignite in the surrounding Mátra Mountain and black coal at the base of of Mecsek Mountain. In the past years the press was buzzing because of the closure of some mining centers (e.g.: in Mecsek mountain) and then opening them up again. "In the integration of mining centers and power plants, most of the previously functioning mines had been closed down. (In 2003: Balinka, Budaberke storage, Sajómercse, Mákvölgy, Fe-

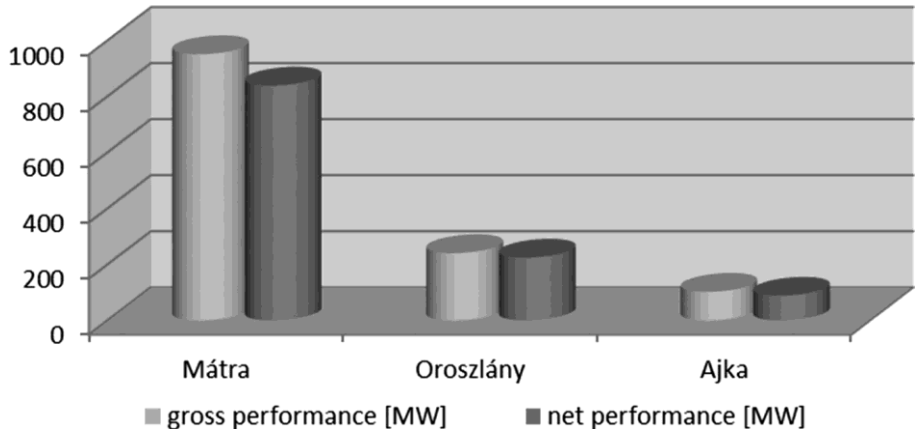
1 Certified energetic engineer and economist, sales energetic engineer, Siemens Zrt., Óbuda University, Ph.D. candidate, e-mail: zelebalazs@gmail.com

ketevölgy, Szuhakálló; in 2004 the openwork in Pécs, Mány, Ármin, Lyukóbánya, Lencsehegy. In 2005, the only deep working mining center was Márkushegy, and besides that Visonta, Bükkábrány, and some other little openworks in Borsod and Nógrád county.)" [2]

Despite these, it is sure that the utilization of these resources was not only significant in the past, but it also could be one of the future's unexploited sources. However, appropriate environmental politics should be worked out and utilized, which in my opinion is a step on the right path in order to reach the goals of sustainable energy provision.

The main goal of this current work is to examine and present the contact and presence ratio of causing factors of unregulated fire cases in Hungary's biggest energy provider and distributor power plant which is based on lignite firing. Basically, in this field, besides the coal dust scattering and the technological hazards, it is mainly the human influencing factor which I am going to examine in depth, based on previous experience.

The opinion about coal firing power plants and the applicability of technologies have changed a little bit in the past few years. It is more and more widespread besides firing lignite to utilize biomass resources, too and this movement contributes to coal technology in a more environmentally gentle form. The essence of the technology is combining the appropriate amount and quality of biomass and coal in line with providing the necessary heating value. Based on these, the clearly coal-used technology is behind us, thus taking the national efficiency data into consideration, not only coal usage, but in line with the fact that biomass utilization has also been integrated into the determination of a power plant's effectiveness. (Graph 1)

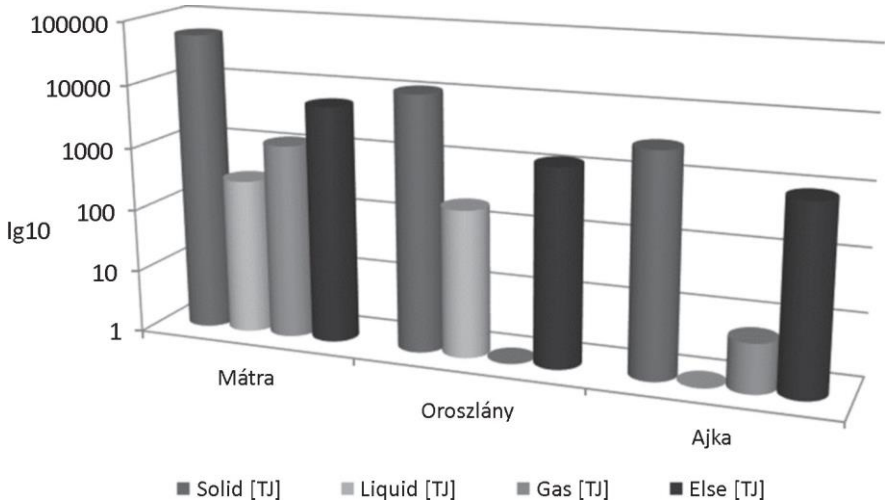


Graph 1. Power Plant energy balance in 2012.
(Own editing – based on MAVIR data. [3])[2]

2 Note: The ratio of the gross electricity power generation of power plants (MWh/a) and the titular, gross performance (MW) is the utilization of the whole power plant area in performance (h/a). (STRÓBL A.: *The analysis of changes in the European and Hungarian electricity-maintenance, providing supply-safety and capacity-research studies*.)

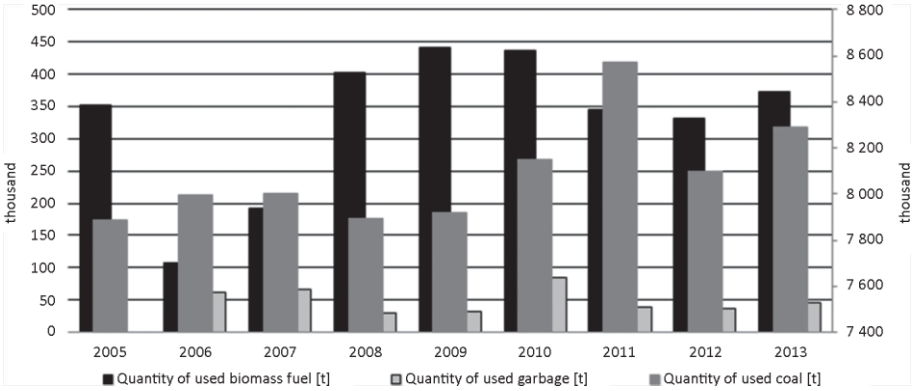
In 2012 the performance of power plants which are based on bio-fuels mixed with mostly carbon sources, can be seen in the first diagram. Based on the sum of these numbers about 1,300 MW performance was within the system, which was a preview of 13% performance of the year 2012. If we take the 2,000 MW gross performance of Paks Power Plant as a basis, – which gave 45.9% of the electric power in the whole country for this current year – these data, besides the nuclear energy utilization can be considered significant values owing to the role of the Mátra Power Plant, too.

On the second diagram, we can see the scattering of energy resource utilization in the plants which transform energy from coal and use it, and the role of Mátra Power Plant is quite outstanding. Thus the examination of this project of mine mostly is about "one of the largest electric power generating plants which is also the biggest coal-firing plant in Hungary" based on the official release of the plant. [4] Analyzing the current situation of the country, a more significant changing method is going on, which affects the power generating plants, basically the production of coal-firing power functions. Considering energy policy aspects and EU directives, the energy maintenance of the country should be provided by more and more renewable resources in the near future. It is set officially in the announcement of the European Commission as the following: "the European Union is going on a good path towards reaching its targets, which means that until 2020, 20% of its energy usage should be provided by renewable energy resources. This initiative is a part of an expansive EU strategy, which is supposed to decrease the climate change. This is absolutely a good thing. Energy from wind, solar, water, tide, geometric and biomass utilization in a larger amount can decrease the European Union's energy import dependence, and it also stimulates innovation and the employment." [5] On the second diagram, based on the 2012 data about the energy consumption of Hungary, continuously based on the previously specified aspects – we can see that how great a role the Mátra power plant has in our country. Besides that we can also see the ratio of other energy resources usage (gas, liquid, etc.), which is on a good path reaching the targets of EU, which helps to maintain a safe and environmentally friendly energy consumption.

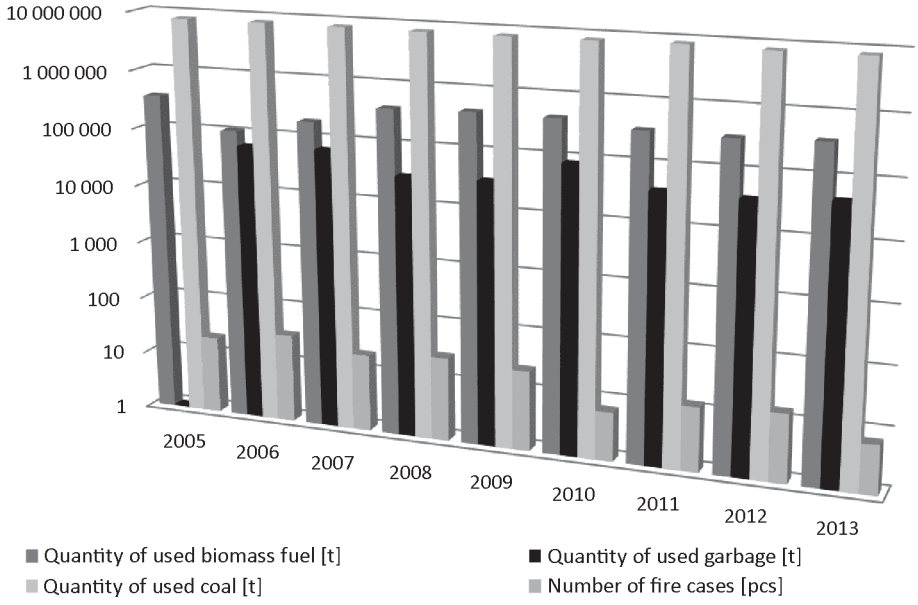


Graph 2. Energy resources utilization "in the biggest power plants remaining today" 2012.
(Own edition – based on MAVIR data. [3])

In the further parts of the analysis I focus on the Mátra Power Plant and I examine the case study from the plant, and I draw the final conclusion from them. You can see on the third diagram the statement about the changes in the distribution of the compact energy fuels in the last 10 years. Based on this it can be seen, that besides the considerably balanced lignite usage, the bio and "other" garbage utilization has appeared on the market and in line with that the utilization ratio has also decreased significantly.



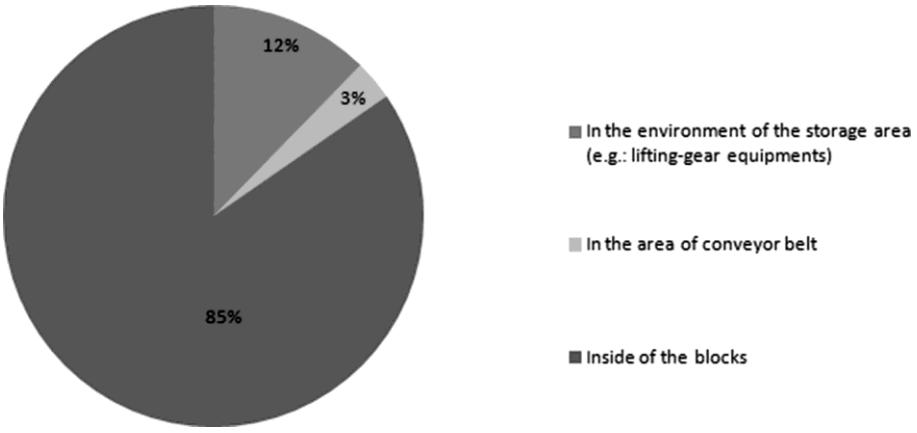
Graph 3. The ratio of used concrete fuels in ME Zrt.
(Own editing – based on Mátra Power Plant Zrt., Department of Thermal Engineering; consultations and reports with associates, January–February 2014.)



Graph 4. The ratio of concrete fuels and fire cases in ME Zrt.
(Own editing – based on Mátra Power Plant Zrt., Department of Thermal Engineering; consultations and reports with associates, January–February 2014.)

However, considering the reports of the plant in the examined period of time, the number of fire cases shows us a decreasing tendency. Hereafter, my examination is focusing on the causes, influencing facts and hazards of these fire cases; I also analyze which cases were the most common ones, and of course whether there is coherence between the distribution of indicators.

Based on my previous research and the reports of the plant, in the last 10 years the most common fire cause was the coal dust, the oil-drainage after technical damage and malfunctions, and the fire cases influenced by human errors. This conclusion is also strengthened by my further analysis, too, in which I examine and present the characteristics of fire cases caused by coal dusting process in the line between the coal storage area and the furnace – it can be seen on Graph 5.



Graph 5. Distribution of fire cases in the Mátra Power Plant.
(Based on Securing Department data, 2000–2012.) [6]

85% of the fire cases were caused by the coal dust spilling/inordinate scattering within the blocks. Other fire cases began directly in the coal storage area, on the conveyor belt and in its surrounding. I confirm these written facts with scientific literature, and besides this I would like to prove my previous statements as well.

Examining international literature resources and American examples, I found an interesting lecture, in which the lecturer called coal "the necessary evil." [7] He meant by this to draw attention to the reaction status of firing coal, thus the thermo process of firing, and especially to the hazardous consequences of the unregulated coal dust burning and explosion effect. "In those places, where the smooth, little granule coal dust can settle, this massed coal dust can ignite – these places can be locations like the equipment beyond the transportation path and the conveyor belt." [7] Besides that, during the transportation of fuel into the furnace, it is also possible to spill and spread dust within the industrial area, which can aggregate and stick into the different tubes and tanks or containers.

The correspondence between the thickness of the coal dust and the thermo reaction (the process of the ignition) depends on the followings.

$$T_{ex} = \frac{H \times A_{tot}}{87{,}5\mathrm{p} \times A_{dust}}$$

where:

- T_{ex} = thickness of coal dust layer, (which contributes to explosion hazard);
- A_{tot} = the whole floorspace (20,000ft^3 the upper measuring limit); [7]
- H = the height of floorspace/building (ft);
- p = the volume of spilled dust (lb/ft^3);
- A_{dust} = the surfaces where dust can accumulate within the building (ft^2).

It is also specified in the publication that human hazards and fire safety protection – (maintenance defaults, smoking prohibitions, compulsory prevention trainings, etc.) – play a great role in the formation of fire cases and explosions caused by coal dust. Thus in the following I examine the scattering and influencing factors of these cases based on the previously mentioned domestic power plant example. Furthermore, as a new theme, I compare the different alternatives with the occurring factors of other causes. However, before I start this, I shall sum up in the first table the causes of the most common fire cases which occurred in this system. Thus the fire cases of the largest Hungarian lignite firing power plant gave the basis to my further examinations. In this process an example, a case study of a Namibian coal firing power plant – with a performance of 150–800 MW – assisted my work. This study categorizes the fire cases caused by mechanical failures into "low hazardous factor" category, however we would rather talk about the more significant influencing factors. [8] Thus probably, the quality, age and other natural factors of fuel – here it is coal – that can be a basis of further analysis. (For example: what kind of influencing factors are there.)

As a consequence, we cannot make further, significant conclusions based on the experience of these one or two plants. However, during a power plant investment in Hungary or abroad in the future, the theories summed up here can be quite usable based on the different firing fuels.

Besides the safety equipment and different solutions, it is worth giving some attention to the factors of the human hazard component, and examining the role of it in the prevention of fire cases and of course the causing influence of it, too.

In the past few years, the plant has deepened the relationship between education and the employees – the decreased number of fire cases also shows that. Thus it can be proven – if we take an equal mechanical failure rate into consideration – that the human hazard factor has influenced the number of occurred fire cases.

3 Note: based on USA measuring system: yard^3 = 27 ft^3 ≈ 0.7646 m^3 in SI dimensions.

Table 1. Distribution of fire cases in the Mátra Power Plant.
(Own edition – based on Securing Department data, 2000–2012.) [6] [9]

Causing object of fire cases (Mátrai Erőmű Zrt.)	**Number of fire cases (pcs.) (2000–2012)**
1. Fire cases caused by unregulated coal dust spilling (coal storage area, within the blocks)	130
2. Fire cases caused by mechanical failures or oil leak (basically turbine oil leak)	66
3. Fire cases caused by other human errors (e.g.: defaults of work and fire protection rules during maintenance welding – assembly scaffold caught on fire)	58
SUM	254

The number of fire cases based on the influencing factors is present basically because of the coal dust spilling. However, further fire cases are caused by mechanical failures or oil leaks – basically turbine oil leaks and the ambient temperature-caused fires. Finally, the human hazardous facts also influence fire cases, which occurred because of the lack of attention and other inappropriate behavior (e.g. inappropriate application of work and safety protection rules, or not complying with them).

My basic target in this publication was to examine the eventuated different power plant fire cases, thus I analyze the connection between the number of occurred fire cases and the lost production capacity. (Table 2)

Based on the fire examination reports of the plant, I can sum up in the following table the role of important human hazards in different cases, like production capacity and the loss frequency ratio. We can observe the events caused by human hazards in addition to other different plant problems – examining a more than 5-year-interval. (Table 1)

However the larger and more dangerous fire cases were handled in time by the professional personnel and fire protection system quite well and fast, so the further malfunction and production loss was avoided. (Table 2)

Table 2. Different types and distribution of fire cases 2000–2005. [6] [9]
(Own edition.)

Main fire cases (Mátra Power Plant Zrt.)	**Examined year**	**Number of fire cases**	**Plant malfunction / production loss (pcs)**	**Causes discovered in the examination**
Caused by unregulated coal dust distribution (in the coal storage area and its surrounding)	2000	0	-	Massed coal dust, spontaneous combustion
	2001	3	-	
	2002	0	-	
	2003	0	-	
	2004	2	-	
	2005	4	-	
Caused by coal dust spilling (inside the blocks, on the conveyor belt, and its surrounding)	2000	2	-	Massed coal dust, spontaneous combustion
	2001	8	-	
	2002	7	-	
	2003	9	1	
	2004	5	-	
	2005	10	-	
Caused by technical or electronic failure (e.g.: oil spilling, electronic malfunction)	2000	14	-	Regular audits (control), maintenance
	2001	10	-	
	2002	9	4/1	
	2003	2	-	
	2004	5	-	
	2005	4	-	
Caused by other human failures (e.g.: defaults of work and fire protection rules during maintenance welding – assembly scaffold caught on fire)	2000	8	1	Human behavior, defaults, failures, inobservance causes (consequence: personalized educational training program, strict regulation of open flame usage)
	2001	2	-	
	2002	3	-	
	2003	4	-	
	2004	5	-	
	2005	6	-	

It can be said, that there is no connection between the production loss and the human influence ratio, consequently the power plant fits the general mechanical and performance regulations from a safety technical point of view, the fire safety system and the general protection mechanism are appropriate. However, the human factor (human hazards) can significantly influence the outcome of the different situations in a positive and also a negative way, thus it can be considered an important influencing factor.

A further consideration is how the human factor and the different safety fire protection equipment are connected and work.

Based on the plant reports, the following factors contributed to the start of a fire: human behavior, human hazards, inobservance causes. Besides this as a consequence, a development of a new educational training program, the strict regulation of open flame usage or the previously planned work and fire protection training and lectures were the factors which contributed to avoiding the occurrence of fire. Furthermore, a quite common case is hazards caused by coal dust spilling, which needs to be looked out for in the future as well. Actually, it has a quite great role in the international literature, despite the fact that we have to take into consideration the different environmental factors and the technical/mechanical influences.

Finally, I examined those human factors which are the most common cases before and during the formation of a fire.

Human beings, as a basic element of a great whole system, control the technology and regulate the different methods, and can also be a causing component of an accident which disrupts the balance and causes production loss. This is based on a scientific statement in the following wording: human performance basically can influence the complex technical system and the safety and responsibility level of the mechanical equipment. The different safety and risk analyzing reports say that the appropriate handling of the human interactions can be a key factor in understanding the different accidents and their risk components. The aim of the human responsibility analysis (HRA) is to identify the different human interactions, analyze them, and after the thorough examination, build them into the safety regulations, training tutorials. It is not enough to measure the possibility of the success and failure numbers, it is also important to make suggestions in order to develop the human performance.

It is important to emphasize the factors which are more significant in the process of development. These factors can be the following: human and machine harmonization, development of processes and educational training system, harmonization of working expectations with human skills, and the decreasing of the effect of human errors which correlate with each other.

It was also mentioned before, how the systematization of the human capacity and performance can play a significant role. With this, we can avoid the danger on a long-term basis for the safety of complex mechanical equipment caused by human damage factors.

The human factor has also been analyzed in the field of nuclear power plant maintenance and utilization, too, where it had been determined, that "it is a complex system, which is built on technological and human factors, and this duality can be seen in the safety issues as well." [10]

This examination of mine shows the mentioned fire cases between 2000 and 2005, and it can be seen from these cases what kind of causal side-effects were responsible for the fires, besides the basic coal-dust spillage.

The review of these events can be performed in the future, too. Furthermore, to avoid these accidents, it is also suggested that the safety level be raised, and to decrease the human damage factor, as much as possible. The previously mentioned American lecture about the coal-dust combustion has also contributed to this statement. However, for example, it does not write about the distribution ratio, which can be different, depending on the location and different coal-type utilization.

It can be seen, that during the transportation process from the storage area to the furnace it is more common that a fire is formatted because of coal-dust spillage. Furthermore it is also more common that it caused more significant damage and danger – or even a catastrophe – inside the blocks in connection with the electric system.

Searching for international risk analysis and publications I could determine, that there are some statements and examinations, [8] which analyzes the different risk factors within the power plant environment. However, there was no examinations or analysis regarding the fire cases and their later effects caused by events between the transport and burning process – whether it was caused by coal-dust spillage, mechanical breakdowns, or human hazards. The decreasing alternatives of the previously examined and analyzed fire cases can be seen in the next graph, which also shows the possible formation places of fires and the most common formation causes in coal-firing power plants, as well. [8]

Location of fire cases, considering the departments within the plant:

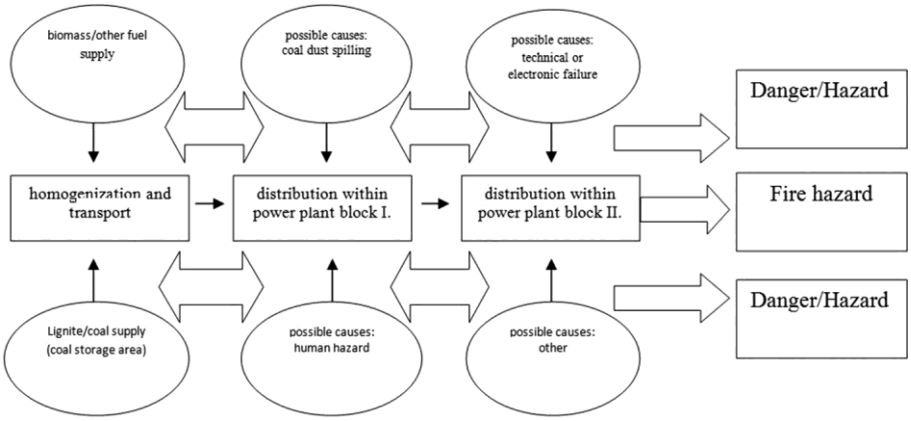


Figure 1. Location of fire cases and the most common causal factors during the resource-transportation process of the power plant. (Own edition.)

Summary

In this publication I analyzed and presented the performance and position in the Hungarian energy production structure of domestic combining coal and biomass fired power plants based on today's present knowledge. I confirmed with strategic figures their place in the Hungarian energy industry and in addition to this I demonstrated the different cases in the field of safety and fire protection area. Furthermore I analyzed the hazards induced by engineering operational failure, formation of fire cases and of course the influencing role of human factors. Based on these information I suggested different, safety increasing problem solving possibilities.

I also analyzed the role of the human factor and made an analytical comparison based on international experience and different cases of a Hungarian power plant. Besides the human errors I identified the importance of human positive thinking, the situational awareness and the ability to act fast and accurately in a dangerous situation.

Analyzing the different fire cases the interaction of complex and multiple factors and interacting components can be seen. That is why we can tell that in the whole system the role of human impact role is an important issue besides the general engineering and technical failure influence. Additionally, in my opinion, the domino principle (interacting influence

role) is also present in these cases, which can be another interesting topic in the future as a new investigation.

All in all, if the number of cases show almost equal distribution, and even if the man-ma- chine-environment system has a huge influencing impact, quality and quantity of the fuel is almost the most important factor, in that the firing process and unregulated, uncontrolled hazards may depend on. To avoid and prevent accidents there are some current developments which can be solutions to the problems, so increased security and fire protection regulations, structural and targeted educational and training policy and good resource management can be key issues besides the general technical compliances and regulations.

References

[1] ŐSZ J.: *Megújuló energiaforrások hasznosítása.* Budapest: BME Energetic Equipments and Systems Department, s.a. http://energia.bme.hu/~kaszas/Energiapolitika/OszJanos/Meg%C3%BAjul%C3%B3%20energiaforr%C3%A1sok.pdf (downloaded: 14 12 2013)

[2] KAJATI Gy.: *The Post-Socialist Transformation of the Hungarian Electricity Industry*. Debrecen: Debreceni Egyetem, 2008. (Ph.D. dissertation)

[3] MAVIR ZRT.: *A Magyar Villamosenergia-rendszer közép- és hosszú távú forrásoldali kapacitásfejlesztése.* Budapest, 2013. www.mavir.hu/documents/10258/15461/Forr%C3%A1selemz%C3%A9s_2013.pdf/0a51f06c-73e7-4607-b582-00d3b1434837 (downloaded: 12 01 2014)

[4] MÁTRAI ERŐMŰ ZRT.: *Official Webpage (online).* www.mert.hu/hu (downloaded: 25 01 2014)

[5] EUROPEAN COMMISSION: *Uniós terv a megújuló energia-felhasználás növeléséért.* 08 06 2012. http://ec.europa.eu/news/energy/120608_hu.htm (downloaded: 11 01 2014)

[6] ZELE B., MÁTRAI ERŐMŰ ZRT.: *Safety Department – Mechanic Section Reports and Consultations. Reports with Plant Associates.* Visonta: January–February, 2014.

[7] CLIFT, J.: *Combustible Dust*. – Lecture sheet online. Hazard Control Technologies. www.hct-world.com (downloaded: 04 01 2014)

[8] NAMPOWER COAL-FIRED POWER STATION, NAMIBIA ENVIRONMENTAL: *Environmental and Socio-Economic, Risk Assessment Report*. Reference Number: 5975; 15 May 2012. Prepared by Aurecon South Africa (Pty) Ltd. (09 01 2014)

[9] SZABADOS G. T.: *Natural and Human Factors at Hazards of Mining Process and their Control – The Role of Environmental and Human Factors in the Mining Hazards and the Protection against them.* Miskolc: Miskolci Egyetem, 2011. (Ph.D. dissertation)

[10] KOVÁCS J.: Human Factors Role on Risk Analysis of Complex Systems. *Bolyai Szemle*, XVI 2 (2007). portal.zmne.hu/download/bjkmk/bsz/bszemle2007/2/bsz2007_2.html (downloaded: 12 01 2014)

AARMS Vol. 14, No. 2 (2015) 173–185. **10.32565/aarms.2015.2.4**

The Questions of Piracy in the Light of International Law and the Responsibility of a Failed State

HARKAI István[1]

These days, one of the most significant issue is maritime piracy and armed robbery; it poses a high threat against international peace and security. These crimes can occur anywhere on the high seas, but the most infected areas are the western basin of the Indian Ocean, the Gulf of Aden, Southeast Asia, or the Gulf of Guinea, where piracy causes many troubles to world trade. The crime of piracy calls for a strong and substantive answer. In this paper, the author tries to look for legal and non-legal devices against piracy and tries to give an answer to the question whether we can take a failed state to account in international law?
Keywords: *Somalia, piracy, jurisdiction, responsibility, failed state*

Introduction

The most infected territory for pirate-activity is the so-called Horn of Africa, namely the Gulf of Aden, and the western basin of the Indian Ocean, but robbers are present at the coast of Southeast Asia, and significant pirate activity is evolving in the Gulf of Guinea.

The owners of vessels attacked by maritime bandits and the Flag States try to act against piracy and attempt to drive back and discourage these activities. The efforts against Somali pirates have been significantly successful. The act of piracy is one of the oldest "professions" in the world caused by deep-rooted social and political problems. This is the reason for the international community not being able to eliminate the international crime of piracy and armed robbery despite all the endeavors to heal the causes, yet these have not been enough to permanently abolish piracy.

History of Piracy and Maritime Robbery

History of crimes committed on seas and commercial shipping were born in the same age. The so-called *Sea Peoples* ravaged the coast of the Eastern Mediterranean in the 14th century before Christ. [1] Merchants of ancient Greek city-states were exposed to looting on the Aegean Sea. *Plutarch* also urged actions against maritime bandits and emphasized that the coastal state could only claim supremacy over the part of the sea belonging to its authority if the Greek states establish security. [2] Not all Hellenic citizens considered pirates offenders. *Hieronymus of Cardia* designated pirate-activity as an "honorable entrepreneurial activity", where pirates are mercenaries, brothers-in-arms. [3]

1 dr. jur., e-mail: harkai.istvan89@gmail.com

The security of sailing was an important strategic issue in the Roman era as well. According to Strabo, the historian-geographer with Greek descendants, two types of groups existed. One of them was a civilized one and used the seas for peaceful purposes. The other one was a group of barbarians, who committed various misdeeds. [4]

On the eve of the *Great Migration*, Goth invaders pillaged along the seashore of the Black Sea, and reached Crete and Cyprus. Franks and Germans harassed the trade routes on the North Sea. Patrick the monk was abducted by Celtic pirates from Wales and dragged to Ireland. Later he was revered as a saint. [5]

The Vikings were the most notorious maritime raiders in the *Middle Ages*. They bothered not only the coastal regions but internal lands as well. They gladly marched deep inside the attacked countries using rivers as water roads. [6] Byzantine territorial waters were threatened by the expansionist Arabic sailors. The Byzantine Empire had quite a developed legal system. In the *Rhodian Sea Law (Lex Rhodia)*, the rules of commerce, navigation and defense against pirates were collected. [2]

The real golden age of piracy was in the 16–17th century. In this era not only the so-called *Jolly Roger*, the characteristic pirate flag was born, but the three main types of piracy – *privateers, buccaneers and corsairs* – as well. The "authentic" pirates belonged to the first category. They were authorized by their ruler, who gave them the so-called *letter of marque*, to attack hostile vessels. The buccaneers were organized and controlled from the bases on the West Indies. The corsairs were Muslim robbers of the Mediterranean Sea in the 16–19th centuries. Their headquarters were located in Algiers, Tunis, Tripoli and Morocco on the north coast of Africa. They were also authorized to attack Christian sails. Their Christian opponents were the Corsairs of Malta. The grandmaster of the Knights of Saint John allowed them to raid the Muslim ships. [6]

The golden age of organized piracy ended in 1816 with the bombing of Algiers, in Southeast Asia pirate activity was defeated by the Ducht, while on the South China Sea it was defeated by the English fleet. Naval powers banned the issue of letters of marque and proscribed the privateers in the Declaration of Paris in 1856. [6]

The Term Piracy

If we would like to define the modern *term* piracy, we have to invoke Article 100–107 of the Convention on the Law of the Sea, which was accepted in 1982. [7] The term is consuetudinary and not every country has joined to the United Nations Convention on the Law of the Sea (UNCLOS) Treaty of Montego Bay. [8]

The modern definition has many antecedents in legal history. It is a basic fact that the offenders of such a crime have to be considered *hostis humani generis*, the enemies of mankind. Because they are beyond legal protection, each and every person is allowed to act against them. [9] Ancient Greek sources created the first terms for the trading people of the eastern Mediterranean in the 2nd century BC; that was the *peirato.* Romans used a similar word, *peiraton*, which meant free sailors who stood apart from the scope of the law. [10] The famous orator and jurist consul *Marcus Tullius Cicero* alleged that the pirates are *enemies of every community*. [10]

Among the scientists of the late Middle Ages[2] *John of Trevisa*, a teacher at the University of Oxford in England was the first, who translated the word *piratae* as *"see theves (sea thieves)".* [10]

In the 17th century, English and Dutch legal literature dealt in depth with the "pirate-question". According to the definition of *Cornelis van Bynkershoek*, pirates are the ones robbing on the high seas and looting on the mainland without the permission of a sovereign power. In the terminology of *Charles Molloy* the pirates are *sea rovers*, *enemies of the entire human race*. They confront not only mankind, but they act against a specific state as well. Later on, legal experts extended the terminology with *aggressiveness* and *robbery*. [9]

During the development of the English legal definition of piracy, the term gradually extended with elements from the field of criminal law. First, we have to mention *criminal intent,* so the *straight intent (doluc directus)* of the pirates has to be directed at asportation (felonious removal) of the attacked ships and the possessions shipboard.[3] When two opposing hostile states loot each other's ships, it has to be considered sea robbery. However, it is obviously not piracy, because in a state of war plunder is allowed. Dozens of such acts occurred during the American Civil War or in World War II. [11]

The protected legal interest is not only the protection of property, but also, the peace of mankind as well, [9] the order and peace of the high seas. [9]

The first attempt of the *codification* of piracy made by the Committee of Experts of the League of Nations was in 1926. In 1930, the *22nd Article* of the *London Naval Treaty* extended the rules, regarding submarines. Until 1956, there was a debate on the question whether naval ships could commit piracy, when the *International Law Commission of the UN* brought the debate to an end, declaring, "piracy could only be committed by privately owned ships, not by warships." [9]

Codifiers of the 20th century argued about the unlawfully appropriation. From their point of view, it was not necessary to ascertain the guiltiness that the intention of pirates cover the *animus furandi* because the motivation for crime could be anything else, for example hatred or vengeance.

The first declared terminology of piracy appeared in the 15th Article of Convention on the High Seas in 1958. This definition is confined to the private acts, which were committed against private ships. [12] The *Agreement of Nyon* from 1937 gives us a wider expression when considering submarines as surface warships. [9]

The 1982 UNCLOS Treaty integrated the rules of the Treaty of 1958 and its developments into a single frame. But there was no common denomination in two important questions. One of them was the animus furandi, the question of private interest, the other one was the place where the crime was committed, the high seas, because the coastal water is under the criminal jurisdiction of the coastal state, where the act could only be considered piracy if the criminal law of the state contains the statutory definition of piracy.[4]

---

2 In the Middle Ages contemporary languages referred to the sea robbers according to their nationalities, so for example the Vikings, who committed many of pirate acts. English resources used the term *Dani piratae* for the first time in the 14th century.

3 Sir Matthew Hale, Sir Edward East, James Kent.

4 This is mainly relevant in case of Southeast Asia, because most of those crimes which were committed in the region occurred on the coastal waters.

The Article 101 of UNCLOS Treaty classifies the following acts as piracy:

Piracy consists of any of the following acts:

a) any illegal acts of violence or detention, or any act of depredation, committed for private ends by the crew or the passengers of a private ship or a private aircraft, and directed:

i) on the high seas, against another ship or aircraft, or against persons or property on board such ship or aircraft;

ii) against a ship, aircraft, persons or property in a place outside the jurisdiction of any State (…). [13]

Moreover, the norm knows the formation of incitement, but it does not consist of the attempt of a crime, which is considered by the legal practice materialized when the act entered into the period of attempt. The preparation is also missing.

The crime of piracy is a sort of "opened statutory definition", because felony, which is hurting the protected legal interest, could be realised in countless ways, not to mention that the intention of asportation is not declared squarely, so the effect of the crime and the time when the legal interest is harmed is unclear.

The term of high seas as a place of committing the crime should not form a subject of a debate, because it is clear that the crimes committed on coastal waters exclusively belong to the jurisdiction of the coastal state, while the high seas, which are *res communis omnium usus*, fall under universal jurisdiction.

The Types of Piracy

The *first* and easiest mode of attack is when pirates step on board a ship (*boarding*), plunder the crew, and then leave. In the *second* case, pirates deprive the crew of their properties, and then take the whole cargo. Sea robbers ambush vessels in the early morning hours with well organised, 6–7 – occasionally more (even 70) – membered-armed groups. The mother ship with an advanced navigation system carries the equipment and the fuel. The bandits approach the target with powerboats and climb up on the stern. The so-called *phantom ships* belong to the *third type.* Pirates take not only the cargo, but also the vessel with the whole crew on it. After this, they sell the cargo and ask ransom for the hostages. The stolen ship will be repainted, renamed and re-catalogued in a foreign country. [14]

According to the definition of the *International Maritime Organization,* there are more specific differences as well. We can make a distinction between *common piracy* and *political piracy*. [15] Common piracy is when pirates attack a vessel only for the sake of private interests. These are the so-called *low-level armed robberies*, which are brought to effect near the seashore; the bandits use boats and cold weapons or small caliber handguns. The *medium-level armed assault and robbery is* escorted by a mother ship and committed by well-armed pirates. The "capital crime" is the *"major criminal highjack",* the hijack of the attacked ship. To carry out this type of attack, sea thieves need detailed plans, ample resources, trained and armed attackers, and last but not least, land management. [15]

Political piracy is founded on political purposes and motivations; the intent of gaining profit is only a subsidiary option. These actions are rather committed by groups, which are linked to terrorist organizations. [15]

Nevertheless, we can identify interesting connections between the two categories. The two big Somali Islamist rebel groups, *Al Shabab* and the *Hizbul Islam* are in rapport with the lords of pirate-companies, who support them from the abundant prizes coming from pirate-activity.[5] [15]

Causes of the Phenomenon

Primarily, piracy is a "subsistence" crime. Attacks are committed by residents of regions where the population lives in relative poverty. In addition, the closeness of the sea is also a basic condition. Pirates need headquarters and supplies, which can easily be provided by the population of the coastal area.

But why exactly has Somalia has become *Tortuga*, a "pirate paradise", in the last decade? We can answer this question with three quite simple arguments. One of them is the geographical location of Somalia, which is situated on the Horn of Africa, in a perfect geostrategic position, which makes it easier to control the area, or at least collect pieces of information about the merchant convoys sailing along the Somali coasts. The other reason is anarchy. It is beyond dispute that Somalia is a failed state, without administration of justice. However, if there were some kind of central jurisdiction, it could restrain criminal activities. This statement is demonstrated by that six-month period when the Islamic Courts practically abolished piracy, but when the Islamist government fell, sea bandits reorganized themselves and the attacks continued. [16]

Somalia is characterized by political instability; the federal government is only able to control the capital, Mogadishu. Two thirds of the young population is unemployed, the households have to live on only two dollars per day. [17] The whole population is dependent on the international food supply, which is more than 150 tons every year. Regrettably, international aid shipments are popular targets of the pirates. [17]

After the long civil war,[6] since 1991, on the coastal waters – lacking Somali coast guards – the main European, Asian and African companies started to exploit and pollute the coastal waters. The Somali people, deprived of their livelihood, tried to protect themselves as much as they could, so the first pirates came from among the fishermen who knew the surrounding waters well and were familiar with navigation. [17]

Later on, a significant part of the coastguard joined the fishermen. In a short period of time, the little groups formed into small clans with a hierarchy. The clans from the region near *Kismayoo* recognize the primacy of the clans of *Harardheere* and *Hobyo*. [16] The clans are independent of tribal and ethnic bonds; it is hard to guess their membership.

In one band, around 50 members can be found, and there are some groups with Pakistani and Bantu fishermen. [16] The "pawns" are the local fishermen who support the actions with their local knowledge. The "bishops" are the former militiamen, they are competent in the use of weapons, and they execute the attacks or defend the bases on the mainland. The "chess masters" are the engineering and planning experts who gain dates from the databases of freighter companies and work out the details of the missions. [16] The pirates have moles in positions of authority who receive money from the ransom in lieu of the valuable pieces of information leaked by them. [17]

5 Occasionally we can detect some similarities between piracy and terrorist acts committed on sea in the light of perpetration. Politically motivated rebellions and uprisings also take after the crime of piracy.

6 Which was followed by a serious poverty demanding at least 200 thousand lives.

The clans have built up quite a developed heartland from the rich spoils; they are continuously upgrading their technology. [17] It is a curiosity that a pirate "stock exchange" runs, where the spoils are sold and people can make investments, which can be useful in pirate activity. Once a woman who received an RPG rocket launcher as alimony, sold it on the market for 75 thousand dollars. [17]

Pirate activity seriously redounds in the GDP of Somalia.[7] This is so true that – in the African context – piracy provides a relatively high standard of living not only for the pirates and warlords of the clans, but for the average people as well. [18] The most seductive fact for pirates is not really the value of the stolen goods, but rather the ransom given in return for the hostages. [16] Members of an attacking team could earn around 30 thousand dollars per capita. Much more money goes into the pocket of the leaders of clans. Participants of a particular attack could receive even 30% of the ransom, the militia defending the pirate haunts take 10% in return for their service, and last but not least, the local community also obtain in planty (around 20%). 20% is the reinvestment in the "pirate venture". 10% goes into the pocket of Al-Shabab, which refers to a real connection between pirate companies and terrorist organizations.[8] [16] Not only the Somali Al-Shabab or Hyzbul Islam are related to pirate clans, but probably Al-Kaida is also connected to them and receives financial support for the attacks on different targets. [19]

The relationship between pirates and terrorists is pretty paradoxical. As I have already mentioned the Islamic courts declared jihad against sea robbers and extinguished piracy in the past. Nevertheless, in the last couple of years, after the fall of the Islamist government, the activity has been renewed and is flourishing; although Sharia still prohibits abduction, punishing hostages or piracy itself. [20] Another interesting fact is that, usually, pirates do not follow any kind of religious or political ideology; but instead, making profit is their main purpose. This fact also emphasises the difference between terrorists and pirates. [20]

The main seat of pirates used to be the port of Mogadishu and its outskirts. Later on, it moved to Puntland, then further into the region of the Gulf of Aden. Puntland[9] is a separatist province of the collapsed Somalia, which was created by the Clan Harti in 1998. [21] The "province-state" has a relatively stable government. Puntland is the biggest citadel of piracy and receives the highest share in the profit of crimes committed on the seas.

Pirates try to explain and translate their acts into the language of international law. Apart from the fact that they consider their acts as a retributive mission against the Western and Asian ships fishing and polluting the sea in an unlawful way, Somalia does not acknowledge the effective and valid rules of the international law of the sea. Instead, they presume that the zone of influence spreads out 200 sea miles from the coastline, the ships sailing across the Somali waters have to pay "duty", but they regularly miss it, so the supposed premise of the pirates is completely legal.[10] The name of the two biggest clans shows us how seriously this previous statement was meant by pirates. The biggest company is the *Somali Marines* from Eyl, Puntland, which refers to itself as the *"Defenders of Somali Territorial Waters."* The

7 Apart from the oil, because the exploitation is slowed due to the civil war, and also apart from fishing, where the biggest importer is Yemen. Since the beginning of the civil war, fish export has drastically fallen.

8 The Muslim political movements acted against pirate activity, they captured some pirate haunts, but they were not able to cross the boarders of Puntland. Today they live side by side respecting each other.

9 The territory of the province is 212 thousand km^2, the population is almost 4 million.

10 The second approach is shown in the film drama Captain Phillips directed by Paul Greengrass in 2013, which is based on a true story.

other grouping is the *"National Volunteer Coast Guard",* the headquarter of which is located in Kismayoo. [21] On the level of communication, the government of Puntland promised that it would act hard against piracy, but they would never abandon a profit-productive "branch". Moreover, warlords from Puntland also maintain pirate-militias. [21] In addition, foreign "financial circles", "business groups", and even some members of the federal government benefit from the "profit". [19]

The *social influence* of piracy is reflected in the fact that the local residents esteem pirates as defenders of their country confronting the exploiting powers. The clans pay a dividend to the local leaders and finance communal investments. [22] From the rest of the money, there is plenty for alcoholic beverages and for a special narcotic herb, the *khat* chewed during action.

Defense beyond Law

In the history of piracy, the world-powers used the devices of diplomacy, paid protection rackets, made contracts, and as a final solution used armed force against pirates. Until the end of 18th century, the Knights of Saint John of Malta kept up a convoy service against the corsairs of North Africa. [23]

Since Somalia is not able to counteract effectively, the most obvious solution is that the freighters prepare themselves against pirate attacks. [24] It is hard to climb up the higher stern or sidewall and the captains can create waves with maneuvers to capsize the boats. They can install water cannons, barbed wire, electric fences, or they can employ armed guards. [22] The expenses for defenses make freight transport more expensive. According to some estimations, pirate activity causes 18 billion dollars loss each year for the world economy. [20]

It would be a more effective method if the vessels formed convoys escorted by warships. The ships are in continuous connection with American, British or other naval forces. Besides, there is a protocol in case of a pirate attack, which has to be followed in danger. The increase of military presence is the best practice, but it is only enough to heal the symptoms, it does not solve the whole problem.

Somali territorial waters are primarily used by European, American and Asian traders, thus *NATO* and the *European Union* have a prominent role. NATO has two permanent maritime groups, one of which is the so-called *Standing NATO Maritime Group*. [25]

NATO actions are granted not only on a conceptual level, but also have won political formulation during the informal meeting of defense ministers in Budapest in 2008, where the naval units of NATO got a mandate for the protection of vessels of the *World Food Programme*. [25] Man-of-wars of the participant states provide not only armed attendance, but also, they execute patrol on the neighbouring waters with Somalia as well. [25]

As the biggest supporter of Somalia, the European Union would like to ensure the safe and certain arrival of its aids; therefore, they established the *military coordination action programme (NAVCO)* [26] whereby naval units were installed on the threatened waters of the Indian Ocean and the Gulf of Aden. [25] In November 2008, a joint military operation was launched by the EU, under the codename *Atalanta EUNAVFOR Somalia*, [27] with more than 20 ships and fighters and a 1 800 member crew. [25] Although Hungary does not have sea forces, it delegated an IT non-commissioned officer assisting in the mapping of attacks.[11] [18]

11 Four other soldiers take part in the training of Somali governmental forces. The Training Mission was started by the NATO. [18: 34]

There is no worthwhile international military co-operation without Americans, who created three several battle groups for the defense of shipping routes. The mission covers 1.1 million square miles; its duty is to act against the bandits, to facilitate global maritime security and the freedom of navigation. [25]

The pirate activity substantially aims at the most profitable oil carriers. After hijacking the biggest Saudi tanker, one of the biggest oil exporting countries, Saudi Arabia joined the international co-operation. The Saudi foreign secretary labeled the pirates as *"disease need to be exterminate."* [19]

Legal Defense

The United Nations labeled Somali piracy as a threat against international peace and security, thus based on universal jurisdiction, all of the states are entitled and obliged to act against piracy, as long as its legal system defines and punishes the crime of piracy. [8] These crimes are so very dangerous for society that it endangers the whole international community. Even if the acting state has no territorial, national or any other special jurisdiction, it is allowed to act against these offences. [28] *Piracy and war crimes* traditionally belong to this category of *delicts.* [29]

The text of UNCLOS Treaty disposes that states are allowed to act based on universal jurisdiction even on the high seas or the territory of the state, but this second option raises the question of harming sovereignty. It is a fundamental principle that every state has the right to proceed against its own citizens on its own territory according to its own law in the light of its own criminal law supremacy. Every other state is obliged to avoid breaching these rights. Chasing and arresting Somali pirates on the coastal waters or on the main land certainly infringes the sovereignty of Somalia unless it consented to the infringement. The principle *volenti non-fit injuria* excludes illegality. The permission has to be based on free will and it has to be given by a governmental body or person. [30] Somalia has made this kind of declaration when entering into agreement with the EU.

Aggression against pirates raises further questions as well, because by the 21st century the prohibition of the use of armed forces is generally accepted. Since piracy falls under special judgment, the use of armed forces is permitted. [8] Especially since the UN Security Council declared pirate activity as a threat against international peace and security, it makes Chapter VII Article 43 of UN Charter applicable. This rule gives permission to use armed force or duress, but it is necessary to have the acceptance of the "host" state, which was given by Somalia in 2008. [25]

In the light of the above mentioned, international missions are allowed to intrude into Somali territorial waters, patrol and halt the suspicious watercraft and arrest their crews. [25] The procedure has to be necessary and proportional. [8] Combat against sea robbers mean *law enforcement* operations, where the classic rules of martial law are not applied, because these are not military actions among states. [9] This statement is only true for the cases occurring on international waters, on the main land the Geneva law could be applied, because the pursuer state is performing operations on the territory of another state, which refers to Article 2 of 1949 Geneva Convention. [8]

Questions of Responsibility

Criminal jurisdiction could be based on the *territorial principle*, which is the place where the crime was committed. [29] According to the *principle of nationality (or active-personality/ nationality of offender's principle)* [29] against a Somali citizen, only Somalia is allowed to act. However, among pirates we can not only find Somali thieves, but for example Pakistani ones as well and the Somali authorities do not have the assertion of a right to act against them.

The nationality of the ship, which was the *instrument of the crime*, also creates a problem. Article 91 of UNCLOS Treaty says *"ships have the nationality of the state whose flag they are entitled to fly. There must exist a genuine link between the state and the ship."* [29] According to the *passive personality principle*, a state has jurisdiction over the offender of those crimes, which impact the citizens of the state in the past or in the future. [29] The *principle of protection* authorizes the state to exercise jurisdiction over those offenders who are not the citizens of the state, but violate or endanger the security of the state with their action. [29]

Establishing *individual criminal responsibility* of pirates would be the duty of Somalia in the first place, but at the moment, a central power able to bring the offenders to justice, does not exist. [25]

In most cases, the countries giving effect to the duress on pirates do not want to account for pirate activities before their own courts, so the international community sought a kind of solution, which could give effect to criminal sanctions.

The *Djibouti Code of Conduct*, which was accepted in January 2009, prescribed joint actions, information exchange for the participant Eastern African countries. Moreover, it consists of rules of investigation and prosecution. Furthermore, the Code prescribes that the plaintiffs of pirate acts have to be accommodated properly and the authorities have to promote their return home as soon as possible. [25] For financing the commitments, the participant states established funds and assistant offices. [8]

The most dedicated to calling pirates to account is Kenya. Under the agreements with the United States, Great Britain, the European Union and China, Kenya takes over and conducts procedures against the pirates who were arrested by international joint operations, in turn for the partners assisting in the modernization of the prison network and the administration of justice. [8]

Besides Kenya, the Seychelles made similar contracts, but the insular state asked a higher price of the EU than Kenya – complete financial, personal, logistic and infrastructural support. [25]

Passing over pirates could be operable, and indeed, it is operable, because numerous sentences have already been adjudicated, which are currently under execution, even if they are raising doubts. According to the UNCLOS Treaty, if a warship captures a pirate, the flag state has jurisdiction over the arrested sea robber. On the other hand, there is no *expressis verbis* prohibition in UNCLOS Treaty regarding the deliverance of pirates to a third state; however, it does not permit it pronouncedly. [8]

Is Somalia Responsible for the Pirate Activity?

During the codification of the international legal responsibility, the *International Law Commission of the United Nations* endeavored to pass comprehensive regulations among the international community. The complete draft elaborated the consuetudinary standards, and although it does not have binding force as a treaty, it does have a basis of reference without doubt. [30] According to the *Crawford draft,* any unlawful or wrongful act of a state raises international legal responsibility if it is *imputable* and *injures the international commitments of the state.* [30] The injury could be the harm of *jus cogens norm* itself or could offend the interests of another states, business organizations or individuals.

The offending act could be *proactive behavior,* or *omission*. Because piracy is counteractive to international law, the state where such crimes are regularly committed has to do everything possible in order to stop criminal activity. Thus Somalia has legislative and law enforcement liabilities, at least theoretically. Practically there are three options to establish responsibility for Somalia for pirate activity. Firstly, if the infringement is imputable to the legislative power: for example, if the Federal Parliament of Somalia made a law, which would legitimate piracy that would be against international legal norms. On this ground – apart from the fact that Somalia has a different legal opinion about the maritime zones belonging to its sovereignty — the federal government, recognized by the international community, shall not be responsible, because as far as their power allows it, they try to prosecute the crimes committed in the country or on seas.

The case of the Somali coast guard is even more interesting. This armed force, which is responsible for the security of the Somali coast, belongs to the Somali *executive power*, and its function is to act against sea bandits. But it hardly does its duty *(infringement carried out by omission),* moreover, it is involved in pirate actions on the most infected coast lines, which would establish state responsibility, not to mention that pirate clans refer to themselves as voluntary coast guards, but of course, this does not mean that they are acting in the name of the federal government.

Is Somalia a state at all?

According to the current status of international law, the answer to this question is definitely "yes". Somalia has all the criteria in order to be recognized as a state. It has *territory,* [30] which is suitable for human life, it has *population,* [30] which means that there are people living there permanently, and they are linked to the state with a nationality or another status. In addition, it is necessary to have an *independent governing power*. [30] A plus condition is *recognition,* which might be constitutive or declarative. [29] Recognition is an objective factual situation (the fulfillment of the three conditions at the same time); it is when the international community acknowledges the three criteria together, of course spiced with a huge amount of the great power's politics.

But the existence of a state is not only a moment fixed by the law. It has not only criteria, but periods of existence as well. A state can come into existence and also, it may cease to exist. The cessation of a state has also different types, from fusion to division. [29] In my opinion, in the case of Somalia, we are talking about this latter version. After becoming

independent of the colonial rule and being recognised by the international community, the former British and Italian Somaliland simply *divided among the clans of the Somali nation*, and the international community has not responded to this on the level of law, yet. Therefore, Somalia, as it is legaly recognised today, is only a *legal fiction*, in fact, it does not exist. [30] Therefore, because it is a failed state, which means not able to fulfill the requirements of the minimum criteria of a state, Somalia shall not be responsible for the violation of the international law.

Considering Somalia as a *failed state* makes it harder to challenge Somalia for the violation of international law, and makes it easier to understand the whole "pirate story" and the "state" system. Somalia was the first with 113.9 points with *Very High Alert* on the list of *Failed States Index* [31] published by *Found for Peace* in 2013.[12] Somalia is the most referred to example of a failed state, because it fulfills all the conditions. Its government is not able to control the whole territory and maintain the administrational system efficiently, the majority of the population does not recognize its legitimacy. It is incapable of providing the basic public needs, and it does not have singular power over the enforcement agencies. [33]

In order to reach the phase of failure, the *existence of internal war and insurgency*, ethnic or religious conflicts is essential. After the collapse of the administration system and the infrastructure, the remains of the state have to face humanitarian disaster, the flood of migrants and poverty and most importantly, total economic deterioration. [34] The failure of a state has different phases, too. However, getting from a stable or a strong phase into a failed status also has an effect on international peace and security.

Somalia, as a recognized country by the international community, practically does not exist. The country has been replaced by clans, which have created and consolidated their own territories (Puntland, Somaliland, etc.) after the long civil war. These regions are practicing authority as a *quasi state*. Due to the militias, relative peace reigns, but for the time being they have not succeeded in building an operable, self-supporting state. There is a chance for Somalia to form a federal state in the future, which is the purpose of the independent regions in the long run, but there is no agreement as to the solution.

The deeper and deeper integration of piracy into the society, the tough struggle of Al-Shabab, the immeasurable corruption and the lack of raw materials make the situation more and more complicated.

Summary

Piracy is a phenomenon that can be proved since man first ventured onto the seas to sail. There has been no success in finding an effective remedy. On the coasts of the high seas, there will always be underdeveloped regions, which will not be able to maintain public order and control their territories and population properly. The international crime of piracy is well defined, acting against it is assured. As a failed state and the biggest citadel of piracy and in the long run Somalia will not be able to prevent piracy or call the pirates to account, even if its international legal responsibility was established.

12 Hungary is on the 141th place with 47.6 points in the Stable category.

References

[1] HÓMAN B., SZEKFŰ Gy., KERÉNYI K.: *Egyetemes történet*. Budapest: Magyar Szemle Társaság, 1935–1937. (Hóman Bálint munkái, digitális kiadás: Arcanum Adatbázis Kft., 2003.) http://mek.oszk.hu/07100/07139/html/index.html (downloaded: 25 01 2014)

[2] CSATLÓS E.: A római kor és a középkor tengerjoga. *De iurisprudentia et iure publico*, VI 3 (2012).

[3] SOUZA, P. de: *Piracy in the Graeco-Roman World*. Cambridge: Cambridge University Press, 1999. http://catdir.loc.gov/catdir/samples/cam032/99013647.pdf (downloaded: 25 01 2014)

[4] SOUZA, P. de: Ancient Rome and the Pirates. *History Today*, 51 7 (2007). www.historytoday.com/philip-souza/ancient-rome-and-pirates (downloaded: 25 01 2014)

[5] MARITIME CONNECTOR: *History of Piracy*. http://maritime-connector.com/wiki/history-of-piracy (downloaded: 25 01 2014)

[6] NATIONAL MUSEUM OF THE ROYAL NAVY: *Piracy – A Brief History of Piracy.* Portsmouth, 2002. www.royalnavalmuseum.org/info_sheets_piracy.htm (downloaded: 25 01 2014)

[7] *United Nations Convention on the Law of the Sea – UNCLOS – 1982.* www.un.org/depts/los/convention_agreements/convention_overview_convention.htm (downloaded: 25 01 2014)

[8] LATTMANN T.: A szomáliai kalózkodás kérdése a nemzetközi jogban és azon túl. *Nemzet és Biztonság,* 9 (2010), 66–75.

[9] VARGA A. F.: Kísérletek és eredmények a kalózkodás bűncselekményének definiálásában. *Hadtudományi Szemle*, VI 1 (2011), 30–43.

[10] RUBIN, A. P.: *The Law of Piracy*. Newport, Rhode Island: Naval War College Press, 1988.

[11] KAJTÁR I.: C.S.S. „Alabama" – Adalék a tengeri háborúk jogtörténetéhez. *Jura*, VI 1 (1997), 14–20.

[12] *Convention on the High Seas – 1958.* https://treaties.un.org/Pages/ViewDetails.aspx?src=TREATY&mtdsg_no=XXI-2&chapter=21&lang=en (downloaded: 25 01 2014)

[13] A MAGYAR KÖZTÁRSASÁG KORMÁNYA: *H/5558. számú országgyűlési határozati javaslat az Egyesült Nemzetek Montego Bayben, 1982. december 10-én aláírásra megnyitott Tengerjogi Egyezménye megerősítéséről, valamint XI. részének végrehajtásával kapcsolatos, New Yorkban, 1994. július 24-én aláírásra megnyitott Megállapodáshoz történő csatlakozásról.* www.parlament.hu/irom36/5558/5558.htm (downloaded: 17 02 2014)

[14] McDANIEL, M. S.: *The Serious Problem of Modern High Seas Piracy*. Chicago, 2000. (Presentation). www.cargolaw.com/presentations_pirates.html#type_attacks (downloaded 02 02 2014)

[15] GRAF, A.: Countering Piracy and Maritime Terrorism in South East Asia and off the Horn of Africa. *PiraT-Working Papers on Maritime Security,* 5 (2011).

[16] BESENYŐ J., KISS Á. P.: Kelet Afrika tengeri farkasai – A szomáliai kalózok fénykora. *Afrika Tanulmányok,* 3–4 (2009), 46–58.

[17] MARSAI V.: A szomáliai kalózok és az EU Atalanta-missziója. *Nemzet és biztonság*, IV 4 (2011), 66–76.

[18] BESENYŐ J.: A szomáliai stabilitás és a nemzetközi közösség. *Afrika Tanulmányok*, V 1 (2011), 29–41.

[19] BESENYŐ J.: A szomáliai kalózkérdés: Milyen érdekek mentén? *Mediterrán Világ*, 13 (2010), 43–79.
[20] KIS Á. P., BESENYŐ J., RESPERGER I.: *Országismertető Szomália*. Második, bővített kiadás. Budapest: Honvéd Vezérkar Tudományos Kutatóhely és MH GEOSZ, 2014.
[21] RAVASZ Á.: Kalózkodás Szomáliában: Puntland és a bukott államiság. *Kül-Világ*, V 3–4 (2008), 86–104.
[22] FÁBIÁN É.: A francia haditengerészet Afrika szarvának kalózvizein. *Nemzet és biztonság*, IV 1 (2011), 18–27.
[23] KAJTÁR I.: A Mediterránum tengerjogának történetéhez. *Jogtörténeti szemle*, VI 4 (2004), 71–77.
[24] INTERNATIONAL MARITIME ORGANIZATION: Ship Protection Measures. (Section 8.) In. *Piracy and Armed Robbery against Ships in Waters off the Coast of Somalia – Best Management Practices for Protection against Somalia Based Piracy.* London, 2011. www.imo.org/MediaCentre/HotTopics/piracy/Documents/1339.pdf (downloaded 17 02 2014)
[25] VARGA A. F.: Nemzetközi küzdelem a szomáliai kalózkodás ellen. *Hadtudomány*, Elektronikus szám (2011). www.mhtt.eu/hadtudomany/2011/2011_elektronikus/2011_e_14.pdf (downloaded 17 02 2014)
[26] COUNCIL OF THE EUROPEAN UNION: Council Joint Action 2008/749/CFSP of 19 September 2008 on the European Union Military Coordination Action in Support of UN Security Council Resolution 1816. (EU NAVCO). *Official Journal of the European Union*, L 252/39 (2008). http://eur-lex.europa.eu/legal-content/EN/TXT/ HTML/?uri=CELEX:32008E0749&from=EN (downloaded 17 02 2014)
[27] COUNCIL OF THE EUROPEAN UNION: Council Joint Action 2008/851/CFSP of 10 November 2008 on a European Union Military Operation to Contribute to the Deterrence, Prevention and Repression of Acts of Piracy and Armed Robbery off the Somali Coast. *Official Journal of the European Union,* L 301/33 (2008). http://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32008E0851&from=EN (downloaded 17 02 2014)
[28] WISE, E. M., PODGOR, E. S., CLARK, R. S.: *International Criminal Law: Cases and Materials*. Second Edition, New York City: LexisNexis, 2004.
[29] SHAW, M. N.: *Nemzetközi jog*. Budapest: Complex Publisher, 2008.
[30] KOVÁCS P.: *Nemzetközi közjog*. Budapest: Osiris Publisher, 2011.
[31] *Failed States Index IX2013*. http://library.fundforpeace.org/library/cfsir1306-failedstatesindex2013-06l.pdf (downloaded: 08 02 2014)
[33] WAGNER P.: *A „bukott államiság" és Afganisztán (Holtzer Lóránt emlékére)*. *Kül-Világ,* III 2 (2006), 37–54.
[34] JAMAL, A. R.: *Identifying Causes of State failure: The Case of Somalia. Failed States in Sub-Saharan Africa*. Konstanz: Universität Konstanz Politik- und Verwaltungswissenschaften, 2013.

AARMS Vol. 14, No. 2 (2015) 187–200. **10.32565/aarms.2015.2.5**

Selection of Authentication Systems for Hungarian Health Care, Based on Physiological Study

Part I.

The Biometric Systems

BARÁTH Artur[1]

The actuality of the topic indicated in the title comes from more and more events where the verification of identity might be required. Health-care is an important part of the critical national infrastructures. A primary task of the protection of critical information infrastructures consists in the access control of data managed by IT system where the identity-authentication forms an important part. For this reason, the technologies considered suitable for use and their supporting means are listed.

After examining and comparing the relevant parameters, the optimum solution for the authentication procedures in the Hungarian health care system is specified.

First of all, the modern and efficient biometric identification processes are examined; however, the possession-based Radio Frequency Identification (RFID), as an additional system, is also studied. Due to the well-known problems of passwords and chip cards, they were deliberately omitted during the system planning and efforts were made to exclude the human factors from the planned authentication system as much as possible.

Keywords*: authentication, biometry, RFID, health care, human factors*

Introduction

This study makes an effort to find a solution for identification and access management problems of the health care institutions in Hungary. As is generally known, the above indicated field faces serious problems in data security and access authorization issues that need to be remedied as soon as possible. For this, a technology that offers a proper level of protection and the introduction and operation of which fail to carry glaringly high costs is required; a further important aspects is that the authentication problems caused by people are excluded as far as possible.

The work is aimed at selecting the system from among those available that are best suited for the purpose and condition defined above. To achieve this, the procedures worth considering are listed, their properties and operation are described; finally, based on their examination and a summation of the results, the most suitable one is selected.

1 e-mail: magicman@mailbox.hu

The Importance of Data Protection and Interpretation of Related Concepts

In the 21st century, great value is attached to the protection of personal data. It cannot be accepted that our personal data is stolen by unauthorized persons. The professionals engaged in the protection of personal data should not turn a blind eye to the explosion-like growth in the number of biometric devices. The development and increasing use of such technologies relying on authentication identification needs special attention. The significant novelty in this identification system is that they can be compared not only by people but also by IT devices developed for this purpose. This service requires the cooperation of the identified person, without any other persons, as this process takes place automatically. The specific characteristics of the given person are programmed into the device that will be able to recognize and identify the person in the future, based on a chip or portrait.

In my opinion, the protection of our prized possessions is indispensable today, whether it be our data stored in the computer or physical devices. There are a lot of possibilities to protect and delimit them from certain persons. Within these, we can talk about different levels of protection. There are safe or less safe solutions; the technology to be used is determined by the importance of the data or object to be protected. First, I want to deal with the means serving the protection of physical assets, although these methods are also used to protect other data todays. This will be described later.

It is necessary to clarify certain indispensable definitions before listing possibilities.

Authentication

Authentication is a security service. It is aimed at identifying the origin of a message or request in a reliable way. The classic method is how people identify themselves to the computer (and how computers also identify each other), it consists of entering passwords. Essentially, the authentication serves the verification of individual authority and identification of the user. In other words, it determines which users (e.g. employees) have access to what information and how far their competence extends. [1] The ratios of "Security level/cost" are grouped into categories as follows:

- Low-end (low-end, low-cost): it is aimed at creating an appropriate security level by using solutions as cheap and as economical as possible.
- High-end (top of the line): it is used in systems that require a high level of security. The disadvantage is that it is considered to be an expensive solution.

Biometrics

Biometrics consists of the identification of people based on one or more essential physical or behavioral pattern. Biometrics is used as a method of identity management and access control, primarily in the field of IT. [1] Biometrics can be categorized into two main groups:

- Physiological-biometrics: perhaps, it is this that is considered to be the most popular biometric means. They are connected with the form of individual parts of the human body in all cases. These include: examination of face (geometry, skin design, heat

map, smile, dynamic characteristics), lip (figure, imprint, movement), anthropometry (physical dimensions such as height, head length, shoulder width, etc.), bioelectric field, skin spectroscopy,[2] DNA/DNS, teeth (bite, teeth X-ray photo), ear (geometry, geometry of auricular canal, echo), iris, retina, hand geometry (hands, fingers), nail (nail bed), heart rate, odor/aroma, soles, toes (ridge, phalanges, wrinkle, wrinkle joints, topography), vascular (on fingers, hands, palms, wrists).

- Behavioral biometrics: a less popular biometric method. Its essence is to examine the movements performed by the human body and behavior forms expressed by the individual, e.g. use of keyboard, sound (speech, laughter), gait, handwriting (dynamic signature, signature sound) and dynamics of using the mouse.

The identification systems are used to establish authority. This may be done by entering a given computer system, or accessing some critical data. The entry or access is always managed by some system that may be human resource-based or electronic, or a still more IT based one. In the following, the IT line will be followed. Obviously, there are several possible solutions to achieve our goals; these shall be arranged on the basis of several criteria in order to arrive at a correct conclusion at the end of the studies. In almost all cases, personal identification is used for releasing the protection. To do this, three methods are used apart from the vanishingly small exceptions: [2]

- knowledge-based;
- held-based object;
- biometric-feature based.

The safety in this order increases from top to bottom. As is known, there is no perfect security, but the use of these facilities enables obtaining a level of safety that is acceptable in respect to both economy and safety. The above three methods are described in detail as follows:

Knowledge-based

Some sort of user name, password, PIN combination are used to identify ourselves. A better known and more frequent solution is the use of a password that, in any case, shall include at least three ones of those listed below:

- small letters and capitals;
- number;
- special characters.

In addition, the length of the password is also an important factor, as the more characters, the more time and work are required for decoding. The minimum length generally accepted in professional circles amounts to 16 characters.

In the PIN (not the same PIN code, that is used by credit cards) identification, the secret is usually a four, six or eight-digit decimal number, possibly a five-character alphanumeric sequence. The four-character PIN represents 102, the six-character 106, and the eight-character 108 variants. The alphanumeric string is usually based on the 26 letters of the English alphabet and ten numbers, represents a five-character password with (26+10)5, that is 60,466,176 variations. From among the three identification methods, it is the last one, in that it is the feeblest.

2 Solution method of imaging technology.

Held object-based

In this case, an object can be kept to identify ourselves. In most cases, this is a chip card, less often an RFID chip; however, a more and more popular method is to insert an implant under the skin, and it may also be a device of a RFID system. [3]

A chip card is a card kept by us and presented to the reader when we want to identify ourselves. The reader identifies the card by using radio frequency technology. A 112-bit key is used nowadays with 3-DES encryption. This is included in the chip card. The system compares it to the keys stored in the SQL database; thus, the identification takes place. A weak point of such systems is the link between the reader and the database as it is susceptible to be successfully attacked by certain viruses (SQL injection).[3]

- There are many types of chip-cards and systems using this method:
- operating frequency (125 kHz, 13.56 MHz, 430 MHz, 862–956 MHz (UHF), 2.45 GHz, etc.;
- reading range (<0.05 m, 0.05–0.5 m, 0.5–2 m, >2 m);
- power supply mode (passive, active and partially passive);
- operations allowed (factory programmed, writable once, rewritable, writable/readable).

Security of chips:

- Open: simple fixed code fix, readable and interpretable (or unstructured) data content.
- Encoded: readable, cannot be interpreted, but can be copied.
- Data access protected (password protected): access using password; unable being copied in itself.
- Encrypted: it is the most secure at present.

An improvement of identification by using chip-card is to insert an implant consisting of a chip in the human body. It is the Very-Chip in the United States of America that is the only one allowed to be used as a human RFID implant. The method of identification is the same as in the case of having a card with ourselves. This includes a 16-character length code for the purpose of identification. [3]

3 http://en.wikipedia.org/wiki/SQL_injection (downloaded: 14 03 2015)

Biometric Systems – General

In respect to security, the biometric identifiers also have some kind of order shown in Figure 1.

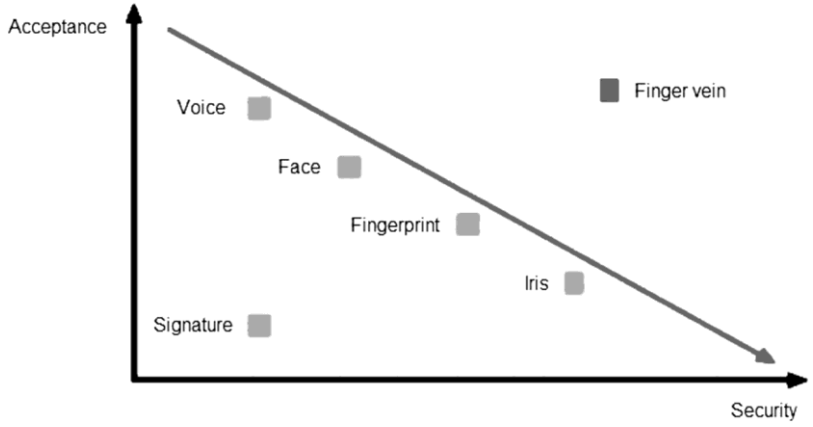


Figure 1. Security of finger vein. (Made by the author.)

The most prominent ones of these are the fingerprint and iris identification. Consequently, it is these that shall be better examined in the following. First, two questions are answered that come up in nearly all cases when using biometric methods and provide proof that the world is indeed going in this direction.

The replies to the two questions below further strengthen the legitimacy of biometric identification: Is it true whether biometric identification will become the basis of identity verification in the future?

Yes, because the traditional identification methods – such as passwords or various identity cards – are based on "what you know" or "what you have". On the other hand, the biometric identification methods are based on "what you are"; on anatomical features (face, fingerprint or iris recognition), or behavioral patterns (signatures or gate). These procedures are much safer than traditional methods. In contrast to the passwords and identification documents or cards, the biometric methods are much more difficult to be deceived by guessing, distribution, lending, copying or counterfeiting.

Why will the biometric methods spread soon in the very near future?

Still, as both the demand and supply are given. On the demand side there is a growing demand for safety and a growing fear of counterfeiting. The states want to know who crosses their borders, to which they pay social benefits. The companies want to control who enters their premises, who are allowed to access their webpages and databases. And what is of key importance: the social acceptance of technology is also growing slowly; more and more use biometric methods in order to ensure identity and to prevent theft, to protect their laptops or phones or in case of payment. On the supply side, small-size, low-cost and fast sensors were able to be developed as a result of the sudden technologic development in the recent past that can be fully automated and are suitable to be used for "real time" imaging. In the following, the functions of the individual procedures are described.

These identification systems can be defined as the totality of processes and devices that uses the physical properties of persons that can be measured by using some kind of technique with the view of determining personal identity or identification. During identification, a comparison of the body of the individual takes place. In this case, two different answers are expected of the system: yes or no. During personal identification, the system recognizes the person by distinguishing its data from stored ones of others. In the first case, two data are compared, while in the latter, one item of data is compared with other stored data.

The more and more and widening use of biometric systems is justified by the special features of biometric data. Data of people can be considered to be universal in the sense that everybody is in possession of such data. The best solution for identification or distinction is represented by DNA-samples, fingerprints and retinal images; however, the geometry of face, the hand and the voice can also be used for this purpose. Everybody has physical characteristics of this kind. These data are individual and different, so the conclusion that there are no two ones alike can be drawn. In time, these data suffer no change. Of course it is not true for all biometric data. Taking these identification systems in a broader sense, it is not only the physical properties of individuals that can be taken into account, but the behavior as well. Factors describing the behavior include e.g. the signature and movement of persons. Part of biometric identification systems combine the use of certain data types e.g. fingerprint and voice identification, while others use methods that take the knowledge of individuals into account e.g. whether they know their password or PIN code associated with their bank card etc. [4]

Basic requirements of these systems are:

- they shall be able to be interpreted and used by everybody;
- uniqueness (no two persons meet the same conditions);
- time stability (static characteristics for long time);
- measurability (preferably high-performance, fast);
- difficulty being deceived.

Figure 2 shows the general block diagram of an average biometric test. In respect of function, the steps indicated are the minimum required ones that have to take place.

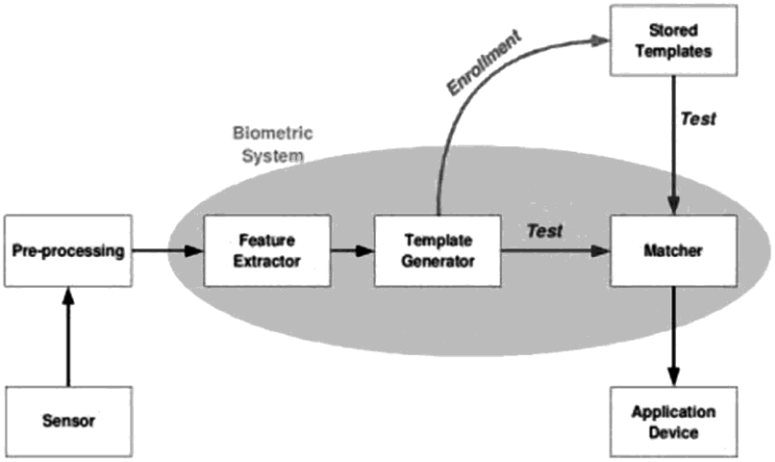


Figure 2. Biometric test's block diagram. [6]

In the block diagram, the operation of biometric identification can be followed. The sensor data are subject to pre-processing (e.g. noise filtering) and the unwanted data are filtered out of the data set (feature extraction). Based on the remaining data, a sample is made up that can be compared with the samples stored in the database; then, the results are forwarded for further processing (alarm, notification etc.). Figure 3 shows the elements that form a biometric identification system.

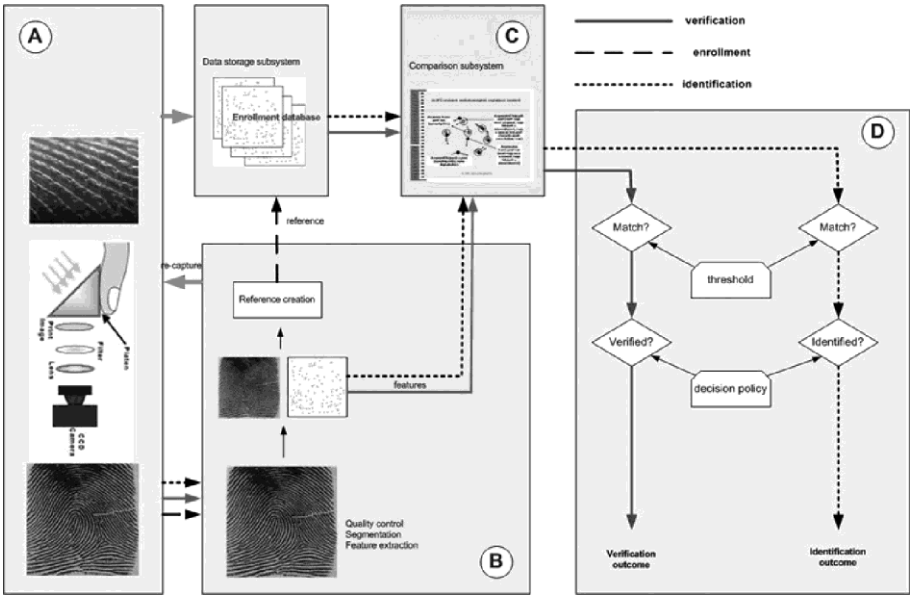


Figure 3. Elements and process of biometric identification.[4]

In general, a biometric identification system consists of four main parts. First, a sensor is required to record the biometric pattern – be it fingerprints, iris texture, handwriting, voice pattern, or whatever. Then, the specific features of the sample that facilitate the identification shall be selected; a database that includes examples characteristic to the individual biometric features is required. Finally, a device for comparison that compares the individual samples with those stored in the database is required. [5]

In terms of reliability, several studies have been done in the past that draw conclusions as follows.

The identification accuracy is determined by two indicators:

- False Accept Rate (FAR) – this index shows the frequency of admitting NON-authorized users to enter the system.
- False Reject Rate (FRR) – this index shows the frequency of rejecting authorized users to enter the system.

4 Source: unknown.

Table 1. Reliability comparison. (Made by the author.)

Face recognition	2,000:1
Voice identification	500:1
Fingertips identification	1,000,000:1
Iris scan	10,000,000:1
Retina scan	10,000,000:1

Table 1 shows a comparison of several FAR indices (how many false identification fails to a right one).

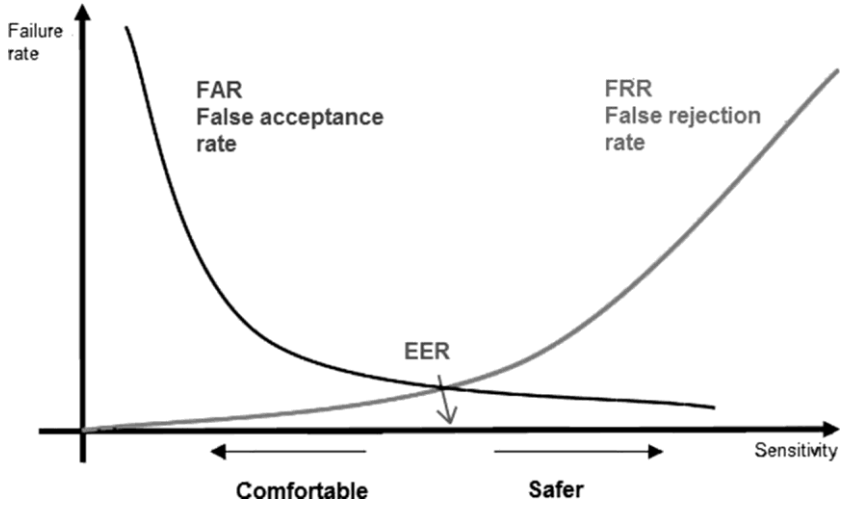


Figure 4. Equal Error Rate (EER) point declaration. (Made by the author.)

The FAR and FRR together represent two curves intersecting at a point. Figure 4 shows: this point is called EER. Table 2 shows a comparison of the relative accuracy of individual tests i.e. the EER values calculated with the FAR and FRR values.

Table 2. EER points list. (Made by the author.)

Face recognition	200:1
Voice identification	50:1
Fingertips identification	500:1
Iris scan	131,000:1
Retina scan	10,000,000+:1

The FRR value of modern fingerprint identification systems lies around 0.01 per cent (one error falls to ten thousand events); while their FAR value around 0.1 per cent (one error falls to a thousand events). Of course, the actual performance of a biometric system depends on several factors. In addition to the biometric characteristics, much depends on the sensitivity of the sensor, size of the database and a variety of other factors (temperature, humidity, indoor or outdoor use, and so on). Various aspects have importance in case of different uses. [4]

Identification of Fingerprints

The safe use of fingerprints in data protection is based on individuality and permanence. There are no two people, not even twins, with the same fingerprints. First, consider the two kinds of imprints classification structured by Henry (1900) and Galton (1892).

In the Henry structure, it is the numbers of ranges, while in the Galton structure the characteristic signs that are competent. Although the uniqueness or quality of fingerprints do not serve as a basis of Henry structure unlike ranges and unique properties. Therefore, to establish whether two fingertips are identical or not, an appropriate number of (privileged) points and/or properties should be located in the same place:

- Beginning and ending ridge;
- Bifurcation;
- Island;
- Enclosure.

This list has been expanded and refined over time. According to the proposal made by the American National Standards Institute it became as follows:

- Ending;
- Bifurcation;
- Mixing;
- Undefined.

The most common system today is the Minutiae-Coordinate model (used by the FBI). This system uses two properties i.e. ending and bifurcation for the identification. It interprets the points of properties as a coordinate each and uses them as a basis for displaying and verifying the identification.

The subsequent Sparrous model uses a structural representing mode for drawing the fingerprints, which made it possible to recognize the print even if it is distorted.

A complete fingerprint contains about 100 ridge bifurcations and endings i.e. so called minutiae points. The fingerprint identifiers compare 30 to 60 minutiae of the given pattern. The resulting image can also be used for identification either directly (global analysis) or by creating a fingerprint code based on the distance of minutiae points from each other or the angle formed by the connecting straight lines. The patterns formed may be a size between 100 and 1500 bytes, depending on the algorithm and the number of identified points; of course, they are also closely related to system reliability.

The Possible Methods of Touch Based Biometric Identification

Handprint identification: it is not as general a use in biometric identification; instead, it occurs primarily at the scenes of offences. In case of identification, the wrinkled location of main lines, the ridges and tissue samples on the palm are analyzed. Following a precise analysis, the information similar to the fingerprint sample is obtained.

The essence of hand geometry identification is that, while analyzing the samples taken of the form and surface of the hand, the length and width of fingers, the width of the hand as well as the scale of the palm and fingers are taken into account. Four positioning pins are used to place the palm in the same position for scanning. There are devices without positioning pins that examine particular specific values under microscope; systems of this kind are used e.g. in work time recording systems. Their advantage is the possibility of integrating with other systems.

Palm and finger identification is a relatively new method. Functionally the finger or palm is illuminated by infrared light that is reflected by the different tissues in different intensity due to differences in absorption. The light is absorbed by blood flowing in the veins much better that results in the exact mapping of veins. The advantage is that the operation is not affected by surface lesions of the skin and is nearly impossible to be deceived.

Fingerprint identification: the finger is placed on the scanner that, by measuring the surface temperature and moisture, verifies whether the finger is live. Then, the identification takes place.

From among the four possibilities described above, the fingerprint test system is selected. It is for reasons that the identification is intended to be used not only for access authorization but also for log-in control into computer systems. In such cases, the palm scanner raises very serious scruples in respect to practical applications. Fingerprint scanning is more appropriate even in respect to economy; in fact, purchase of physical devices represents lower costs. Last but not least, the selected procedure is more advantageous even in respect of hygiene.

According to the decision, the fingerprint test system is considered in more detail. A number of technologic solutions exist in fingerprint scanning systems. Certain identifiers are capable of examining whether the fingers belong to the hand of a living person; this is done by analyzing the temperature and humidity of fingers. The similarity between the devices consists in that the uniqueness of ridges on fingers is analyzed, stored, and finally, compared to specify that the ridges on the fingers are of the individual named and stored, and then it eventually finds the comparison identified. There are two versions of the technique, optical and non-optical methods.

Optical Devices

The fingerprint is mapped onto the surface of an image processor. The image processor is a Complementary Metal-Oxide Semiconductor (CMOS) or Charge-Coupled Device (CCD) element. LEDs are used to illuminate the ridges. Currently used types are as follows:

- Total reflection type: in this case the fingers are placed on the surface of a prism, and the picture obtained during lighting is mapped onto the surface of an image processing device.

- Diffraction type: similar to total reflection-procedure, except that the prism is replaced by a Fresnel lens.
- Chip sensor: a finger is placed onto the surface of the sensor and the signal obtained is transferred via fiber optic cable to the image processing device.
- Thermal analysis: in this case, the device shall not be touched, instead, the finger pulled away in front of it. This device reads the information broken up into slices to form the image. The built-in sensor detects the temperature difference between ridges. So far, few manufacturers have made an attempt in this field; yet, the part played by this device in the market is expected to rise and its current high price is expected to be reduced. It can also be used under extreme circumstances.

Non-Optical Devices

- Radio frequency signal is sent to the finger that reflects it to the receiving sensor. This procedure is also capable of forming a depth image of the fingers and the ridges.
- Principle of capacitive sensor: the finger is placed onto the surface of a sensor mounted with small capacitors. The sensor detects the capacity differences between grooves and ridges. This signal is converted into an electric signal for evaluation. The sensor consists of a conducting layer of lattice structure that generates an electromagnetic field over itself. Touching it with a finger means some of the charge is removed from the position in which it is capable of being sensed by the device.
- The sensor emits ultrasonic signals to the finger. The reflected wave signals are used to generate a depth image.
- Pressure sensor: a piezo-electric matrix is built-in below the sensor surface that detects the unevenness of the finger surface to generate an image.
- The E-field technology measures the electric field of the skin. It is suitable for everyday use and insensible to the quality of the fingerprint. This electronic reading creates an electric field between the finger and the semiconductor in contact, which takes up the grooving of the fingerprint. The resulting image is of high purity, three-dimensional, but small size.

The research and development projects affect all the elements of biometric identification systems, thus better and better, more reliable and faster sensors and algorithms are being made. One direction of development is to combine data from different, independent biometric sources, thereby reducing the error rate. The US-VISIT strategy with dual fingerprint verification is one of the possible forms of "multi-biometric" systems.

In light of the knowledge of the above let me state that the fingerprint can and should be introduced for the purpose of identification in the Hungarian health care system; still, a simpler system representing lower costs shall also be selected for areas where the high level of identification ensured by the fingerprint is not required.

In the future, the information supplied by fingerprints will be able to be completed with information supplied by face recognizing software. With this in view, the retina, iris, facial recognition systems are described in the following.

Possible Methods of Biometric Identification

The iris-recognition is made possible by the individual and time invariant pattern of the human iris. The video camera takes a picture of the eye, then following that, the pupil of the eye, eyelid and eyelashes are cut from the picture; the field of the iris is removed. The information included in the unique iris pattern is recorded in digital form. The identification requires the sequence of bits to be compared.

Iris identification is based on the iris of the eye. In this device, the possibility of errors is almost zero. The iris of the eye includes 400 different identification points. The iris never suffers any changes during life. The chance that two people have exactly the same irises is excluded. During identification, the relevant device examines the visible and invisible elements of the iris. The radial pattern of the iris with circles, ditches and crown belongs to the elements visible to the naked eye. The retina-membrane veins become visible when reading in infrared light. During identification, the pupillary reflexes can also be observed, thus the abuses of contact lenses can be excluded. The scanner of iris identification system converts the image of the iris together with all its characteristics (dimples, circles, ditches, crown, and fabric fibers) that make the eye unique into a three-dimensional contour map. After digitizing, the information thus obtained it forms a code of exactly 2,048-digits. This will be compared later on with those stored in the database. Two methods can be used during the examination. Based on taking photographs, two types of reading can be distinguished i.e. active and passive. The active reading requires the active involvement of the user as he/she shall keep his/her eye at a distance of 0.15 to 0.35 m from the camera. In contrast, the passive procedure is more pleasant in respect to the users; in this case the system first determines the position of the eyes by means of a wide-angle camera and focuses another camera on them to perform the reading at a distance of as much as 0.3 to –1 m. In both cases, the reading takes about 1 to 2 seconds, same as the time required for fingerprint identification. Unfortunately, the active devices can raise hygiene problems; while the disadvantage of passive devices is that they require highly sophisticated technical implementation, so that their price is very high.

Retina identification: the retina at the back of the eye is illuminated with infrared light. Its function is similar to the identification systems of finger and palm. The reflected infrared beams are absorbed to a different extent, so they are able to draw the fundus vasculature. Direct contact with the reader shall be established; therefore, there is a high risk of infection. This technology is very rarely used; usually in case of systems requiring high security.

Face recognition in visible light: Face recognition – associated with sample identification procedures – can serve as a good means of identification. Currently, it is used in special cases. It is unable to be used for distinction of perfectly identical twins.

Face thermogram: A photo made by an infrared camera that shows the thermal pattern of the face. The image is unique and combined with a pattern identification algorithm of high complexity – that verifies the relative temperature differences on the face – it offers a technique that is independent of age and health, but the temperature of the body as well. By means of taking nineteen thousand "data points", the method offers an exceptionally safe procedure provided that the costs of process are reduced to an acceptable level. It is of high accuracy and capable of distinguishing seemingly perfectly identical twins even in the dark. Another advantage is its complete discretion. The development of this technology is direct-

ed today to reduce the costs in order to become widely applicable in the identification and authentication procedures. The facial thermogram is the most promising method at present.

Similarly to the other biometric procedures, the processing begins with noise filtering. Noise and disturbing factors include: eyelashes, eyelids, pupils, reflections. Then, the recognition of structure and the generation of code take place. This data set consists of a sequence of properties described in a polar coordinate system taken concentrically with outward motion starting from the pupil of the eye. As a camera that is capable of taking as many as ten pictures in a second at nearly the same focus, it is used to take pictures, it is possible to select the "best" picture with the lowest noise level (eyelashes, eyelids, reflections).

Iris identification has caught up to fingerprint identification in a short time, and became clearly a complement to finger at present. This is because multimodal identification offers higher security (lower false rejection, greater selectivity) on the one hand, and if the fingerprint testing is the primary one, it provides an alternative identification method, on the other hand.

Neither glasses nor contact lenses interfere with the black and white picture. The aging of the template can be compensated for by a simple re-recording (because of aging, illness: cataract, diabetes). It is important that the final image shall be sufficiently detailed so as to make the features used for identification well visible.

Iridology can help to survey the health of the person examined. By means of pupil control, it is possible to demonstrate whether a person is tired, consumed alcohol or used any drug. In addition to its quickness, the key advantage of iris identification is the accuracy – thanks to the iris stability – because the eye is an internal organ, protected, yet visible from outside. Several countries use iris identification. Millions are included in an iris identification system throughout the world for reasons of convenience, which enables them to cross borders without passport.

In the light of the knowledge described above in can be stated that it is the iris identification that would be the best choice at present. Both considering the ratio of price/value and taking the reliability parameters into account, there are currently no better alternatives. Yet, it can be said that its use in a healthcare environment might be problematic, except for identification at a critical level. On the other hand, it is not competitive against the costs of fingerprint testing systems. The development and spread of this technology result in positive changes in this respect, however, I think it unable to be introduced at present.

For this reason, the search for an additional system to the fingerprint testing shall be continued in order to achieve the appropriate level of security. In the second part, an important consideration for the selection is the ease of use, the acceptable cost level and, of course, the proper level of functionality. I consider it worth directing the development towards the RFID technology. By using this, a multi-level identification system could be implemented, that would perform the identification at a standard level by means of RFID while the high-priority access control would use the fingerprint test and RFID in combination.

References

[1] IMRE S. et al.: *Examination of RFID Systems In Terms Of Consumption and Technology.* (In Hungarian), www.rfid.answare.hu:8080/site/kutatasi-erdmenyeink/radios-megoldasok/2006/rfid-rendszerek-vizsgalata-felhasznalas-es-technologia-szempontjabol.pdf/view (downloaded: 12 02 2012)

[2] VARGA D., OLÁH A.: Biometric identification systems. (In Hungarian), *Híradástechnika*, 7 (2004). www.hiradastechnika.hu/data/upload/file/2004/2004_07/HT0407-8.pdf (10 10 2013) [3] BUNYITAI Á.: Today and Tomorrow's Automatic Identification Procedures for Access Control Security Systems Perspective. (In Hungarian), *Hadmérnök*, VI 1 (2011), 22–35. www.hadmernok.hu/2011_1_bunyitai.pdf (downloaded: 11 08 2013)

[4] SZABÓ M. D.: *Biometric Identification and Data Protection*. (In Hungarian), http://szabomat.hu/tanulmany/biometria.pdf (downloaded: 16 08 2013)

[5] KOVÁCS T.: *Biometric Identification*. (In Hungarian), Budapest: BMF, 2009. (College digital note)

[6] http://upload.wikimedia.org/wikipedia/commons/3/3c/Biometric_system_diagram.png (downloaded: 12 10 2014)

AARMS Vol. 14, No. 2 (2015) 201–207. **10.32565/aarms.2015.2.6**

Selection of Authentication Systems for Hungarian Health Care, Based on Physiological Study

Part II.

The RFID Systems

BARÁTH Artur

The actuality of the topic indicated in the title comes from more and more events where the verification of identity might be required. Health-care is an important part of critical national infrastructures. A primary task of the protection of critical information infrastructures consists in the access control of data managed by IT system where the identity-authentication forms an important part. For this reason, the technologies considered suitable for use and their supporting means are listed. After examining and comparing the relevant parameters, the optimum solution for the authentication procedures in the Hungarian health care system is specified.
First of all, the modern and efficient biometric identification processes are examined; however, the possession-based Radio Frequency Identification (RFID) as an additional system is also studied. Due to the well-known problems of passwords and chip cards. They were deliberately omitted during the system planning and efforts were made to exclude the human factors from the planned authentication system as much as possible.
Keywords: *authentication, biometry, RFID, health care, human factors*

Study of RFID Systems

In general, it can be said that the radio frequency identification systems serve for identification of goods, products or persons by using radio frequency data transmission. The communication takes place between the writing/reading device (reader) and the electronic data carrier unit (transponder) at a frequency corresponding to the external conditions and the required reading distance.

Basically, two types of systems in service are distinguished:

- *Passive system* – the transponder has no dedicated power supply; the energy required for the signal transmission is obtained from the electromagnetic field generated by the reader.
- *Active system* – the transponder has its own power supply and emits its identification signal either continuously or on call received from the reader.

Passive System

The data to be identified are stored in a circuit provided by an antenna and a memory (transponder). For data acquisition, a so-called reader is used. The RF3 signal modulated by the control data will be amplified in an amplifier unit and radiated through an antenna system. Within the range of a reader, the radiated signal induces power supply by means of the transponder antenna in the transponder circuit. The transponder addressed radiates a reply signal corresponding to the data read from its memory back to the reader. The reader forwards the signal reflected by the transponder by means of an amplifier and a demodulator to the processing computer. Thus, using the RFID, the necessary information becomes accessible in a simple way, quickly and reliably without physical contact. [1] Figure 1 shows the block diagram of passive system.

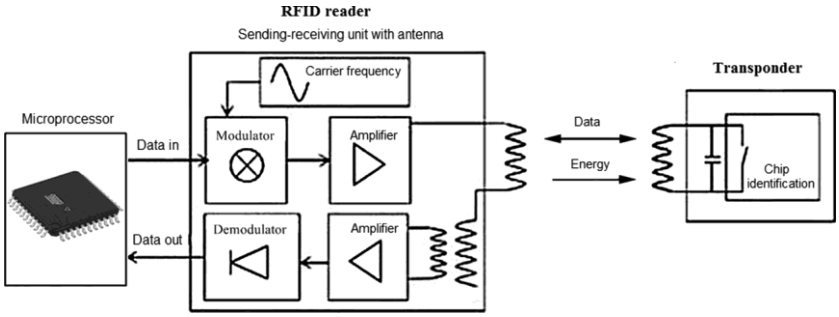


Figure 1. Block diagram of passive system. (Made by the author.)

Active System

Having its own power supply, the transponder radiates its identification signal continuously, or, in a so-called semi-active mode, in reply to the signal radiated by the reader, the transponder radiates a reply signal corresponding to its identifier stored in its memory by using its own power supply to the reader.

The use of RFID technology in access control systems was the prime mover of system development. According to the different security level of access control systems various card types (transponders) were developed that can be characterized by the type of memory, coding of identification data, and the range of reading.

In the operating principle of a system a number of groupings in addition to the two ones already made by power supply to transponder can be used. The most important features of RFID systems are: operating frequency and range of reader as well as the mode of coupling. The range of reading depends on many factors, but the most important ones are the power supply to the transponder and the mode of data transfer to the reader. Various couplings are used for this purpose. [1]

The systems can operate in a very wide frequency range from 13 kHz to 5.8 GHz. The available range extends from a few mm to 15 m in passive transponders and from 20 m to 100 m in active systems.

The large international frequency management organizations allocated four frequency classes for the purpose of RFID:

- low frequency (LF) RFID identification;
- high-frequency (HF) RFID identification;
- ultra-high frequency (UHF) RFID identification;
- microwave frequency identification RFID.

The coupling procedures of transponders specify the way in which the reader and the transponder can contact each other. The different coupling modes can be classified according to the coupling range i.e. close, vicinity and long-range. The range of close coupling lies within 0.01 m, the vicinity one between 0.01 and 1 m while the long range one exceeds one meter.

Connection procedures

The different coupling modes can be classified based on the physical properties of the coupling. Based on this, inductive, capacitive, magnetic, and backscatter couplings are distinguished. The capacitive and magnetic couplings are used in close-coupled systems, inductive coupling usually in vicinity-coupled systems, and the backscatter coupling in long range systems. In the present environment the subject of importance is that the operating conditions of medical devices shall be taken into account, while taking care that the radio frequency communication shall not cause problems during operation. [1]

Inductive coupling: "With the view of power supply to the transponder, the reader antenna generates a high-frequency electromagnetic field. The Cr capacity connected parallel to the reader antenna coil form a resonant circuit with a resonant frequency corresponding to the data transfer frequency of the reader. The transponder antenna-coil and the C1 capacity connected parallel are tuned to the data transfer frequency of the reader, resulting in the voltage drop across the coil takes its maximum." [1] (Figure 2)

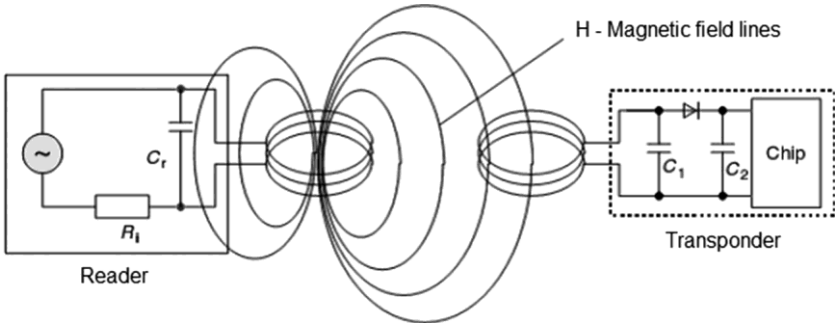


Figure 2. Model of connection. (Made by the author.)

Capacitive coupling: "Capacitive coupling requires no antennas, instead a pair of electrodes. Both the reader and the transponder have a conductor plate each that together form a capacitor. Of course, the realization of coupling requires the two plates to be parallel to each other. The transponder microchip located between two plates, one of which forms a capacitor with the plate of the reader and the other one with the ground. The voltage between the two plates of the transponder serves for power supply to the transponder microchip." [1] (Figure 3)

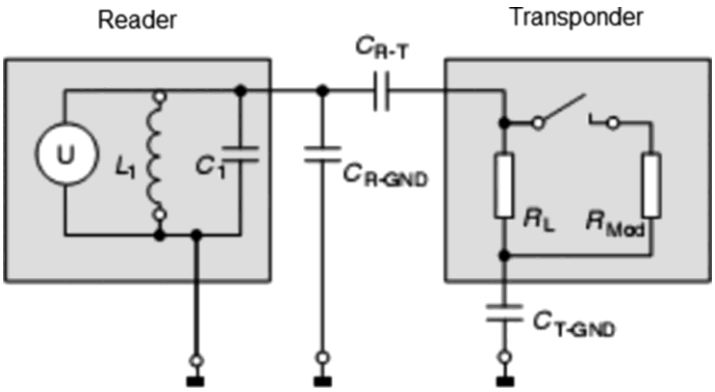


Figure 3. Capacitive coupling. (Made by the author.)

Magnetic coupling: "The magnetic coupling is very similar to the inductive one in that the antenna of transmitter and receiver form a transformer. The difference lies in that the antenna coil of the reader consists of a ferrite core provided with a circle- or U-shaped coil." [1]

Backscatter coupling: "Only a small part of power P1 radiated by the antenna of the reader reaches the antenna of the transponder (due to the section attenuation and undirected antennas) After rectification, the high-frequency voltage generated by the power P1 coming to the transponder is capable of reviving the integrated circuits of the transponder. Part of the power P1 is reflected by the antenna that comes back to the reader as an energy P2. The reflection characteristics of the antenna can be modified by altering the load connected to the antenna. Accordingly, during the transponder to reader communication a loading resistor RL that can be switched on/off is connected in series with the antenna. Switching it on and off, the data stream can be transferred. Thus, the P2 output power can be modulated (modulated backscatter). The reflected power P2 is also attenuated in space according to the outdoor attenuation, and therefore only a small fraction separated from the original signal comes to the reader so as to be able to be processed by the reader." [1]

Kinds of RFID Chips

In the description of RFID technology based identification, the element called a transponder is also very important. The chip carries the data of the person to be identified. During employment, a transponder is assigned to the person. This relationship will be stored in a database and processed using software. Due to cost-effectiveness, typically passive chips will be proposed to be used; yet, in view of the more accurate and more reliable operation, the selection of active chips is preferred.

Passive LF Chip

Passive timing chip, not programmable, contains a factory made unique identifier. The chip is mounted in a special plastic envelope that protects it against external physical impact and ensures practical wearing. It is reusable; therefore it could be used even in the further development in case of patients as well.

Active LF Chip

An active transducer used in the system is equipped with an internal energy source, which results in the production cost being multiplied; in exchange two wires as an antenna are sufficient. In addition, the operation is more efficient by one order of magnitude. Since the position and fixing of the connection points in this device allow more flexible solutions, the system provides more options. As the costs are much higher in case of this system, the application is likely still to come.

Passive UHF Chip

It is the write-once and read-many times chip that seems to be the best alternative. A capacitive coupling to the reader is used; therefore the antenna is designed as an armature with a relatively large surface. These chips were originally developed and produced as identifiers for one of the largest buyer markets of the RFID technology for the purpose of tracking goods in logistic systems. Due to the high volume of orders, it can be purchased from manufacturers at low cost; therefore it can also be used in a non-recurrent way. The efficiency of these chips has deteriorated in a wet environment and therefore their use for measurement causes problems if shaded by the human body or covered by wet or sweaty clothing. This problem can be solved by using a spacer that provides the necessary distance of two to-three mm away from the human body.

Battery Assisted Passive (BAP) Chip

The BAP chips contain a-battery plate built in a passive chip that contributes to the power supply required to wake up the chip and increase the energy level of the reply signal. This allows measurements at distances of as much as 100 m. BAP chips preserve their charge for 2.5 years, then they continue to operate as common passive chips.

Reader (Reader and Data Collector)

To obtain data to be identified and stored on the chip a so-called reader is used to produce the signal that powers the antenna to generate the appropriate electromagnetic field. It is another task to detect the data radiated back by the chip and, by processing them, to obtain the necessary information. The data thus obtained are stored in its memory and transmitted to the computer connected to it, which, in turn, saves the data by means of software into a remote database. Thus, a safe system is implemented, should the backup of data in the remote database fail; data can be found in the storage of a reader and transferred again.

It may have its own power supply, so that the operation is not affected by any power supply failures, yet, the power supply to the other components of the system shall remain operative.

In case of appropriate types of chips, certain RFID readers are also capable of writing not only reading. In practice, this means that information can be added to or deleted from the data stored in the chip's memory.

The proper communication requires an RFID/UHF reader. Depending on its design, it is capable of controlling 4 or 8 antennas. The high data bandwidth and advanced anti-collision algorithm allows for 430 chips/sec reading speed, therefore it is primarily used in applications with a large number of chips. It is powered by internal battery or an external power supply.

In getting acquainted with the technology and equipment one can realize that a system of proper configuration can be developed. In my opinion, the RFID is suitable to be used even as a basic system, i.e. in applications where problems of authority arise in respect to use – e.g. elevator, door instrument. A further advantage is that, in case of proper design, the position of employees wearing individual chips can be known at any time, which is a significant aspect in emergency situations. The possibility of improvement is also given in fact; it is not only the employees that can be provided with RFID but also the patients for the time of treatment in hospital. The biometric system can be used as an additional identification system if required by the circumstances.

Conclusions

The studies have shown that the benefits of biometric systems are clear while their disadvantages are far from those unable to be managed. From among the systems studied, it is the fingerprint testing system that is considered the most appropriate one and its use is proposed as supplementary equipment in the applications as follows:

- computer login in every case;
- entry to any surgical environment;
- access particular high-value assets;
- entry to ICU.

Of course, the certification can happen only if, in addition to the positive result of additional test, the admission from the principal system is also obtained by the person to be identified. It is the RFID technology that is selected as the main authentication system. In my opinion, each and every employee should wear a chip integrated in a wrist strap so as to be at hand any time without taking any trouble with it. In can be used in any case without obstructing the life of the wearer and its loss is properly hindered. As a result of study the structure set on the basis of preliminary expectations has been changed and the application fields of procedures interchanged.

Raising the range of systems described above to a national level, the permeability between institutions will be made feasible due to the possibility of configuring the authority levels at an institutional level. Thereby any employee wearing a RFID wrist strap can have admission to any health care institution in the country; thus those employed by several employers – as it is very frequent in Hungarian health care at present – can be identified by the same wrist strap.

As a further development, the use of a wrist strap could be extended to bed-ridden patients as well, that would improve the possibility of control and safety.

This study fails to define the technology to be applied; in fact, by the time that the study came to the phase of implementation, any and all equipment up-to-date at present would – fortunately – virtually be outdated due to technical development. The detailed specification of equipment shall be made at the time the introduction of the Unified Health-Care Authentication System becomes opportune.

Reference

[1] IMRE S. et al.: *Examination of RFID Systems In Terms Of Consumption and Technology.* (In Hungarian). www.rfid.answare.hu:8080/site/kutatasi-erdmenyeink/radios-megoldasok/2006/rfid-rendszerek-vizsgalata-felhasznalas-es-technologia-szempontjabol.pdf/view (downloaded: 12 02 2012)

AARMS Vol. 14, No. 2 (2015) 209–218. **10.32565/aarms.2015.2.7**

Application of Special Risk Reduction Protective Measures in Combiterminals for Dangerous Goods

CIMER Zsolt,[1] VARGA Ferenc[2]

During the combined transport of dangerous goods, in the territory of combiterminals, transport takes place from one type of transport to another means of transport. During the trans-shipment and the temporary storage of hazardous materials accidents resulting in smaller or larger consequences may occur (and in recent times have occurred as well). Experience has shown that reducing the consequences of such accidents requires the use of some special tools. In our article we present these tools in order to reduce such consequences as well as the technical foundations of their applicability.

Keywords: *combiterminals, disaster management, major accidents involving dangerous substances, device system, technical tools*

Combiterminals for Dangerous Goods

The transportation of dangerous goods – similarly to any other goods – is mainly carried out in a simple way: involving only one mode of transport. In the course of combined (or multimodal) transportation, the transportation task is carried out with the participation of two or more transport modes. Due to the cooperation of different modes of transport in the course of combined goods transportation and establishment of shipment chains, the advantages of each transport mode can be combined while eliminating the disadvantages. Combined transportation can contribute to the reduction of the deterioration of public roads, public traffic jams as well as environmental polluting effects. Such combined transportation requires an intermodal transport unit (e.g. containers), which have to be moved between the different transport modes. In the course of combined transportation the so-called intermodal transport unit containing the goods is moved with the help of special loading equipment. The movement of unit cargo, i.e. the loading from one vehicle to another is carried out in the combiterminals. [1] [2]

1 Szent István University, Institute of Fire Safety and Disaster Management; e-mail: cimer.zsolt@ybl.szie.hu
2 Budapest Disaster Management Directorate, e-mail: ferenc.varga@katved.gov.hu

Picture 1. View of a combined terminal. [3]

The reloading between means of transport in the combined terminals can be carried out by cranes or loading machines.

In case of cranes the rail stands on the ground and the crane structure forms a frame. One of the vertical pillars of the crane is fixed while the other is hinge-joined to the bridge so that the structure is statically determinate, i.e. it does not hinder thermal expansion. Cranes mainly move on special heavy-duty rails, alongside the loading rail-tracks.

Picture 2. Crane. [3]

Every fifth container in the world is moved by Kalmar loaders. These machines are equipped with special frames and are designed to move empty and loaded containers in port terminals. They are special advanced loading machines developed from fork lift trucks with a container holding frame in the place of the fork cart, which holds the loaded container at its four corners. The frame is adjustable; therefore it is not only able to lift 12 m containers, but also 6.1 m and 9.1 m containers. [4] In the below picture a Kalmar loader is lifting a so-called tanktainer (a tank container) from a lorry.

Picture 3. Kalmar loader during operation. [3]

Risk Analysis of Combiterminals

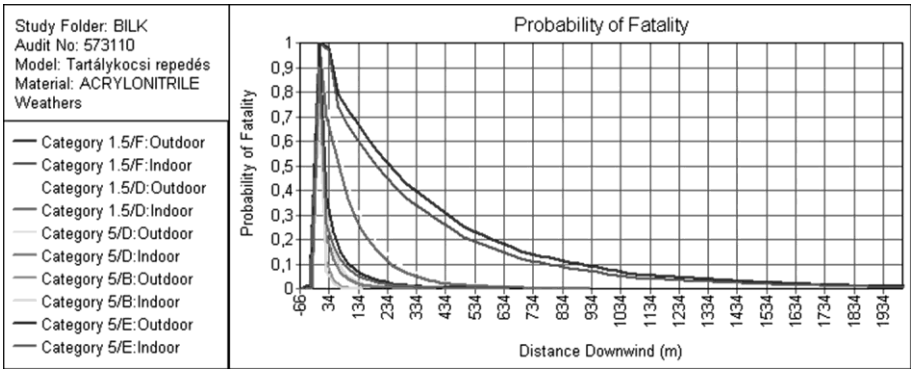
It is a common characteristic of combiterminals that dangerous goods might be present in their territory at any time in line with the current market demands, however, the time of presence is very short; in many cases the transfer takes place immediately (within some hours), but usually in 1–3 days. A direct activity (repackaging, mixing, solution, dilution, chemical reaction, etc.) involving dangerous substances is not carried out in the territory of combiterminals, therefore the certified packaging of dangerous goods is not removed.

In the territory of combiterminals the dangerous substance might be released basically in two ways:

- in case of catastrophic damage to the packaging, by an instant release of the dangerous substance (catastrophic breakage of packaging);
- in case of a less tightness of packaging, by continuous release.

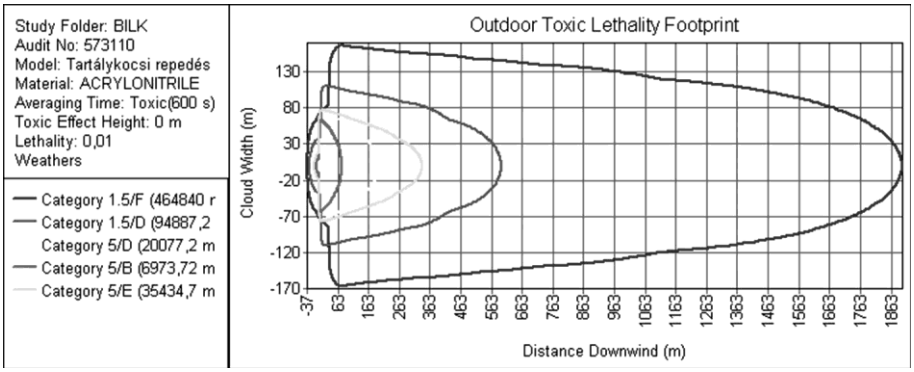
To present an example, with the help of Det Norske Veritas Phast consequence analyzing software, we show the toxic consequences resulting from the catastrophic damage of a 25-ton acrylonitrile (ACN) transport container.

The below figure indicates the variation of probability of fatality depending on the distance (from the container).



Graph 1. Probability of fatality in relation to distance. [5]

Based on the below figure for persons outside the 1,685 m area the probability of fatality is less than 1%.



Graph 2. 1% probability of fatality of persons outside, depending on distance. [5]

Based on the figures, in case of damage to the dangerous goods cargo the fatal consequences might also extend outside the combiterminal, and concentration values might be considered in a greater km distance which could lead to human injuries.

To analyse the possible external causes of the release of dangerous substances we applied the method of fault tree analysis. Fault tree analysis is a "reverse logic technique", which, after the definition of the unwanted event, determines its basic causes and their combinations with the application of logic relations. The fault tree analysis basically consists of the following 4 steps:

1. Definition of the problem (top event).
2. Creation of the fault tree.
3. Solution of the fault tree (determination of the fault event combinations).
4. Grading the fault event combinations. [6] [7]

We define "Dangerous substance accident affecting the civil population" as a top event. Based on the experiences in combiterminals operating in Hungary, we created the fault tree shown in Figure 1, below.

The following three simultaneous conditions have to be met in order for a dangerous substance accident affecting the civil population to happen:

- An accident takes place resulting in the release of a dangerous substance (catastrophic damage of the dangerous goods cargo or the substance flowing).
- The safety system serving for damage control purposes does not work properly.
- There is an inhabited area in the surroundings of the combined terminal.

The catastrophic damage of the dangerous goods cargo can be caused by:

- Damage of the lifting engine (Kalmar loader, crane) during the conveyance of materials (e.g. the revolving structure loosens (Event1) or human error (Event2).
- Fall of the dangerous cargo during storage (Event3, Event4).
- External force during storage (e.g. the Kalmar loader reverses against the dangerous goods cargo) (Event5).

The flow of the dangerous goods cargo can be caused by:

- Technical failure of the packaging (container) (e.g. the extracting stud does not close properly) (Event1) and the failure of check-up (Event6, Event7).
- External force (e.g. the Kalmar loader reverses/hits against the dangerous goods cargo and holes it) (Event8).

The improper operation of safety system serving for damage control purposes can be caused by:

- There is no safety planning; the Operator is not prepared for the management of such events (Event9).
- The Operator is prepared for the damage control on the level of safety infrastructure, but the safety equipment is not used, because:
  - The event is not noticed or it is recognized too late.
  - The protectiveness of safety equipment is under-planned or insufficient.
  - The concerned persons are not able to use the safety equipment, lack of training and practice.

The revealed basic events, i.e. the causes triggering the occurrence of the top event can basically be traced back to the deficiencies of the management system.

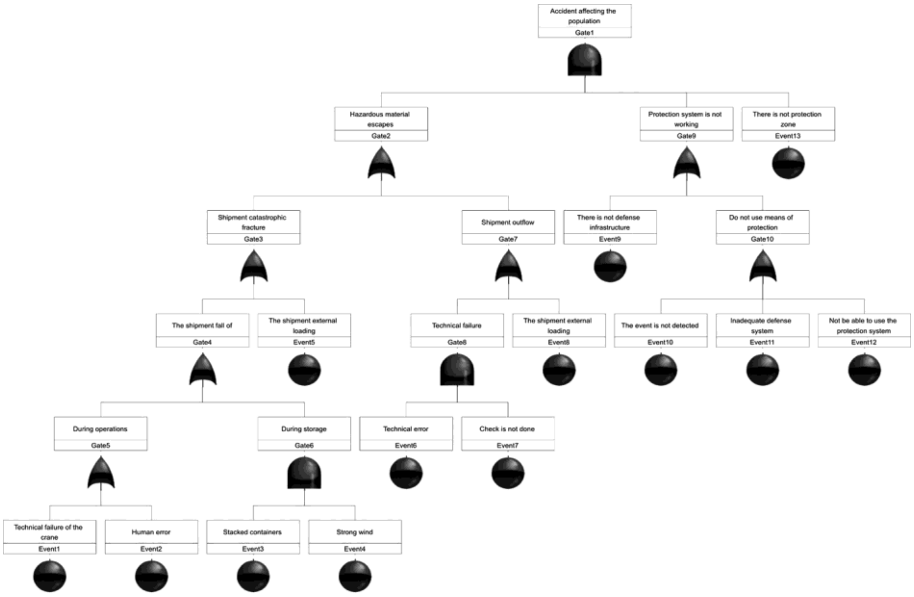


Figure 1. Fault tree: Dangerous substance accident with consequences also outside the combiterminals. [5]

As the above fault tree analysis indicates, in case of combiterminals the causes triggering the occurrence of the top event can generally be traced back to the deficiencies of the management system. The frequency of the top event can be reduced if we reduce the probability and/or frequency of the revealed basic events, which finally entail the top event itself. Such measures reducing the basic events should be included in some internal regulations (as part of the operator's management system).

The extent of consequences in case of an accident can be reduced by safety planning based on hazard analysis. In the frame of safety planning it is necessary to determine the exact measures to be carried out in case of accidents involving dangerous substances as well as the human and material resources required to carry out such measures. The special training of the personnel involved in the safety interventions should also be part of the safety planning, which should be tested by practical exercises. The safety planning should be treated as part of the management system, and as such the results should be included in an internal regulation.

Risk Reducing Tools Applicable in Case of Combiterminals

Considering the characteristics of combiterminals, i.e. that the present types and quantities of dangerous substances are continuously changing, we recommend the following three effective technical solutions be considered in the course of safety planning.

Water shield

A water shield is a special thrust nozzle with a standard Storz-clip usually made of aluminium. A vertical metal plate is joined to the horizontal entering pipe, with a special split between the metal plate and the pipe. Through this split the water streams freely form a water curtain

in the shape of a peacock tail. The water shields have to be set 5–10 m far from the exit point and also take into consideration the wind direction, so that the explosive gas cloud is directed towards the water curtain. No gas can pass the water curtain without mixing; therefore, the water shields have to be set in a way that the neighbouring water curtains overlap but do not collide with each other. Water shields set this way form a contiguous water wall. Usually a single water shield is not enough for efficient protection, as a high amount of escaping gas can pass by the two sides of the water shield. To prevent such a case, it is recommended to set more water shields in a V or U form to demarcate the gas cloud. [8]

The water shield is relatively cheap; it is easy to set up, so it can be used as a mobile tool, or it can be set as a fix structure sparing the installation time, and as such increasing the efficiency of protection. The disadvantage of water shields is that they cannot be used against dangerous substances reacting with water.

Picture 4. Water shield. [9]

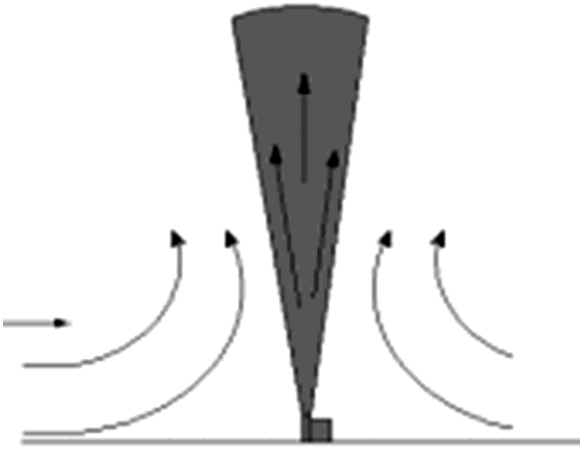


Figure 2. The section of a water shield – profile. [8]

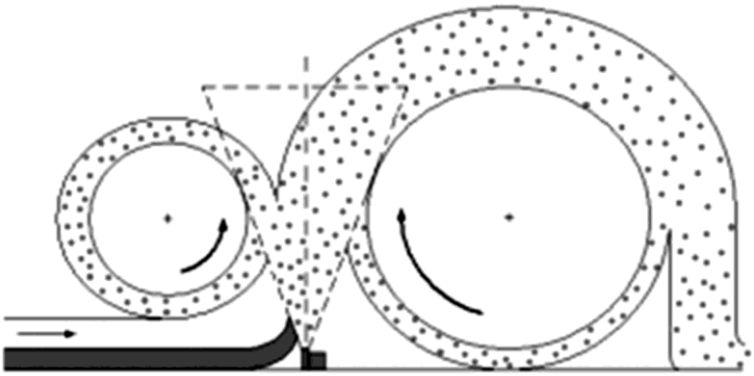


Figure 3. Streams developed around the water shield. [8]

Mobile safety bund

In case of accidents involving dangerous substances the extent of liquid flows can be reduced by applying mobile safety bunds. The mobile bund is a metal structure, on top of which the damaged container can be placed, so that it takes up the leaking substance. The substance getting into the bund can be extracted into an empty container either by gravity or by a pump. The installation of a mobile safety bund is more expensive than a water shield.

Picture 5. Mobile safety bund. [10]

Safety bund connected to an underground collecting system

The storage of dangerous substances can take place in a separate area, which can also function as a safety bund. The storage area has a concrete base covered with an acid-proof, chemical-resistant plastic layer. Its slope ensures the conduct of any potential dangerous substances and rain water into the metal barred drain hole. The liquid is channelled from the drain hole into the underground safety bund with a capacity at least 1.5 times the capacity of the container. The content of the safety bund can be emptied by opening/closing a gate valve.

The installation of the system is expensive; however, due to its very high efficiency the release of dangerous substances through an evaporating pool surface can be minimized.

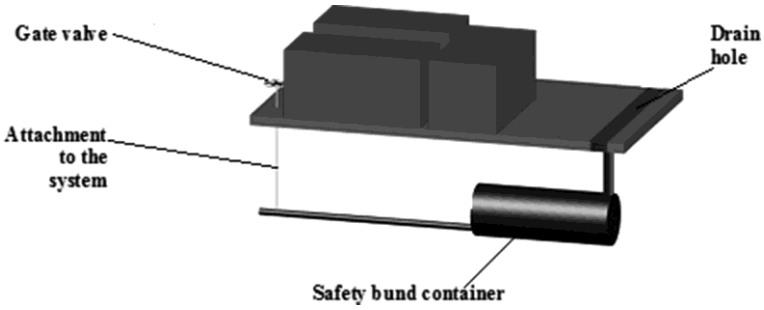


Figure 4. Bund connected to an underground collecting system. [11]

Summary

In the course of the combined transport of dangerous goods the movement/reload of the cargo from one vehicle to another takes place in combiterminals.

It is a common characteristic of combiterminals that any dangerous goods might be present in their territory at any time, however, the time of presence is very short; in many cases the transfer takes place immediately (within one day), but usually a 1–3-day long transfer should be considered.

Based on the hazard analysis it can be concluded, that accidents involving dangerous substances could happen in the territory of combiterminals, which might have an effect even outside the territory of the combiterminal and affect the civilian population as well.

In our opinion, operators of combiterminals, even if they do not fall under the scope of legal regulations on control of major-accidents involving dangerous substances, should implement the conditions of response to accidents involving dangerous substances in their management systems, and they should have efficient technical tools to reduce the potentially damaging consequences.

References

[1] PREZENSZKI J.: *Logistics I.* 11th Edition, Budapest: BME Institute of Post-Graduate Engineering, 2003.

[2] ARNOLD B.: Centres of logistics. *Hobbivasút*, 04 2011. www.hobbivasut.hu/cikkek/a_logisztika_kozpontjai.html (downloaded: 06 03 2015)

[3] *Website of Rail Cargo Terminal*. www.railcargobilk.hu/index.php?lang=hu (downloaded: 19 03 2015)

[4] SUPPLY CHAIN MONITOR: *Special tools*. April 2009. www.scmonitor.hu/archivum/2009/2009-aprilis-specialis-eszkozok (downloaded: 06 03 2015)

[5] SZAKÁL B., CIMER Zs.: *Control of major-accidents involving dangerous substances relating to combined terminals: The science for population protection.* Lázně Bohdaneč: Generalní Reditelství Hasického Sboru CR., 2015.

[6] BOGNÁR B. et al. (Eds.): *Industrial Safety I. Handbook on Implementation Tasks of Operators and Authorities.* (Iparbiztonságtan I.: Kézikönyv az iparbiztonsági üzemeltetői és hatósági feladatok ellátásához.) Budapest: Nemzeti Közszolgálati és Tankönyvkiadó, 2013.

[7] SZAKÁL B. et al.: *Industrial Safety I. Dangerous Substances and their Major Accidents in Industry and Transportation.* (Iparbiztonság I. veszélyes anyagok és súlyos baleseteik az iparban és a szállításban.) Budapest: Korytrade Kft., 2012.

[8] VERES R.: *The recent means of protection against explosive gas cloud of chemical industry.* Pécs: Pécsi Tudományegyetem, 2008. (Dissertation)

[9] *Website of Szolgálatiruha.hu Company Introduction www.facebook.com/szolgalatiruha.hu/photos/a* (downloaded: 19 03 2015)

[10] *Website of Agrotec Magyarország Kft., Company Introduction*. www.agrotec.hu (downloaded: 19 03 2015)

[11] MAJOROS E.: *The possibilities of reducing hazards caused by combined terminals*. Budapest: Budapesti Műszaki és Gazdaságtudományi Egyetem, 2015. (Dissertation)

AARMS Vol. 14, No. 2 (2015) 219–236. **10.32565/aarms.2015.2.8**

Detecting Military Historical Objects by LiDAR Data

JUHÁSZ Attila,[1] NEUBERGER Hajnalka[2]

Light Detection and Ranging (LiDAR) technology has become one of the major remote sensing methods in the last few years. There are several areas, where the scanned 3D point-clouds can be used very efficiently. In our study we review the potential applications of LiDAR data in military historical reconstruction. Previously we defined the major steps of the entire reconstruction process and the – mostly archive – useful data sources. Obviously the base of this kind of investigations must be archive data, but it is an interesting challenge to integrate a cutting edge method into such tasks. LiDAR technology can be very useful, especially in vegetation covered areas, where the conventional remote sensing technologies are mostly inefficient. We shall summarize how laser scanned data can support the different parts of reconstruction work and define the technological steps of LiDAR data processing.

Keywords: *LiDAR, geoinformatics, military history, remote sensing*

The LiDAR Technology

The Operation Principle of LiDAR

Airborne LiDAR is an active remote sensing technology. Its main operation principle is the same as laser telemetry: the scanner, placed on some kind of mobile platform (airplane, helicopter), emits a laser beam and measures the time it takes to return to its source. (Figure 1) The light moves at a constant and known speed, so the instrument can calculate the distance between itself and the target. To calculate coordinates from the raw telemetries we also need the position and the orientation of the laser pulse: the accurate coordinates of the sensor and the angle of the emitted pulse. Usually Global Navigation Satellite System (GNSS) and Inertial Navigation Systems (INS) are applied to determine these data. In practice the laser beam sweeps the surface of the ground across the flight track. [1]

1 Dr., BME Department of Photogrammetry and Geoinformatics; e-mail: juhasz.attila@epito.bme.hu
2 BME Department of Photogrammetry and Geoinformatics; e-mail: neuberger.hajnalka@epito.bme.hu

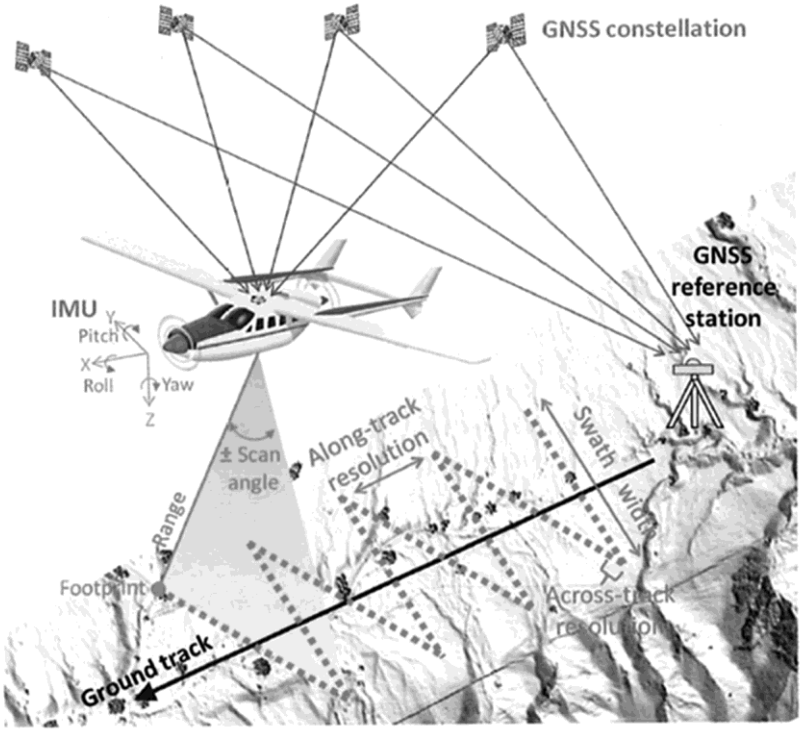


Figure 1. The operation principle of LiDAR. [2]

To understand the technology we need to review some basic characteristics of the laser beam: [3]

- monochromatic,
- coherent,
- very little divergence,
- polarized,
- concentrated energy.

According to the energy emitted and the distance the laser beam disperses; it reaches the ground surface as an ellipse (in practice with about 25–40 cm major axis) called the footprint, and then reflects from this point. Thanks to this characteristic, a single laser beam can reflect from numerous and various levels (ground level, understory level, crown level), this is called multiple returns. The early systems could record only one return, and then the first and last pulse detection was achieved. Nowadays it is possible to record more discrete returns (even 4–6) and the full waveform of the beam can also be detected by some systems. (Figure 2)

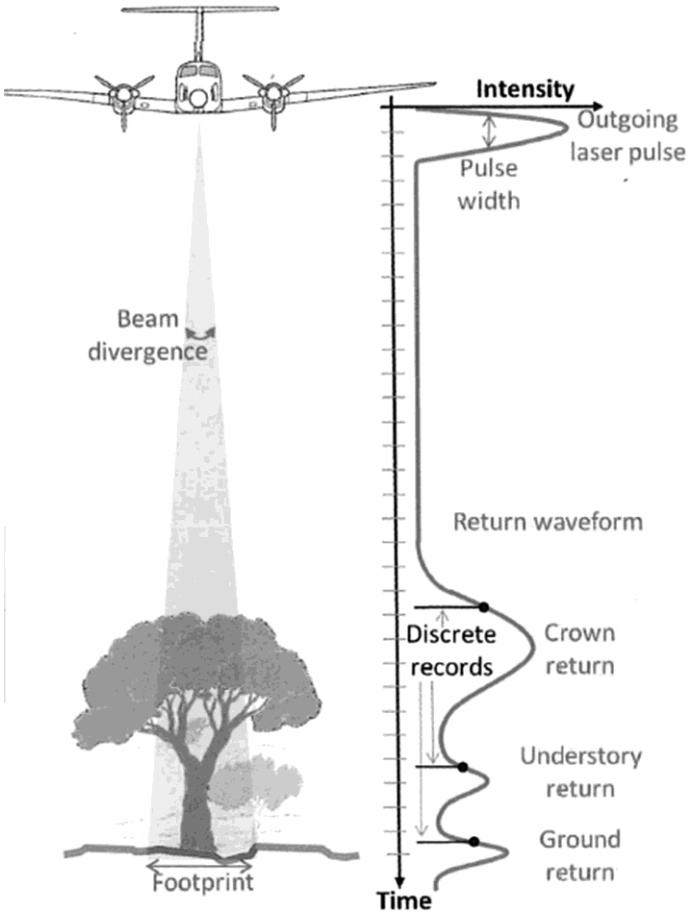


Figure 2. The principle of full waveform detection. [2]

Accuracy is another important characteristic that has to be discussed. Basically, the overall accuracy depends on the accuracy of the components (laser rangefinder, GNSS, INS) and the calibration of the sensor and the navigation system. Omitting the details, we can declare that in practice the accuracy of this measurement is about 15 cm vertically and about 25 cm horizontally. [4]

Airborne LiDAR technology is conventionally applied in civilian areas, like engineering [5] but nowadays it is also used in archaeological projects, [6] [7] even in Hungary. [8] Moreover it plays a more significant role in military missions, [9] education and training support. [10] For example in target positioning, detection of camouflaged and hidden objects, tactical map creation and battlefield simulation.

The LiDAR data

Airborne LiDAR data are usually stored in *.las* files. [11] Instead of a detailed review of this file format we present just a few important and useful pieces of information about the *.las* file for the users. The collected point cloud can be managed as a huge matrix and can be handled in mathematical software. Figure 3 represents sample data from Siófok loaded into MATLAB (matrix laboratory) software. The rows consist of coherent values of the single measurements and the columns represent the followings:

- X (m) UTM (horizontal) coordinate;
- Y (m) UTM (vertical) coordinate;
- Z (m) height above ellipsoid WGS84;
- intensity of discrete returns;
- return number of reflection;
- total number of returns in a single beam;
- 7–262. recorded values of the full-waveform.

Variables - A

A

99273x262 double

	1	2	3	4	5	6	7	8	9	10
28889	2.7745e+05	5.2000e+06	166.7970	49	1	1	4	4	5	5
28890	2.7745e+05	5.2000e+06	167.9070	53	1	1	4	4	4	4
28891	2.7745e+05	5.2000e+06	169.2620	51	1	1	5	5	5	5
28892	2.7745e+05	5.2000e+06	168.9850	35	1	2	4	4	4	4
28893	2.7746e+05	5.2000e+06	154.2160	10	2	2	4	4	4	4
28894	2.7745e+05	5.2000e+06	170.4700	4	1	2	3	3	3	4
28895	2.7746e+05	5.2000e+06	151.8670	51	2	2	3	3	3	4
28896	2.7746e+05	5.2000e+06	169.7320	24	1	3	4	4	4	4
28897	2.7746e+05	5.2000e+06	153.4500	28	2	3	4	4	4	4
28898	2.7746e+05	5.2000e+06	151.6370	1	3	3	4	4	4	4
28899	2.7746e+05	5.2000e+06	169.8470	18	1	3	4	4	4	4
28900	2.7746e+05	5.2000e+06	154.2830	5	2	3	4	4	4	4
28901	2.7746e+05	5.2000e+06	151.5680	1	3	3	4	4	4	4
28902	2.7746e+05	5.2000e+06	157.5460	26	1	1	4	3	3	4
28903	2.7746e+05	5.2000e+06	170.2080	4	1	1	4	4	4	4
28904	2.7746e+05	5.2000e+06	171.8220	59	1	1	4	4	4	4

Figure 3. LiDAR data in MATLAB. [2]

20th Century Hungarian Military Objects

In this chapter we review the potential of applying LiDAR data in the reconstruction of the Hungarian military historical objects and events in the 20th century. First, we briefly summarize the potential military historical events, locations and the typical military defense objects of the century.

There are no existing significant military objects (fortress or fortress parts) connected to World War I, in Hungary. There are some infrastructure and service facilities throughout the country (Budapest, Komárom, Mosonmagyaróvár). These facilities are mostly well known and mapped. The occurrent LiDAR scanning of these buildings can be used as reference data to locate and identify such objects in the investigation of new areas. Additionally, LiDAR measurements can be controlled and validated by field measurements.

As opposed to World War I, a lot of important events of World War II happened in the country. The objects that formed defensive fortifications (lines) built in this period are the potential targets of our research. These defensive lines spanned across the whole country from the north eastern region in a south western direction (Árpád-, Karola-, Attila-, Margit-line). The most important parts of these defense systems were the anti-tank trenches, the infantry trench systems and the various placements (flack, artillery). Based on of the size of these defense objects we can suppose that today's measurements also capture the signs of the larger parts of the defensive lines. The proper sizes of the anti-tank trenches are 8–10 meters width and 4–6 meters depth, so there is a great chance to locate the remains despite several decades of intensive agricultural activities in the concerned areas. Unfortunately, most infantry trenches have already disappeared from these agricultural areas, but there were several cases when the trenches were built next to forests or parallel with roads and dirt roads; some of these areas are still almost untouched. In these cases the trenches are usually covered by dense vegetation (grass, bushes, trees), so it seems reasonable to apply LiDAR survey and the most common process, the digital elevation model generation through vegetation removal. The remains of the trenches have a 1–0.1 meter depth, so it can be declared that the trench point must be well separable from other ground points in the point cloud. Confirming this, Figure 4 represents a foreign example of detected fire trenches on digital terrain model derived from LiDAR data.

Figure 4. Fire trenches located with LiDAR data (black arrow). [12]

Beside the World War II defensive lines, the South Defensive System (SDS) has to be mentioned, which was built in the early 1950's. It was located close to the former Yugoslavian border and was constructed in almost the same way as the earlier defensive lines, mentioned above. The SDS consisted of field- and permanent-fortification parts too.

In addition, we have to note that there have been only a few airborne LiDAR measurements carried out in Hungary until now. Unfortunately, none of these measurements were accomplished to reconstruct military defensive objects from the 20th century. However, there are some point clouds created by museums and national parks which include areas where potential defensive objects can be found. In these cases there is a possibility to locate these trenches or placements and to validate the known information based on laser scanned data.

Potential Applications of LiDAR Data

In this chapter we briefly review the point cloud processing, interpretation and visualization possibilities, which can effectively support military historical object reconstruction. As mentioned previously, the earlier types of the LiDAR sensors could detect just one reflected signal. However, the main analysis and interpretation processes can be carried out on this kind of (containing only first pulse reflections) point cloud too.

The first and probably the most important process is the classification of the point cloud. The primary goal of this step is to select ground points, i.e. deriving the digital terrain model (DTM). The definition of the accurate DTM is essential in case of archaeological and military historical studies. The identification of the relatively small offsets is the key task during the micro-relief analysis, thus as a dense point cloud should be used when possible. There are various known processes to select the ground points. The most frequently used weighting algorithms based on the relative vertical offsets and the selection depending on the slope or curvature. We have to mention that in most cases the software works in "black box" mode, so the users only have options to modify a few parameters or define typical land covers to achieve the best result.

Numerous interpolation methods can be applied in creating the elevation model, resulting in barely different models. However, we can declare that the most important issue is the quality of the selected interpolation method, because it has the greatest influence on the result. The most frequently applied interpolation methods are as follows:

- Nearest Neighbour,
- Natural Neighbour,
- Linear interpolation, like TIN,
- Inverse Distance Weighting,
- Spline based solutions,
- Radial Basis Function,
- Kriging.

Just like the interpolation methods, the various edge detection algorithms have a highlighted role in the military objects reconstruction. Applying these algorithms enables the enhancement, the accuracy and reliability of object detection, measuring the typical sizes and supporting the identification. Here are some typical edge detection methods:

- Roberts filter (calculate the direction and the value of the gradients, convolution filter, highlight edges in diagonal direction).
- Sobel filter (calculate the direction and the value of the gradients, convolution filter, all direction, post-processing needed: thresholding).
- Laplace filter (approximate the 2nd derivate value [0 at the edges], all direction, detecting blurred edges).

Obviously there are several other edge detectors. Usually it is recommended to apply pre- or post-processing along with these algorithms. For example, a smoothing filter can be useful before the edge detection, but we have to apply it carefully because it can delete the important small vertical differences from the data, which are necessary in most reconstruction cases. On the other hand the post-processing of the filtered data (e.g. thresholding, edge thinning) can significantly improve the interpretation. Besides the filters and edge detectors, there are simple visualization and analysis methods that are also able to simplify our task. Geographic information system (GIS) and computer-aided design (CAD) software has the ability to create cross and longitudinal sections, an extrusion of the height values or vary the illumination direction. These techniques can facilitate detection, identification, measurement and representation of military objects. Applying the appropriate parameters the visualization of the point cloud can be more plastic and the important height differences can be highlighted. Figure 5 shows a point cloud which contains of World War I trenches near Kemmelberg (Belgium). There are different edge detectors and illumination directions used to improve the interpretation of the objects (see arrows). [13]

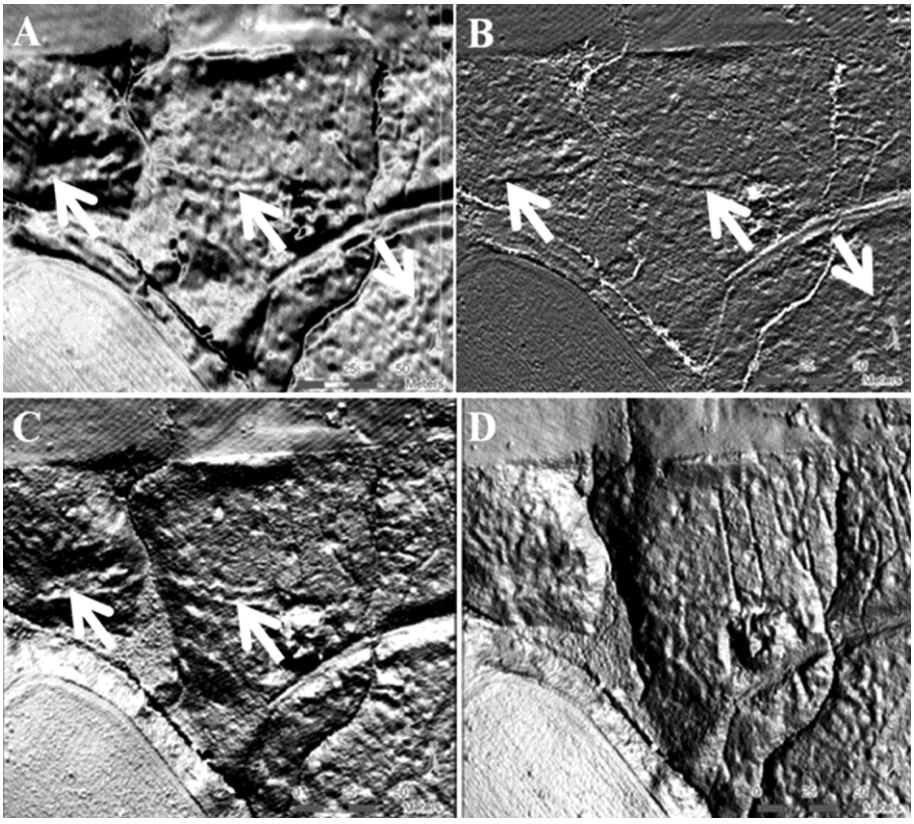


Figure 5. Detecting trenches near Kemmelberg (Belgium). Applying Sobel (A) and Laplace (B) filters and Northern (C) and Southern (D) illumination direction. [13]

Besides detecting the reflected laser beam pulses, the sensors record intensity values. This value depends on the color, material, texture and surface of the object from which the pulse backscattered. [14] Usually, these intensity data do not improve the military historical reconstruction process, because it does not carry extra information on the objects. Generally the investigated objects are made from concrete or soil, covered by vegetation.

As mentioned earlier, current LiDAR sensors are able to record the full waveform of the reflected pulse, in parallel with the discrete returns. In this case two measurements (the discrete and the full waveform) are derived independently from two different digitizers, which use different scales. [15] Processing full waveform measurements are often referred to as a potential procedure to gather more information about land cover and terrain objects, compared to the discrete detection methods. The research of the full waveform data is one of the hot topics of the LiDAR community; there are remarkable results in forested and in urban areas, but considering our aims, where the most reliable digital elevation model (DEM) is needed, these results are not so relevant.

The selection, classification and the representation of the LiDAR data is achieved by specific tools, but the latest versions of GIS software also contain many useful functions. In addition, there are other useful features in GIS software to improve the reconstruction: [16]

- manage various information from different data sources in a uniform reference system;
- generate DEM;
- interpolation methods;
- database connection;
- queries, spatial analysis;
- 2D and 3D visualization, animation.

So it seems to make sense to manage our LiDAR data in GIS environment that opens the way of integrating image and map data, increasing the effectiveness of environmental-, military object- and military event reconstructions. The GIS database enables the management a number of attribute information and attached documentation.

In conclusion, the LiDAR measurements and methodology are really useful in archaeology and military historical reconstruction, in object detection, identification or locating potential research territories. To achieve this, we need to set appropriately (if possible) the parameters and the time of data acquisition and carefully select the interpretation and representation methods. The detailed reasons for LiDAR data application are as follows:

- Object detection (according to vegetation cover):
  - open area:
    - investigate micro relief,
    - investigate intensity values.
  - vegetation covered area:
    - select foliage free period (if it is possible),
    - remove the vegetation, create DEM,
    - investigate micro relief.
- Identification (to achieve accurate object sizes):
  - the densest point cloud,
  - the most effective interpolation,
  - the most effective edge detector.
- Representation (the most expediential and esthetic visualization):
  - illumination direction,
  - height data extrusion.

LiDAR Supported Reconstruction of a South Defensive System (SDS) Fortification

Creating Uniform Base Map

In this chapter, we investigate a typical part of the SDS near Sátorhely and Majs settlements. Firstly, we carried out a traditional military historical reconstruction based on conventional data sources (images, maps), then applied LiDAR data where it was available (the LiDAR file contained only discrete reflections). During the reconstruction process we used GIS software. The base data were georeferred in Quantum GIS (QGIS); all the available maps, aerial images and satellite images were transformed into the Hungarian EOV (Egységes Országos Vetület – Unified National Projection) coordinate system. During the transformation we had to consider that the aerial and satellite images were not orthorectificated, so both are distorted. There were a few meters inaccuracy, but according to the reconstruction's demands, it was appropriate.

The Reconstruction Process

The first part of a complete military historical reconstruction is the environmental investigation of the period area. [17] The environment always has a great influence on the type, position and size of the military defensive system. The environmental reconstruction has two main steps. The past conditions can be derived from the recent state of the environment, from maps. (Figure 6) We applied ArcGIS software in this process, which has several cartographic functions, particularly in creating legends, so it is suitable for high quality mapping. The map objects were topologically corrected in GIS.

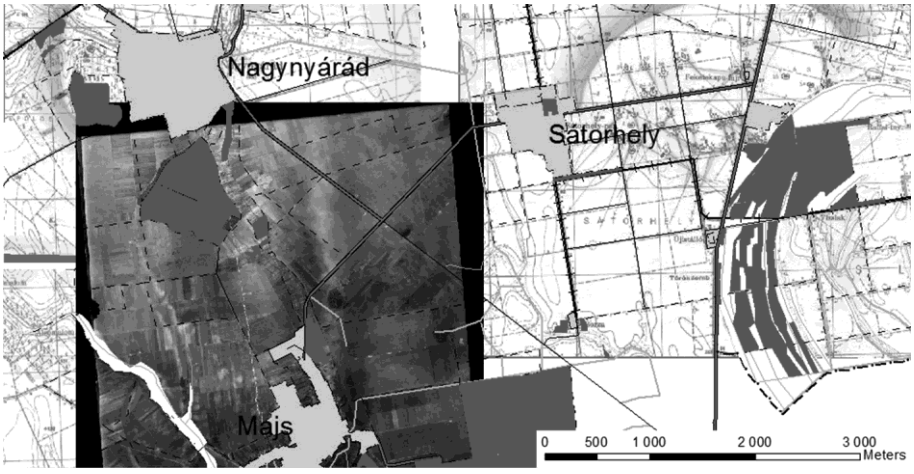


Figure 6. Environmental reconstruction based on aerial images and maps. [18]

A major part of environmental reconstruction is processing height data. Both topographic maps (contour lines) and LiDAR measurements can be the sources of elevation data. GIS also enables 3D modeling. Unfortunately in our study the scanned points did not access from the whole investigated area (Figure 7), so we also made a 3D model from the digitized contour lines. (Figure 8)

The defending capability of the defensive fortifications is mainly based on its location, depth and defense activity. As it is clearly visible in the investigated part of the SDS, the system was built to effectively stop and repulse the enemy with artillery, tanks and other engineering equipment, together with utilization of the terrain's characteristics. The bases of defense were the battalion defensive areas (e.g. Majs and Nagynyárád) and the subordinate company positions (e.g. Sátorhely) with connected fire control and engineering equipment systems. The battalion defensive areas were about 2 km in width and 1.5–2 km in depth. The company standing-grounds were 800-1,000 meters in width and 400–600 meters in depth [19].

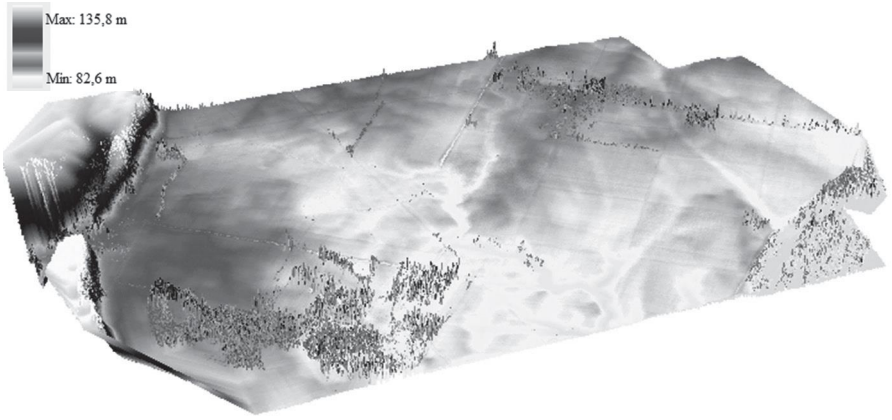


Figure 7. Digital surface model (DSM) from the LiDAR data. [18]

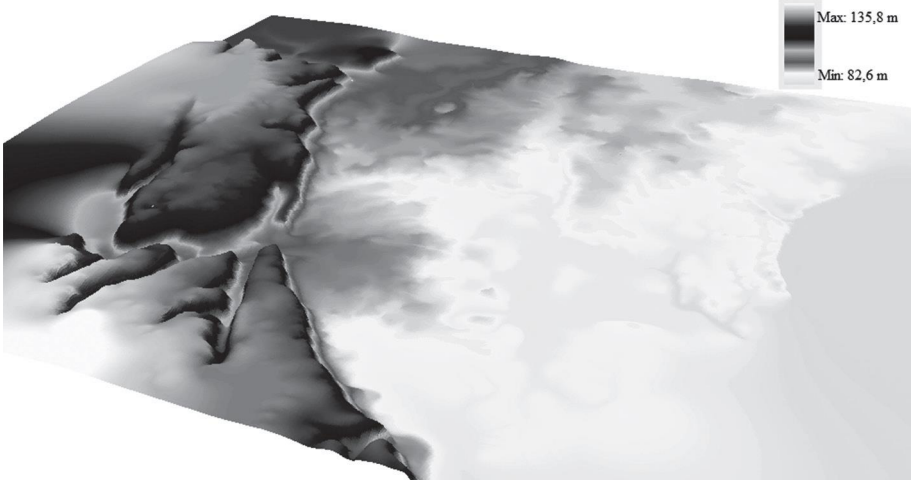


Figure 8. DEM from contour lines. [18]

In the following, we review the typical defensive objects and their major characteristics, which can be detected and identified.

Fire trench: a 0.6–1 meter wide and 1–1.5 meter deep infantry dugout, it enables safe and covered movement, and shooting at the enemy. Based on their typical shape and tracing, these trenches are relatively easy to identify. The reasons for the shape of fire trenches are the prevention of the enemy side fire and the reduction of an explosion's destructive impact.

Anti-tank trench: an 8–10 meter wide, 4–6 meter deep trench, it is usually these that are tens or even hundreds of kilometers long. Also they have a typical shape with linear parts and relatively hard breakpoints.

Various placements, bunkers: built in open or covered form alone or integrated in a defensive system. Because of their small (typically few meters in diameter) size, we can manage these objects as points in our investigations.

In our case, we had the chance to identify anti-tank trenches, because of the area's current land use (agricultural production). The point objects were digitized on the basis of the 1957 maps and the attributes came from the connected (paper) tables. There was a significant offset between the map based digitized point locations and the image based object locations, so we modified them based on the aerial images. As Figure 9 shows, the errors (see arrows) were not regular (offset between the numbered points coming from the map and the little light patches on the image), so it was probably not the image to map transformation that caused them. The next step was trench digitizing based only on the image data (Figure 10) and in some cases on the LiDAR data. (Figure 11) Figure 12 shows the part of the military object reconstruction with a special legend.

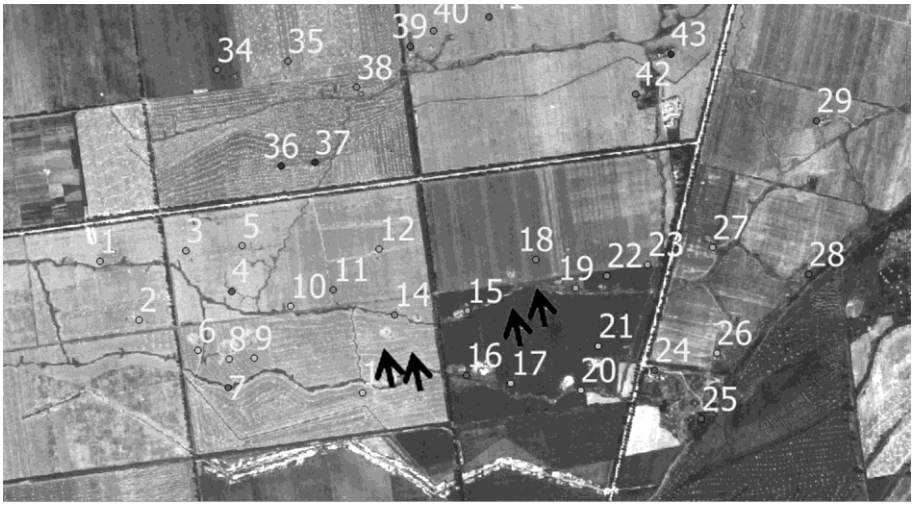


Figure 9. The offsets between the points coming from the map and the aerial image. [18]

Figure 10. Detecting and digitizing fire- and anti-tank trenches from an aerial image. [18]

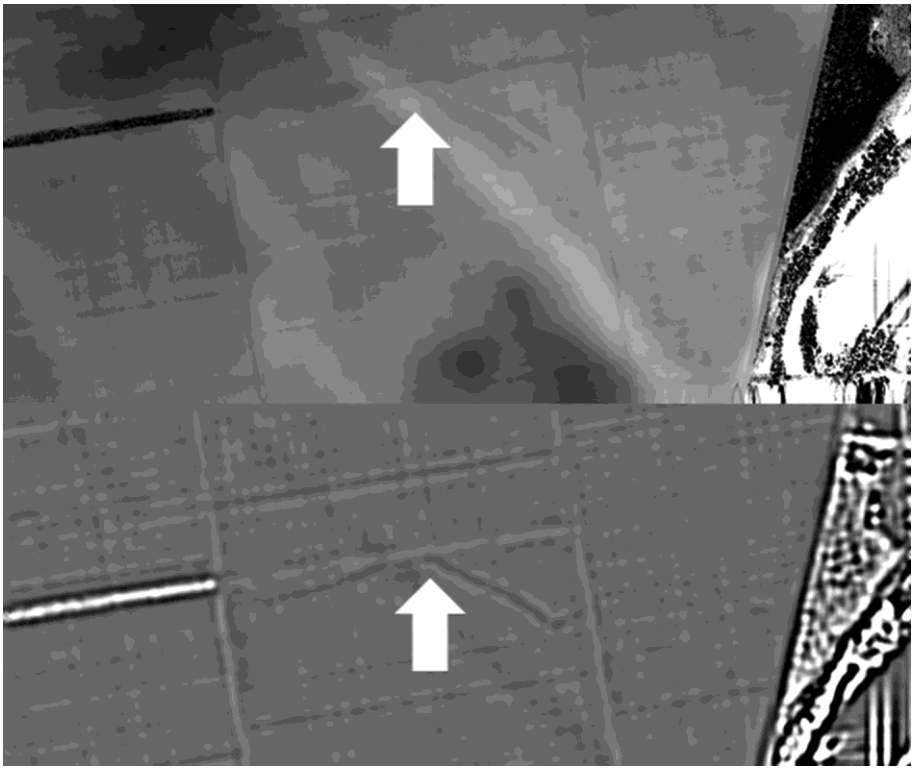


Figure 11. Anti-tank trench detected by kriging (above) and by Laplace filtering (bottom). [18]

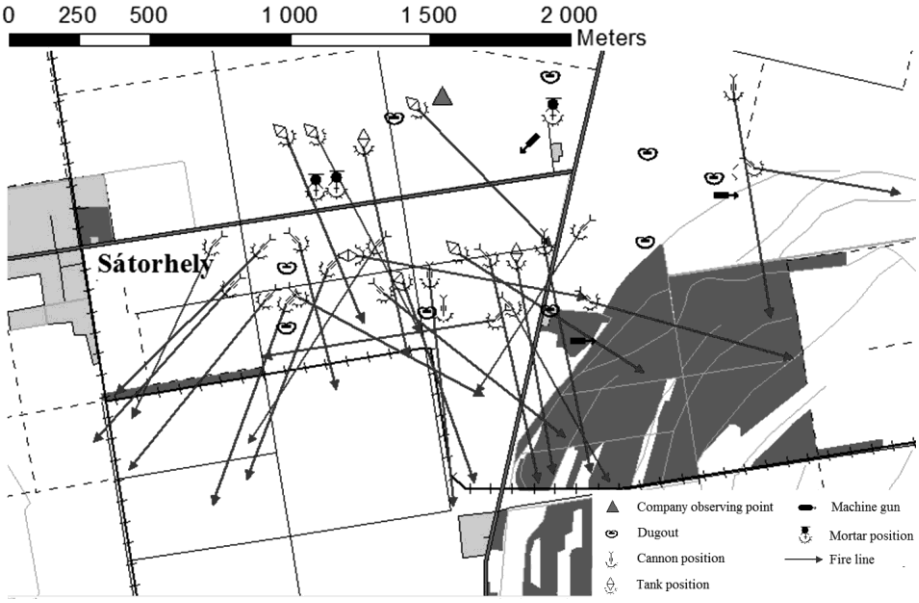


Figure 12. Part of the military object reconstruction. [18]

Since there were not any military events in this area in the 1950's, we studied the potential assault directions as an event reconstruction. In fact, we investigated the possible directions for heavy vehicles and for the infantry according to the accessibility and the environment. In addition, we carried out some spatial GIS analysis of the area, in the following sections we provide examples for both investigations.

The Investigation of Accessibility (Heavy Vehicle)

The forested areas, the wetlands, the wider streams and the steeper slopes (>20–25°) cannot be passed by heavy vehicles and tanks. Therefore, we derived the slope map from the terrain model and represented the planimetric data on it. As a result it seems that there are steep slopes at the western and south-western side of the area and wetlands and forests are located on the eastern side. As potential assault directions we defined the following two tracks: (Figure 13)

- 1st track: Tanks can pass through open terrain from south to north as far as the defensive trenches. However, the fortifications could stop them.
- 2nd track: Moving on the side-roads, the tanks can go around the defensive line and reach Sátorhely. To avoid this case, the defense forces should destroy the bridges and create road obstacles.

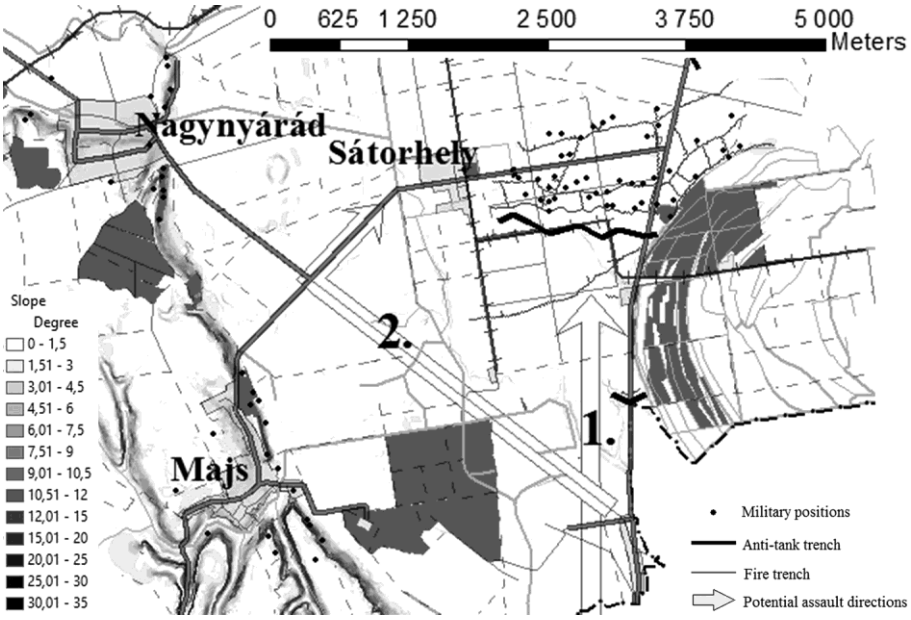


Figure 13. The slope map with the potential assault directions (heavy vehicle). [18]

The Viewshed Map

It is important to build an appropriate observing network to keep the fortification elements under control and to direct the fire. At least the roads connecting the subunits must be observed. Usually, according to the terrain and environment, these observation points were located on the side of ridges in well camouflaged locations. Figure 14 represents the controlled territory from the Majs battalion observing point (marked by black triangle next to Majs), we assumed a 130 meters height. According to the prior expectation, the roads and almost the whole defensive system are under control. To derive a better viewshed map we have to consider the followings:

- the accurate position of the real observing point;
- DSM usage instead of DEM;
- refining analysis parameters.

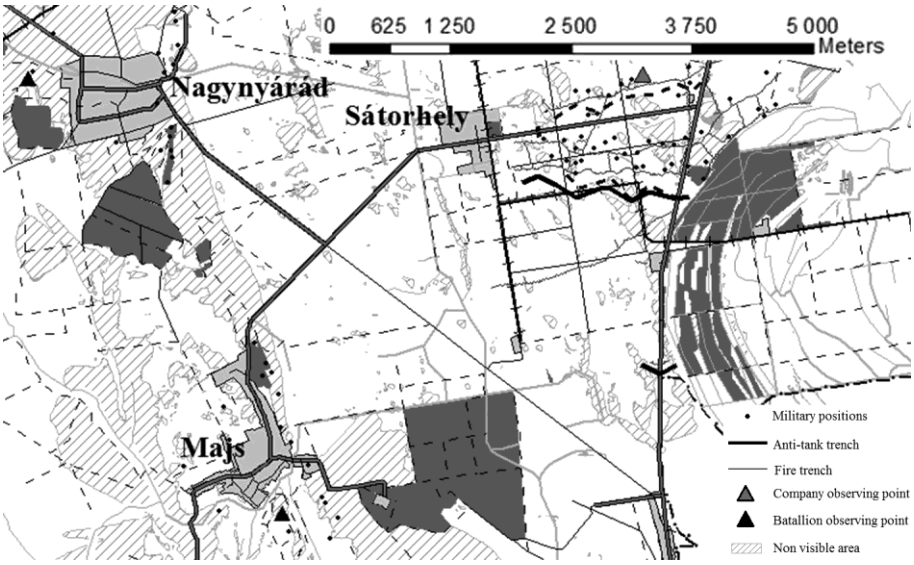


Figure 14. The viewshed map derived from Majs observing point (black triangle next to Majs). [18]

Finally, we review the typical LiDAR data processing steps in practice, supporting military historical reconstruction:

- generating manageable data (clipping, resampling);
- ground point selection;
- classification;
- DSM and DEM generation (Figure 16);
- visualization;
- GIS analysis (spatial, attribute-based);
- archaeological and military historical reconstruction.

Conclusions

In this portion we shall summarize our experiences of applying LiDAR data in military historical reconstructions. The elementary advantages coming from the technology itself:

- direct 3D solution;
- homogeneous point distribution;
- high accuracy;
- high resolution;
- undistorted measurement;
- directly enabled micro relief;
- fast data processing.

Through the military object reconstruction the anti-tank trenches can be detected almost in every cases and in case of beneficial circumstances (undisturbed area), the fire trenches and the placements can also be recognized. Further information can be derived by the simple coloring of the point cloud or creating cross sections or clippings. In our case, the estimated

depth of the detected anti-tank trench is 30 cm, as seen on the Figure 15. The potential applications of LiDAR data in object reconstruction are:

- potential research territory detection / selection, even in forested areas;
- direct 3D object measurements;
- further information from the recorded intensity values;
- archiving objects and finds by scanning.

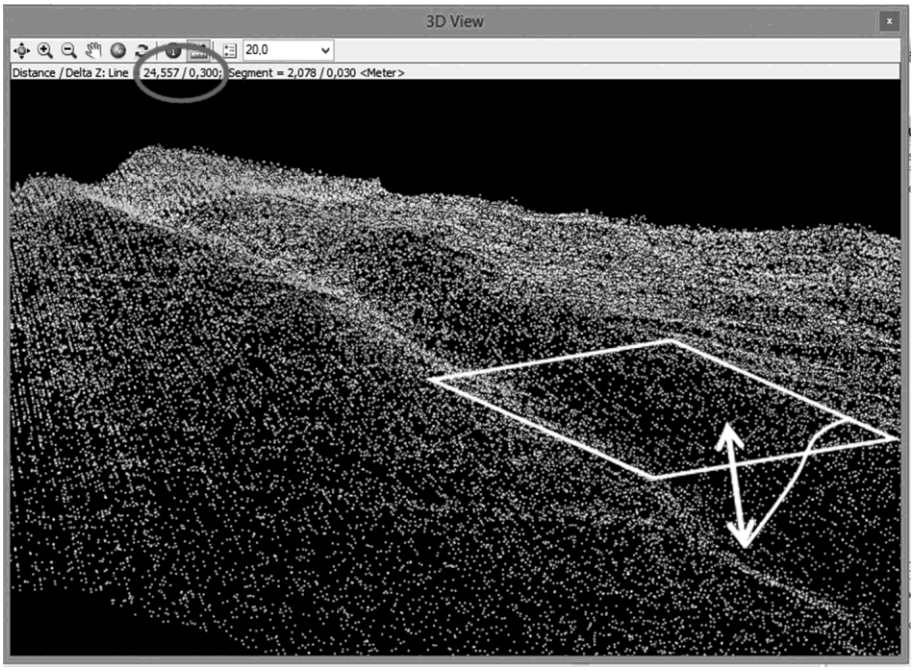


Figure 15. 3D view of an anti-tank trench part to measure the depth. [18]

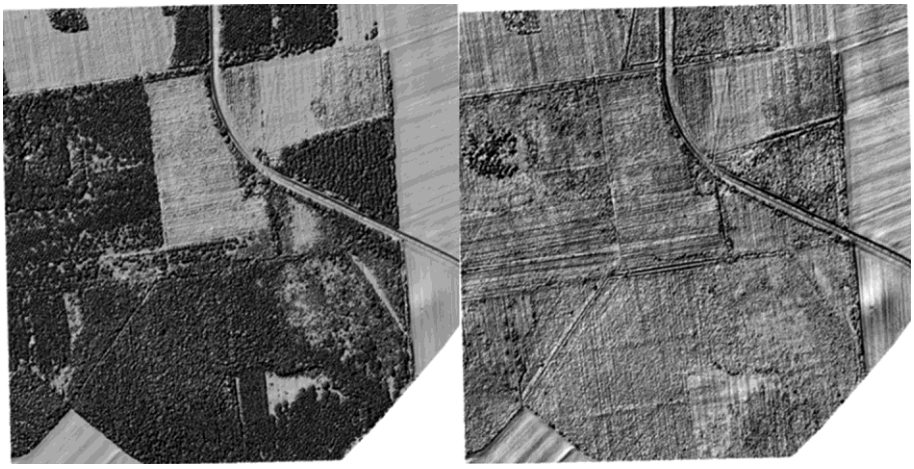


Figure 16. DSM (left) and DEM (right) of the investigated area. [18]

References

[1] BARSI Á., DETREKŐI Á., LOVAS T., TÓVÁRI D.: Adatgyűjtés légi lézerletapogatással. *Geodézia és Kartográfia*, 7 (2003). www.fomi.hu/honlap/magyar/szaklap/2003/07/2.pdf (downloaded: 06 03 2015)

[2] TAMÁS J., FLÓRIÁN T.: *Geoinformatics*. Debrecen: Debreceni Egyetem, 2008. www.tankonyvtar.hu/en/tartalom/tamop425/0032_terinformatika/ch05.html (downloaded: 06 03 2015)

[3] BOR Z.: A mindentudó fénysugár: a lézer. In. *Mindentudás egyeteme.* 2003. II. szemeszter http://mindentudas.hu/el%C5%91ad%C3%A1sok/tudom%C3%A1nyter%C3%BCletek/term%C3%A9szettudom%C3%A1ny/103-fizikai-tudom%C3%A1nyok/6008-a-mindentudo-fenysugar-a-lezer.html (downloaded: 06 03 2015)

[4] VERŐNÉ W. M.: *A lézer alapú távérzékelés*. Sopron: Nyugat-magyarországi Egyetem, 2010. www.tankonyvtar.hu/hu/tartalom/tamop425/0027_FOI3/ch01.html (downloaded: 06 03 2015)

[5] LiDAR: *The uses of LiDAR, LiDAR-UK*. www.lidar-uk.com/usage-of-lidar/ (downloaded: 06.03.2015)

[6] CROW, P.: *Savernake forest LiDAR.* Swindon: English Heritage, 2006. https://www.english-heritage.org.uk/professional/research/landscapes-and-areas/national-mapping-programme/savernake-forest-nmp/savernake-forest-lidar/ (downloaded: 06 03 2015)

[7] LAWRIE, B.: Beyon Angkor: How lasers revealed a lost city. *BBC News*, 23 09 2014. www.bbc.com/news/magazine-29245289 (downloaded: 06 03 2015)

[8] GÁTI Cs.: *A Pécs melletti Jakab-hegy őskori földvár LiDAR adatainak régészeti célú kiértékelése és térinformatikai integrálása*. s.l., 2014.

[9] *BuckEye*. Alexandria: Army Geospatial Center. www.agc.army.mil/Media/FactSheets/FactSheetArticleView/tabid/11913/Article/480897/buckeye.aspx (downloaded: 06 03 2015)

[10] AIR FORCE RESERVE COMMAND: Special Mission, Special Team: IMAs step up to serve on critical 3-D imaging project. *Citizen Airman*, 11 15 2012. www.citamn.afrc.af.mil/features/story.asp?id=123326431 (downloaded: 06 03 2015)

[11] ASPRS: *LAS Specification Version 1.3 – R11. 2010.* Bethesda: The American Society for Photogrammetry & Remote Sensing, 2010. www.asprs.org/a/society/committees/standards/LAS_1_3_r11.pdf (downloaded: 06 03 2015)

[12] JOHNES, D. M. (Ed.): *3D Laser Scanning for Heritage*. (2nd Edition). Swindon: English Heritage, 2011. www.english-heritage.org.uk/publications/3d-laser-scanning-heritage2/3D_Laser_Scanning_final_low-res.pdf (downloaded: 06 03 2015)

[13] STAL, C. et al.: Kemmelberg (Belgium) Case Study: Comparison of DTM Analysis Methods for the Detection of Relicts from the First World War. *30th EARSeL Symposium*, Remote Sensing for Science, Education, and Natural and Cultural Heritage. Paris: UNESCO, 2010. www.conferences.earsel.org/system/uploads/asset/file/46/earsel-symposium-2010_2-02.pdf (downloaded: 06 03 2015)

[14] LOVAS T., BERÉNYI A., BARSI Á.: *Lézerszkennelés*. Budapest: TERC, 2012.

[15] NEUBERGER H.: *Beépített terület felszínborítottságának vizsgálat teljes hullámalakos lézerszkenneléssel.* Budapest: BME, 2014.

[16] JUHÁSZ A.: *A lézerszkennelés lehetőségei a modernt kori régészetben*. Budapest: Pannon Palatinus, 2014. www.pannonpalatinus.hu/?p=3134 (downloaded: 06 03 2015)
[17] JUHÁSZ A.: *A XIX–XX. századi tábori erődítések a Kárpát-medencében – Hadtörténeti rekonstrukció térinformatikával.* Budapest: Tina Könyvkiadó, 2004.
[18] NEUBERGER H.: XX. századi katonai objektumrekonstrukció LiDAR adatok felhasználásával. Budapest: BME, 2014.
[19] RAVASZ I. (Ed.): *Betonba zárt hidegháború*. Budapest: HM Hadtörténeti Intézet és Múzeum, 2010.

AARMS Vol. 14, No. 2 (2015) 237–247. **10.32565/aarms.2015.2.9**

With Renewables for Energy Security

HAUBER György,[1] HONVÁRI Patrícia[2]

Taking into account the possible future exhaustion of fossil energy sources, the actual and near danger of climate change, the drastic increase of the greenhouse gases in the last 200 years, as well as the growing need for sustainable development, consumption and liveable environment, the increasing necessity of renewable energy sources becomes clear. Utilization of these energy sources have to acquire a bigger role in the field of energy supply, in order to enhance the energy security of Hungary, to decline the energy import dependence, to reduce the negative environmental impacts, and to recover the economy. The world's hunger for energy is growing exponentially; this is why it is crucial to establish feasibility scenarios in the next decades, which are able to meet these expectations, and to increase the safety of the energy supply.
Keywords: *energy security, fossil fuels, global climate change, security of supply, renewable energy, logistic system*

Introduction

In the 21st century the excessive exploitation of the fossil fuels is continuing, endangering, on one hand the energy security of the world, and increasing the negative consequences of climate change on the other hand. At the moment, the world cannot get rid of its dependency on petroleum and natural gas. The allocation of the resources is unequal, while the economic and political systems of the energy owners are questionable from several points of view. A solution for the vulnerable situation of the consuming-importing countries is still awaited. It would take time to elaborate strategic alternatives, which decrease the dependency and answer global climate change as the most serious challenge created by the growing utilization of fossil energy sources. A solution has to be found both on global and local level. On one hand it is necessary for the European Union to have a comprehensive energy strategy that is able to reduce this dependency; on the other hand every single member state and region has to find a possible way out of this tight corner. The aim is the diversification of the sources and accessibility, the formation of appropriate safety stocks, the preparation of a scenario, and a practice in order to react quickly and efficiently in unexpected situations. Another necessary and possible alternative of the diversification is the utilization of the renewable energy sources, the application in a wider range, because in the next few decades the greenhouse gas emissions have to be cut by three-quarters in favour of sustainable development. It is important to understand, eventually, that we should not dominate the planet, but that we are a part of it, and we cannot take more from the environment than it is able to produce. This can ensure that future generations will also have the possibility for development.

1 National University of Public Service, Faculty of Law Enforcement, college teaching assistant; e-mail: hauber.gyorgy@uni-nke.hu
2 Hungarian Academy of Sciences, Centre for Economic and Regional Studies, research assistant; Széchenyi István University, Doctoral School of Regional and Economic Sciences, Ph.D. student; e-mail: honvari.patricia@rkk.hu

Geopolitical Aspects of Fossil Fuels and the Question of Import Dependency

Energy correspondingly means political and economic power; there is an on-going, fierce struggle for its possession nowadays. This threat has effects on both a national and international level, which crucially concern the energy security both of Europe and Hungary. The world enhances the uncertainty with a lot of factors today. The different ethnic, religious, or civilization conflicts do not diminish. Wars, civil wars, dictatorships, economic crises aggravate the situation. New threats have also appeared, like terrorism or the information war. In addition to all that, the unequal distribution of the energy sources raises further problems. Social structures of the countries rich in energy sources are also leading in the direction of insecurity. It is enough to think about the armed conflict between Russia and the Ukraine, which is leading to tense relations with the EU, or about antidemocratic countries, like Azerbaijan, Turkmenistan or Kazakhstan. Similar threats can be recognized in the Middle East as well (this is the place where 65% of the world's petroleum sources are found), like the nuclear program of Iran, the Israel-Palestinian conflict, the recent terror attacks in Paris and many other outbreaks. The more energy dependent a country is, the more it is exposed to these dangers.

The main energy supply of the EU can be endangered if large-scale import is centred in the hands of only a few partners. In 2012 more than three-quarter (76.8%) of the natural gas import of the EU–28 originated from Russia, Norway and Algeria. This way, the import sources were even more focused in these three countries than in the last two years. Shrinkage in the production of coal, lignite, petroleum, natural gas and nuclear energy caused the European Union to increasingly rely on energy imports in order to satisfy the demand. This situation has stabilized after the economic crisis; however the energy import of the EU–28 is still the equivalent of nearly 922.8 millions of tons of oil[3] above export. Within the European Union the average rate of the energy dependency[4] is 54%, which is relatively high. There are some countries where this rate approaches or even exceeds 100% (Malta, Ireland, Cyprus, Luxemburg), and there are some others which are in a more favourable position, like Estonia (12%) or Romania (21%). The import dependency of Hungary is practically equal to the Union's average with 52%. [1] For countries with a high rate (such as Hungary) it is a top priority to dispose a long-term, computable strategy, which is able to reduce this exposure.

"Energy became the currency of political and economic power, this determines the power ranking of the nations, and this is the measure of the success and material development. This is why the access to energy is the top priority in the 21st century. This is the leading principle of every single government's foreign policy, the none plus ultra of the global energy industry, whose success is dependent on the following: the discovery, the extraction and the selling of coal, oil and natural gas (and the end product, the electricity) in increasing quantity. However, it is easy to establish at first sight, that despite every success, our energy management is stumbling. Oil industry is one of the most vulnerable economic sectors, as it is exposed to the sudden price increases, as well as to corrupt, tyrannical and labile political systems. Though

3 TOE: tonne of oil equivalent, is a unit of energy. It gives the mass of the burned crude oil, which is essential to the production of a given quantity of energy.

4 The rate of energy dependency is a ratio, which is interpreted as the quotient of the net energy import and the gross consumption.

natural gas is cleaner than oil, its transportation is excessively expensive, while the coal is cheap and easily accessible, however it pollutes the environment insomuch, that it causes the death of millions of people yearly." [2] These thoughts support well the power of the energy-rich countries, the vulnerability of other nations, as well as the importance of the energy security. The fights for the resources are intensifying nowadays. This danger should not be underestimated, as the "individual" interests seem to overwrite the interest of the humanity, and instead of enjoying a win-win strategy, the world can easily get involved in a lose-lose situation.

The Question of Dearth

According to the statement of the International Energy Agency, oil has a share of 40% in the energy sector, and at the moment it seems, that until the depletion of the stocks, this situation will not change significantly. Many argue, that the stocks will run out already by the middle of this century, however newer and newer oil fields are discovered and other (though less effective) resources are found. This so called non-conventional petroleum and natural gas extraction can bring new perspectives and potentially rearrange the world's energy map. These new developments have led to the fact that the peak-oil situation according to the Hubert rule did not occur in 2010. [3] However, in his book called *Out of Gas*, David Goodstein describes a threatening energy crisis, which is designated as the end of the oil age. This would not mean the complete depletion of the oil sources, but assumes a situation, where the oil extraction cannot keep pace with energy consumption anymore. [4] Decrease of the stocks puts the extracting and transit countries in a better position, while making the consuming-importing countries more and more vulnerable. This can lead to the rearrangement of the balance of power; otherwise it can increase the tension and create war situations.

Another significant energy source of the world is natural gas, which is responsible for 25% of energy consumption. It is more advantageous than petroleum in many ways, as the available stocks are far bigger; it is a cheaper, more convenient and cleaner energy source. Unfortunately the same problems occur with the utilization of natural gas, as 60% of the stocks are possessed by the United States, Russia and some member states of the CIS,[5] in addition the supplier roads are far more vulnerable.

In order to increase the supply security, it is necessary to analyze the question of the gas storage capacity. In the beginning of the heating season the size of the reserves in the four underground gas storage areas of Hungary was 4.368 billion cubic meters. However the consumption reached 4.2 billion cubic meters, and as a consequence during the period 2.5 billion cubic meters of gas has been removed from the reserves. The inflow rate is 12.8 million cubic meters per day altogether. [5] Theoretically, it is said that our reserves would be enough for 90–100 days, however the problem is, that in the past year, by the time of the heating season, the storage level was only around 40%, and even the inflow rate often lags behind the desirable rate. According to the Hungarian Natural Gas Storage Ltd. the technological level is fine and development is steady. The strategic moves in the following years have to include the following: reaching the appropriate inflow rate, as well as building further strategic gas storage and developing the existing ones.

5 Commonwealth of Independent States

The worldwide consumption of coal is experiencing a renaissance nowadays. The resources are sufficient for another 150 years, this is one of the cheapest energy sources, and as it is almost everywhere available, it can become one of the main energy resources of the developing world. In the past decade coal consumption grew by around 70%, and currently it covers 30% of the world's energy supply. The yearly growth of the coal based power generation in China alone is higher than the total renewable energy capacity of the 25 OECD countries altogether. [6] On average, one coal power plant is being installed per day worldwide. According to researchers, the coal power plants are responsible for 10% of the total carbon dioxide emissions. The health care costs due to air contamination caused by coal burning are reaching 20 billion euros according to cautious estimations, while the number of the lost workdays because of diseases related to coal burning exceeds 4 million. If in this case we ignore the external costs, the world's energy supply can easily go astray.

The danger can appear at any point in the supply chain: in the production, in the transportation, in the distribution, in the trade or in the consumption. It is not indifferent whether the resulting disturbances only affect a few elements, or having a larger scale, if they are temporary or durable, and how their impacts can be modelled. It is important to prepare for every element of danger in time, in order to avoid a global energy crisis. [3]

There is a more philosophical, however not less relevant approach to fossil energy use. Fossil fuels are labelled as income and not as capital, and according to economic rationality, which keeps in mind the maximization of income. The question is, can a "business" be feasible in the long run that absorbs its own capital because of the hunger for income. As a result of this conceptual accounting, the vanishing energy sources should be labelled as capital, and it would be obvious to minimize the present exhaustion, and to search for other, alternative production methods and lifestyles. [7]

Global Climate Change

Global climate change does not mean a risk of access to energy, however it has not foreseen and not estimated consequences. For global warming, two gases are basically responsible: carbon-dioxide and methane. The biggest "sin" of carbon-dioxide is that its atmospheric length of stay is between 20 and 150 years. This means, that the actions of the next few decades can fundamentally influence the global climate of the century end. The carbon-dioxide emissions in the middle of the last century were around 1,600 million tons, which grew fivefold until the beginning of the 21st century. The measurement of the single gases' atmospheric concentration is the part per million (ppm). In the last 400,000 years, this concentration never exceeded 300 ppm, however since the industrial revolution it is growing exponentially, and in 2014 it repeatedly exceeded the psychological limit of 400 ppm. If the emissions continue to follow this curve, the 550 ppm can be reached by 2035. [8]

This growing concentration is due to human activity, and not to any other reasons. However, according to the skeptics, things will get better and better, and there is no point in talking about an energy crisis. [9] In my opinion, research and examination has proved that these scientists are on the wrong track. There are some extremist theories, according to which, global climate change has already become an irreversible process, and destruction is unavoidable. Based on the Gaia-theory from James Lovelock, the Earth is one huge super organism, generating a self-regulating mechanism, it accommodates and develops, reacts to

the occurring changes. Previously, Lovelock believed in the self-healing of the Earth, but during a visit to England in 2004, he got an insight into the newest results of climate research, and based on these results, he claims that we are beyond the point where the processes would be reversible. [10] The consequences in the light of current research are only probable. As we look further into the future, forecasts become more and more uncertain. However, it is undoubted, that we are facing a worldwide problem which can lead to a global catastrophe in extreme cases. Based on the current grade of the greenhouse emissions, by the end of the century we can reach the temperature warming 3 degrees, which is considered to be critical and will launch irreversible processes. It is unequivocal that mankind has to face the biggest worldwide market failure so far and needs to elaborate a long-term, comprehensive strategy in order to handle of the problem.

What is the grade of the carbon-dioxide concentration which is still sustainable and is not disproportionately expensive? Most scientists determine this value around 450–550 ppm, which is only feasible if the global emissions reach their maximum in the next 10–20 years, and afterwards decline by 2–3% yearly. Taking that into account, by 2050, economic performance can be multiplied compared to today's level, then this measurement requires the drastic decrease of the emissions per GDP unit. The aim is the change-over to a lower greenhouse gas emitting economy, the establishment of climate awareness (perception), the decrease of the impact (mitigation), and effective conformation to the changed circumstances (adaptation). In order to function with these properly, the policies, the national governments, the regional economic units have to elaborate, facilitate and extort the unified market feedbacks, they have to overcome the market failures by taking into account the equality and equity.

What costs do the constraints on emission-reduction have, in addition to the installation as well as the dissemination of the new, non-coal based technologies? According to particular reports (like the BAU[6] scenario) comparing the social costs of the coal against the relief and the costs of the mitigation, this would bring in most of the branches a net benefit. Based on the Stern Review,[7] 1% of the total world GDP would be needed to change-over to the stabilization path. However, serious problems can occur from the fact that these costs are not equally distributed. At the moment, the poorest countries are mostly compromised. These nations are in a disadvantaged geographical position, as their climate is generally warmer, they are more exposed to drought, and they suffer more from the yearly change of rainfall, they are highly dependent on the agricultural sector, thus the climate change, the weather extremities affect them more sensitively. They do not have an appropriate health care system, in general they are struggling with political difficulties, the rate of corruption is high, the public administration is weak, and the GDP per capita is relatively low, which further complicates the adaptation to climate change. According to the UN High Commissioner for Refugees, if the extent of the warming does not change by 2050, the number of climate refugees can reach a number between 250 million and 1 billion. This can lead to excessively serious social, economic and ethnic tensions, and entails a high security risk. However, these are the countries that contribute only slightly to global warming. It would be justifiable, expected and equitable, if the world's leading economic powers gave a helping hand and disencumbered these nations. [11]

6 Behavioral Analysis Unit

7 The Stern Review was published on 30th October 2006. To the request of the British Prime Minister Tony Blair, Sir Nicholas Stern, former leading economist of the World Bank made an analysis about the climate change, and the possible economic effects.

Security of Supply

In this section, those factors are reviewed briefly which can endanger the security of supply on a national level. It is expected and likely, that on the fossil energy market serious problems can emerge in the future.[8] Increase of energy-demand will enhance the powerful import dependency. At the moment, the extent of diversification is low, coupled with high and strongly fluctuating energy prices. In the case of the producing and transit countries a high security risk is observed. The threat caused by climate change is rising. Energy efficiency and the introduction of strategic green energy use are extremely slow and circumstantial.

The highest priority of Hungarian energy policy is the intensification of the supply security. According to this, balanced and durable partnership with countries and organizations rich in energy resources, a suitable energy source structure, and the diversification of the energy import and transportation routes are needed. During her visit to Hungary, German Chancellor Angela Merkell recalled, that 30% of German natural gas necessities are provided by Russia. Although Hungary is much more dependent on Russian gas, Russian energy is a vital question both to Germany as well as to other European countries. According to her, the European gas pipelines has to be connected, and an opening into Azerbaijan would be essential, in order to decrease the dependency of the Union on Russian gas. Furthermore, it is important to evolve the appropriate security stocks, supply chains and reacting protocols related to the strategic energy sources, the maintenance and development of the existing infrastructure, and the formation of logistic service centers. The decrease of the energy consumption is also pointing towards the higher security, which can be realized by the increase of energy efficiency.

According to a 2014 autumn report of the International Energy Agency, the world will utilize only one third of the available and economically recoverable energy efficiency possibilities by 2035, if we do not change our methods. The issue of energy efficiency is underestimated everywhere, although this is the only field where economic growth can be reached without the increase of energy consumption. The American Council for an Energy Efficient Economy established in a study that the cheapest way to ensure new power generating capacities is not to build new power plants, but to increase the efficiency. With the energy efficiency programs one kilowatt-hour energy can be produced with half the costs, compared to the power plant. It is worth mentioning the research in the field of behavioral economics and choice architecture, which has reached effective results by making the energy "visible", as well as the policy measures which played a great role in awareness rising in society, making the consumers able to accept and apply the new energy sources in their narrower environment. [12] Another option of diversification is the higher utilization of green energy sources. A way out from the vulnerable import-dependency for Hungary can be the application and dissemination of renewable energy sources, both on the household level (solar panels, solar collectors, geothermal heating systems), and on the business level.

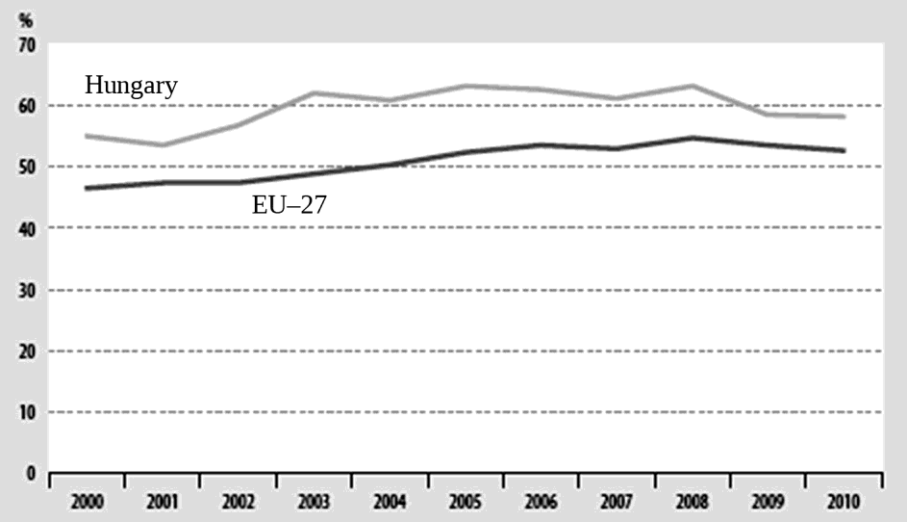
8 Hungary's long term contract with Russia on gas supply will expire this year, which can be the biggest problem of this year as regarding the tense situation. Energy needs of both the Hungarian economy as well as Hungarian households are fundamentally based on natural gas.

Hungary and the Renewable Energy Strategy

A lot can be heard about the importance of the renewable energy sources nowadays. However, it is still questionable, what kind of role will renewable energy have in the future. Can it provide an alternative to fossil-based, excessively wasteful global energy consumption? It is a fact, that in the global world, as well as in the European Union the alternative energy sources already have an increasing role, not only in terms of utilization, but also regarding the Union directives, guidelines and policies. It is a fact, that Hungarian utilization is showing a continuous increasing tendency in the past years. And it is also a fact that more and more rural settlements are opening towards renewable energy sources, creating the base of their local economy development. Naturally, the utilization of the renewables can only work well when the national strategies, action plans and programs also confirm this need. Although the New Széchenyi Plan, the National Rural Strategy or the National Development 2030 strategic documents are all drawing up the necessary increase of the renewable rate, the concrete activities and political will is proving to be insufficient.

Hungarian energy production is fundamentally based on the nuclear power plant in Paks (38%), while the biggest proportion of the production and import is given by natural gas and the petroleum. [13] The long-term energy strategy of Hungary does not intend to change this trend, which is confirmed by the security of supply and the lack of resources for the necessary investments; although it mentions the need to increase the share of the renewable energy sources (to 14.65% by 2020). [14] However, the present trend is not competitive, not safe and not sustainable.

The figure below shows that Hungary is in a similar position according to the energy import dependency as the Union's average, although in the last years it has exceeded it slightly. (Graph 1) This means, that in Hungary – similar to the EU – the increase of the supply security and the diversification of the energy sources is at a crucial standpoint.



Graph 1. Energy import dependency of Hungary and the EU (%), 2010. [15: 192]

Within import, the volume of hydrocarbons is the highest, especially natural gas, with 80% originating from import, primarily from Russia. 37% of the Hungarian energy consumption is based on natural gas; this is the second biggest share in the EU. In accordance with the decrease of the energy import according to the European Union directive (2009/28/EC) Hungary has to increase the share of renewable energy sources to 13% within energy consumption. The government has set to go beyond this and set a target of 14.65% by 2020 in its National Renewable Energy Action Plan. The primary production from renewable energy sources was 9.6% in 2012, which has doubled during the last 10 years; however it is still well below the set target rate. It needs to be also mentioned that the most important renewable energy source is biomass, giving almost 80% of the renewable energy production. Although the utilization of other renewable energy sources is increasing gradually, their application is spreading very slowly. There is no significant progress in the utilization of geothermal energy. The absolute worth of wind energy and biofuels is growing rapidly, however their significance is still slight. The current utilization of solar energy is still only a fragment of the capacity, according to the potential solar radiation, its application is spreading slowly. [13]

What are those factors which can guide the renewable energy policy of Hungary, and why would it be worthwhile opening towards higher utilization? The security of supply can be mentioned in the first place, as Hungary is highly dependent on energy imports, although it is given the renewable energy potential, which could increase energy security. On the other hand, the environmental sustainability and climate protection is also a determining factor, as the utilization of renewable energy sources can decrease the application of the fossil energy sources and as a consequence the carbon-dioxide emissions as well. The environmental and nature conservation aspects are also mentioned in the National Renewable Energy Action Plan. Furthermore, an important factor is agriculture and rural development. In case of the agricultural sector, it is essential to take into consideration the aspects of sustainability. Considering rural development, it needs to be emphasized that a determinant proportion of the renewable energy can be produced in rural areas. This is why the utilization of the new energy sources should play an important role in rural development policies in the future. Development of the green economy can also be a driving factor, as the application of the renewables together with the energy efficiency programs can give a basis for a new green sector economy. [16] Last, but not least the investments in the renewable energy sources have the positive returns, that when the economy is moving in a greener direction, it also has an effect on employment also. [17] [18] However, it is not only about the increase of the absolute amount of jobs. New workplaces will be created, though there will be jobs which will be replaced, some may disappear, and there will be some professions which will be redefined. The role of education needs to be emphasized, in order to educate an appropriate workforce according to the renewed needs. [19]

As for the long-term energy strategy of Hungary, the expansion of the nuclear power plant in Paks is a foregone issue. The government has decided a long-term directive with this. The energy strategy published in 2012 also counted with this possibility, but emphasizes, that "a new nuclear investment requires significant preparatory work, and a guarantee for appropriate and safe operation according to strict regulations. Regarding the establishment of the new block of power plants detailed public information is needed, in favor of greater social acceptance." [14] The huge investment accepted without an international tender is questionable. How can the 3,000 billion Ft loan help decrease energy import dependency? It is a fact, that

the nuclear power of Paks increases the quantity of nationally produced energy, however it also has to be seen, that the investment and credit is given by a country towards which our dependency is already huge.[9] In addition, regarding the nuclear energy the question of security is always emerging. Nuclear catastrophes are not unprecedented in history, including for example the nuclear accident of Three Mile Island in the United States, Chernobyl, or the most recent Fukushima-accident.

During its communication, the government put a big emphasize on the statement, that nuclear energy is the cheapest energy source, while the most frequently mentioned negative aspect in the utilization of the renewable energy sources is the expensive change-over. Yet, what is the price of the energy? If we only take into account the net costs of the extraction, the nuclear energy – along with the fossil fuel – is indeed economical. However, if we also calculate with the costs of the investment, the long-term commitment undertaken by the loan, the rise of the security risks, the immeasurable and so far not estimated environmental damage caused by the nuclear waste, in case of the fossil fuels the great amount of greenhouse gas emissions, is nuclear and fossil energy still cheaper? The question arises: does the world have a price where the climate change, the pollution, and all the problems coming from this issue – starvation, lack of drinking water, natural disasters, extinction of animals, growing welfare gap are vastly smaller? If we consider these, which energy source would be the "cheapest"? It is important to examine the question, whether Hungary has no other chance to decrease its energy dependency, than to invest in the nuclear energy. It raises the possibility of a false long-term strategic decision. If this is the case, it is crucial to improve the strategic position of Hungary with other, mostly local and bottom-up initiatives. The national guidelines and action plans of the development policy should also confirm this need.

"Think Globally, Act Locally"[10]

The nuclear energy investment in Hungary is a decided question, which can determine our energy strategy in the long run. However, this does not exclude the utilization of the renewable energy potential, and making it a determinant factor in the Hungarian energy sector. It is our possibility and obligation to live with this facility and capacity; however the necessary infrastructural background needs to evolve. The system approach is important, as well as the progress along the unified and integrated strategic principles, but the actual, viable solutions has to be found on the regional and local level first. Fortunately, many effectively and economically operated local green energy supply systems are observed, which can be the basis for further adaptation. However, the establishment of the necessary supply chains cannot be forgotten. The formation of logistic supplier centers is obviously expedient where the energy production activity is in progress in a given geographical area. Accessibility is excessively important, through high quality public roads, air, water and railway routes, as well as through the info communication tools. The infrastructural facilities and possibilities of the given area need to be examined, and a development action plan has to be established. The aim is to elaborate a logistical system which satisfies sustainable development. During the production of the renewable energy intervention constraints occur, as the output of the alternative energy sources is in the most cases non-consistent. The target is the establishment of a problem

9 If we only take into account the natural gas import, Hungary's dependency towards Russia is around 80%.
10 Famous thought of the Scottish urban planner, Sir Patrick Geddes.

solving protocol, the optimized operation, distribution and appropriate utilization of the renewable energy sources through logistic systems, the elaboration of smart network-centric logistic system, which is able to bypass the constant changes of the weather, and compensate the resulting production loss towards the consumers. [20] [21] The elaboration of the local systems and the regional development has several advantages. The investment need is smaller, the delivery and storage costs can be significantly decreased, and the advantage of the local production and utilization is that the energy transportation losses can be eliminated or minimized. Additionally, it can provide a solution to excessive centralization and urbanization. Further benefits of rural development – as mentioned before – are the creation of new jobs, the increased orders at local businesses, and the resulting additional income are also mostly utilized in the given area.

Conclusion

The security of energy supply is dependent on several factors. Unfortunately, Hungary is a small, open country, and not well supplied considering the fossil energies. The energy dependency rate is high, our supply sources are limited to 3–4 countries, and the economic and political structure of these countries also means uncertainty and unpredictability. Many elements of the supply chain can hide danger, and the current war situations only create further problems and questions. In addition to these geopolitical factors, the excessive utilization of the fossil energy sources is the direct trigger of the global climate change, the consequences of which are unpredictable. However, scientists and leading politicians of the world agree that the failure of immediate interventions can cause irreversible processes. This study highlighted that there is a way out for Hungary from this non-promising situation. Possibilities given by the renewable energy sources, the potential of green energy could provide a solution already in the short run. The question is, how much energy are we able to produce, what social and economic costs will it have, and how great a degree of risk is there? How can inequitable production and opposing consumption be synchronized? [22] In this area, our energy portfolio is still unstructured, biomass is giving 80% of green energy production, while the wind-, solar and geothermal energy could also be a significant base of this sector. The most effective argument against the renewables is that their production and operation is not cost-effective at the moment, the return of investment is too long. However, if we took the social costs of the fossil energy sources also into account, the balance would move in favor of the alternative sources. In models based on renewable energy sources we cannot forget about the establishment of the supply systems, about the elaboration of a smart, network-centric logistic system, which would be able to handle the intervention constraints during operation.

It is time for thinking together and activity. Every wasted year increases the risk, and decreases the possibilities for future generations to have a better life. The success of a country having such geopolitical and natural facilities as Hungary is significantly dependent on how it can change the conventional energy sources based economic model to an alternative way. Hungary needs to establish a systemic energy strategy, which means a way out from the hardly resolvable contradictions of the present system.

References

[1] http://ec.europa.eu/eurostat/statistics-explained/index.php/Glossary:Energy_dependency_rate (downloaded: 17 05 2015)

[2] ROBERTS, P.: *The End of Oil: On the Edge of a Perilous New World*. Boston: Mariner Books, 2005.

[3] KATONA T. J.: Az energiabiztonság, mint rendszer. *Energiagazdálkodás,* 54 (2013), 6–10. [4] GOODSTEIN, D.: *Out of Gas. The End of the Age of Oil*. New York City: W.W. Norton and Company, 2004.

[5] *According to the data of the Hungarian Natural Gas Storage Ltd.* www.magyarfoldgaztarolo.hu/hu/Ugyfelek/Adatpublikalas/TarolosiKapacitEsNapiAdat/ Lapok/default.aspx (downloaded: 17 05 2015)

[6] INTERNATIONAL ENERGY AGENCY: *Medium Term Coal Market Report.* www.iea.org/topics/coal/ (downloaded: 17 05 2105)

[7] SCHUMACHER, E. F: *Small is Beautiful*. Budapest: Katalizátor Kiadó, 2014.

[8] STERN, N.: *The Economics of Climate Change: The Stern Review.* Cambridge: Cambridge University Press, 2007. DOI: https://doi.org/10.1017/CBO9780511817434

[9] LOMBORG, B.: *Cool it: the Sceptical Environmentalist's Guide to Global Warming.* Budapest: Typotext Kiadó, 2008.

[10] LOVELOCK, J.: *The Revenge of Gaia; Why the Earth is Fighting Back and How We Can Still Save Humanity.* New York City: Basic Books, 2007.

[11] HAUBER Gy.: An Oversized World – Or the Economic Effects of Global Warming. In. *Tudományos Évkönyv,* Budapest: BGF, 2008, 337–345.

[12] PATAKI Gy. et al.: *Nem növekedés-központú gazdaságpolitikai alternatívák: A fenntartható életmód felé való átmenet szakpolitikai lehetőségei.* Budapest: NFFT, 2013.

[13] KSH: *Magyarország 2011*. Budapest: KSH, 2012. www.ksh.hu/docs/hun/xftp/idoszaki/mo/mo2011.pdf (downloaded: 16 05 2013)

[14] NFM: *Nemzeti Energiastratégia 2030.* Budapest: Nemzeti Fejlesztési Minisztérium, 2012. www.kormany.hu/download/e/19/40000/Energiastrategia.pdf (downloaded: 22 04 2013)

[15] KSH: *A fenntartható fejlődés indikátorai Magyarországon.* Budapest: KSH, 2012. www.ksh.hu/docs/hun/xftp/idoszaki/fenntartfejl/fenntartfejl12.pdf (downloaded: 21 05 2013) [16] INTERNATIONAL ENERGY AGENCY: *Energy Policies of IEA Countries – Hungary, 2011 Review.* Paris: International Energy Agency, 2011.

[17] MAGDA R.: A megújuló energiaforrások szerepe és hatásai a hazai agrárgazdaságban. *Gazdálkodás*, 55 (2011), 575–588. DOI: https://doi.org/10.22004/ag.econ.119936

[18] LUKÁCS G. S.: *Megújuló energia – kitörési lehetőség a szegénységből.* Budapest: Szaktudás Kiadó Ház, 2010.

[19] PÁGER B.: A környezetipar szerepe a regionális gazdaság fejlesztésében. In. BARANYI B., FODOR I. (Eds.), *Környezetipar, Újraiparosítás és Regionalitás Magyarországon*. Pécs- Debrecen: MTA KRTK Regionális Kutatások Intézete, 2012.

[20] ESTÓK S.: Megújuló energiák rendszereinek intelligens logisztikai támogatása. *Hadtudományi Szemle*, 7 1 (2014), 13–21.

[21] ESTÓK S.: Hálózatközpontú integrált interdiszciplináris logisztika. *Bolyai Szemle,* XVIII 3 (2009), 23–33.

[22] MACKAY, D. J. C.: *Sustainable Energy – Without the Hot Air.* Budapest: Typotext Kiadó, 2011.

AARMS Vol. 14, No. 2 (2015) 249–258. **10.32565/aarms.2015.2.10**

The Near and More Distant Future Environment of Air Defence Missions

KRAJNC Zoltán[1]

The author presents in this study the near and more distant future environment of air defence missions. He tries to find adequate answers to three questions in order to avoid prospective developments causing disadvantages for air defence experts and troops realising air defence. These are proliferation, the quick development of air warfare vehicles, and the widening airspace threats represented by civilian vehicles. In brief, air defence as defence activity is always at a disadvantage against the threats arriving from the air, but we must do our best to compensate.
Keywords: *air defence, future threat, asymmetrical threats arriving from airspace, determinants of the future operational environment*

Introduction

*"*It is very hard to predict, especially when we speak about the future*"*, said once Nils Bohr, the famous scientist. This truth is experienced also by researchers who intend to predict the future of our security environment and the character and internal content of the wars (armed conflicts) of the future. Despite the difficulties, the cognition of the probable future and the creation of valid future images have great importance in the field of military affairs for the decision-makers.

Today there are scientifically based research papers using so many techniques to describe possible futures; however, we have to recognise that future research can be very "multi-layered" and can be only sustained with certain plausibility derived from present facts. The problem is very complex due to the fact that since the 1980s the relative stability of international relations has been replaced by very complex and difficult situations. Consequently, prognoses, circumstances and conditions of forecasting possible future scenarios have changed. Future research of scientific quality has to face a new situation: in society, in several fields of political and economic life, as well as in the field of military affairs, the stability and balance that existed earlier are becoming less and less characteristic.

The events and conditions standing for instability can be experienced in more and more fields. According to certain experts we have to calculate the duration of instable periods and take into account the fact that unexpected changes will take place more and more frequently, to the extent that instable conditions may become permanent. [1]

However, if instability is present in parallel with stable processes the culmination points induce epic changes. This means that in the security environment in addition to the old, fading processes, new ones are spreading. A spectacular example nowadays is Iraq, where

1 Lieutenant-colonel Krajnc Zoltán Ph.D., e-mail: zoltan.krajnc@gmail.com

traditional phenomena (weapons, military techniques) belonging to a much earlier generation of warfare are present in parallel with the American warfighting culture using the most modern military technology. Guerrilla warfare based on terrorist actions goes together with network-centric warfare and effect-based operations. [2]

This complex situation induces implies that we will be forced to apply newer and newer methods beside the traditional methods of future research in the description of the security environment of the future. In the future research in security policy and in the forecast of the form and internal contents of prospective wars the governmental and non-profit research centres, also called "think tanks" of the United States are at the forefront. There have been some positive changes to level out American professional dominance in this field, as within the framework of the Allied Command Transformation, not long ago so-called centres of excellence.[2] The main task of these centres is to provide future image and vision for NATO for the purpose of well-balanced medium- and long-term planning. [3]

Military thinking and decision-making processes have always included the claim to be able to do future research and make forecasts. It is enough if we have a glance at operational planning, the sequence in the commander's decision-making process, and we can establish that the prediction of the course of action of the enemy is also based on prognosis-making processes integrating elements of general evolution theory and scenario-making methods. This can be observed much more in the elaboration of medium- and long-term plans of military development. Thinking further about the above mentioned analogy, in the case of planning a longer campaign or a complete war, the necessary decisions can be prepared only based on a future image elaborated with scientific thoroughness. [3] [4]

The analysis of the security environment, the time horizon of the examination of potential threats and trends usually result in studies of medium-term (5–10, perhaps 15 years) and of much longer term (even 20–40 or 50 years) forecasts. [3]

We can encounter the results of the analyses in shorter terms in the first chapters of security strategies, strategic conceptions and allied and national doctrines of integrating character, where the main messages are explained further. Analyses of such terms, usually, if executed by strictly scientific methods are usually made with the use of classical prognosis ("forecasting"). This prolongs future based on the tendencies of the near past and on the experiences and phenomena of reality available up to now. [3]

Here we present the most important elements of future images (strategies, official studies summarising trends, etc.) based on the above summarised methods of future research.

2 The first one of such research centres is the Joint Air Power Competence Centre (JAPCC) established in Germany, whose main assignment is long-term forecast, to discover the trends of air and aerospace warfare, and to contribute to the elaboration of the strategic conceptions of NATO.

The Basic Change of the Strategic Environment: Asymmetrical Threats Arriving from Airspace

The air component responding to the succinctly summarised trends of threats is basically constituted to counter threats arriving from the air. The development of air assault equipment and the process of their proliferation will be explained further in detail. [3] It is the increasing freedom of the transfer of the technology that makes it possible for more and more states and even groups to produce harmful devices, not to speak about the dangers caused by the arms trade.[3]

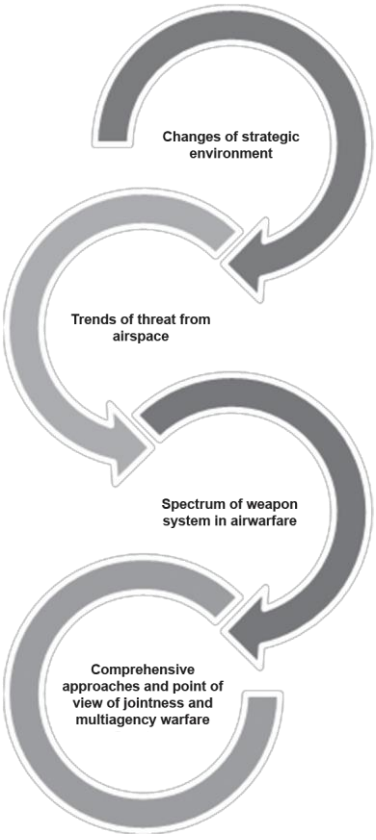


Figure 1. The major determining factors of environment of air defence. (Edited by the author.)

A Terror threat coming from the air cannot be ignored, especially after the attacks against the World Trade Centre and the Pentagon.

3 Last year the news reported an incident when the Spanish Navy "caught" a ship transporting the main elements of ballistic missiles along the Spanish coasts. The addressee was probably one of the Middle-Eastern states or terrorist organisations.

The increase of the terror threat coming from the air can be observed even today, since well-organised terrorist groups after thorough planning, are able to consider every factor. They also may acquire destructive weapons and carriers that are able to cause considerable damage (poisoning, infection, psychical pressure, etc.). They can also use legally available aircraft or civilian aircraft accessed on the ground or hijacked in the air whose application for destruction mean a serious threat.

The situation is exacerbated by the fact that these criminal groups may be able to install warheads (poison, explosive) in "small aircraft" and execute terror attacks with them against appointed targets. Apart from the small aircraft platform, it is imaginable that they will acquire more serious technologies and use helicopters, drones or, according to the darkest version, ballistic missiles. The complexity of such a terror attack and the difficulty of the defence against it are shown by the fact that attacks can be initiated not only from beyond the borders, but also from within the borders of a given country, which is very disturbing as for the calculation of the reaction time.

Airspace Challenges Deriving from Regional Instability

Military challenges and threats in the classical sense have decreased however, instability is a possibility for the appearance of many forms of air threats. The platform, the targets and the destructive material can be similar to the above mentioned: the most likely is the use of small (sport) aircraft.

Certainly we do not only have to account for terror attacks from distant regions, since organised crime (drug related, weapon smugglers, fissile material and refugee smuggling, etc.) and the violation of the international rule of aviation (e.g. its "Balkanian"-type application) can also endanger the security of the airspace of the Western world. [4]

The very quick spread of mass destruction vehicles and their carriers and the technological transfer of their production mean a potential danger to our homeland. It is a well-known fact that the most important terrorist organisations have already made attempts to acquire fissile material, and such materials may be and might have been smuggled out of some successor states of the Soviet Union.

As for our homeland, air threat endangering the Hungarian contingent during military operations outside the area of the country is a very important element. Certainly, it cannot be separated from the threat endangering the coalition forces, since it is an organic part of this threat. Furthermore, the participation of our homeland in such operations can also induce attacks against Hungarian targets from state or sub-state parts of the conflict, which can be very dangerous if the given state (group, organisation) owns nuclear weapons or is even able to produce "dirty bombs".

Nowadays we speak very little about the security risk that evolves with the appearance of portable (portable by man) air defence missiles in illegal hands, and its application against civilian or even military air vehicles or the threat of which can cause very serious damage. [4]

The Spectrum of Air Raid Vehicles as One of the Main Determinants of the Future Operational Environment of Air Defence

Apart from the envisioning of future security challenges, technological and proliferation trends, the making of prognoses can more and more come to the forefront in military future research. This may be the basis for the elaboration of potential air threat alternatives in the preparation of the development of the air force. Although the main assignment of air defence has not changed in the last decades, the spectrum of its threats has considerably been restructured, changes of emphasis have evolved. The earlier understanding of air defence, well-known from the past, can be described as the classical air defence and its version that developed further in an evolutionary way, into "extended air defence".[4]

We can understand why this fairly new concept appeared in the theoretical system of air defence, what is behind it and what qualities it contains through the review of the main categories of air raid vehicles and their tendencies of development. The spectrum of air threat is constituted by air warfare vehicles that we can calculate as potentially dangerous to the objects to be defended. It is practical to analyse, besides the dangers deriving from the application of fixed-wing and rotary-wing aircraft meaning traditional threat with risks deriving from the application of ballistic missiles, manoeuvring robotic aircraft, unmanned aircraft and surface-to-air missiles.

The quick technological development, the harmful proliferation of weapons of mass destruction and their means of delivery can be realised with fairly cheap technology. While in the past both the centre of gravity and the focus of air defence were aimed at the armed air vehicles, aircraft and helicopters controlled by a pilot, nowadays more and more attention needs to be paid to pilotless air vehicles, ballistic missiles and other air vehicles. The attached diagrams make it clear that certain air warfare vehicles mean serious threats. Compared to the prizes of the development and maintenance of fixed-wing aircraft based on high technology, it becomes clear that recent threats can be produced and deployed en-masse at a fairly cheap price thus giving a serious challenge to air defence. [5]

Fixed-Wing Aircraft

Enemy aircraft is the most classical air threat for air defence and the objects to be defended, and they will probably remain a very important element of the air threat spectrum in the future.

We classify the following categories as fixed-wing aircraft: *bombers, fighter-bombers, fighters* and *reconnaissance, cargo and battle aircraft.* According to the newest trends of technological developments air forces do their best to deploy multifunctional, universally applicable aircraft.

Due to the wide range of the applicability of aircraft, they are applied in each segments of air warfare, and they can execute *strategic air operations, attack and defensive counter air operations, air interdiction, close air support, electronic warfare* and *air reconnaissance.*

The designers of air defence can count on enemy fix-winged aircraft attacking targets of big value, like sea harbours, airports, platoon groups, air defence artillery units, command and control centres and logistical bases.

4 *Extended Air Defence – EAD*

The weaponry and ordnance of fixed-wing aircraft provide a very wide spectrum of application as they can carry machine-gun, automatic cannon, guided and non-guided surface-to-surface missiles, air-to-air missiles, reconnaissance and electronic fight containers. Some of them are also able to launch robotic aircraft the air defence must also take into account. [6: 2–8]

Helicopters

Since the Korean War helicopters have been playing an increasing role in warfare. In the beginning they were applied for reconnaissance and artillery fire control, later for the transportation of troops and material. Today they have become usable in the full spectrum of air warfare, as a result of technical and warfare specialisation.

This fact was displayed by the first air raid of the Gulf War in 1991 in which the most important radar points of the Iraqi air defence forces were eliminated with AH–64 Apache helicopters thus paralysing the command and control system of the air defence, with which they decisively contributed to the success of the operation to neutralise the air defence forces.

Attack helicopters were applied by the French for the first time in the Algerian Crisis in 1959. After that the theory of its application, warfare with attack helicopter, started to evolve, and this process was justified by the "attack, transport and armed helicopters" applied en-masse in the Vietnam War. Due to its versatility, survival, and anti-tank ability, helicopters are ideal air vehicles to support warfare and operations of ground forces. In most countries we can consider attack helicopters a cheap, but very effective alternative and the so-called armed (armed, multirole) helicopter the main asset of direct air support for troops.

Another indispensable battle element of mobile air troops and air descent operations is the transport (multi-role) helicopter with which it is possible to transport the forces to the target area, and like this the (special) forces penetrating in the depths of the enemy can disrupt the command and control system of the enemy, occupy objects of key importance and destroy warfare elements of decisive importance.

Search and rescue operations in war time, peace time and air evacuation (the transporting of the casualties home) nowadays cannot be imagined without modern helicopters. With the help of the aircraft with special equipment the air force is able to rescue the crew of the aircraft (helicopter) and to transport the casualties off the battle field.

The development of air defence helicopters is also a present-day development trend, with which the annihilation of aerial vehicles of low velocity (helicopters, aircraft, balloons) is planned.

Military experts recognised early that helicopters can be a prominent vehicle of warfare; therefore it is equipped with a wide range of board weapons and ordnances. It can bear the following weapons (or interior or exterior ordnances): machine gun, automatic cannon, guided and non-guided missiles, grenade launchers, air-to-air missiles, reconnaissance and electronic fight containers.

For ground-based air defence systems the application of helicopters can mean a serious threat as the vehicles attack in pairs, exploit the camouflage opportunities provided by the environment, and are able to approach the target area flying close to the ground. The annihilation of warfare vehicles executing operations mean a very great difficulty to air defence systems operating on the principle of radial velocity measurement (applying Doppler-frequency). The combat against helicopters is supported by the fact that their application depends on the weather and the given time of day, and its relative loudness makes it easier to locate it with traditional devices. [6: 2–6]

Ballistic Missiles

Operation Desert Storm of the Gulf War provided the designers of air defence weapon systems and air defence operations with a very important conclusion. The Iraqi soldiers, despite the relatively small effectiveness of warfare aircraft, applied ballistic missiles and robotic aircraft attacking ground targets with relative success. The experiences of the war revealed what challenges the Republic of Hungary and its allies must face in the future.

The high expenditures of the development, maintenance and application of fixed-wing aircraft and the theoretical possibility of losses to the air forces of NATO member states makes it very attractive for several countries to want to acquire ballistic missiles. The press often calls the missiles a real threat in potential contemporary crisis and describes them as "the long stick of poor countries". Their danger is certainly in their relatively low costs (0.5–1 million USD) and the relative simplicity of the manufacturing technology. This weapon system that can be considered the successor of the V–2 rockets of World War II was developed by the Soviets in the 1950s (under their NATO code name: SCUD), and several of its modifications became widespread throughout the world.

The ballistic missiles can cause destruction to an unacceptable degree by aiming at large population centres, or by the application of non-traditional warheads. Furthermore, it cannot be ignored that the deployment of the missiles or the threatening of their application can have severe psychological effects that were also proved by the Gulf War, when the Iraqi government was threatening Israel with SCUD-missiles.

The ballistic missiles can be applied as vehicles of "first strike" or of the so-called "revenge attack". Their velocity and the variability of their launching make them suitable for unexpected attacks. The ballistic missiles make it possible for the applying countries to project their military forces long distances, beyond their borders, with which they can also re-structure the battle fields, since commanders must use considerable resources to neutralise the rocket threats. It also means a separate problem that the majority of possible aggressor nations in potential crisis areas own weapons of mass destruction together with ballistic missiles as means of delivery.

The vertical launching systems have a high degree of survival ability, especially when they are applied in darkness or at low visibility. The launching vehicles of great mobility are able to occupy hidden launch facilities, making it very difficult for forces and vehicles to locate and annihilate them. The surviving ability of the launching vehicles means that in the case of a possible crisis management or of an operation outside NATO area the rocket threat can stay permanent during the whole involvement.

The warfare ability of these missile systems, due to their low accuracy and precision, is not very relevant; however, we cannot forget about their low range, considerable survival ability, their ability to carry weapons of mass destruction, and their psychical pressure on the population. [6: 2–1]

The defence against ballistic missiles can be really effective only if an active air defence fight is completed with the activity of traditional (special) forces assigned to locate and annihilate the launching vehicles. If the Republic of Hungary would like to meet the challenges represented by ballistic missiles, then it should think about acquiring a missile defence system and the adaptation of the theory of application.

Cruise Missiles (Manoeuvring Robotic Aircraft)

The roots of present-day manoeuvring robotic aircraft, similarly to ballistic missiles, can be found in the German military industry of the Second World War. The V–1 rocket of the Wehrmacht and the Okha type robotic aircraft of the Japanese army served as the basis for the post-war American developments. This vehicle was in the beginning used as air defence deployed by the Air Force and the Navy in the 1950s. The development of robotic aircraft continued, and the vehicles launched from ground-, air- or sea-based systems were developed for strategic (nuclear) attacks during the bipolar world order.

We call robotic aircraft those vehicles that are equipped with wings of different types and categories, stabilizers and engines utilising the oxygen of the atmosphere, flying along a programmed orbit or flying with remote control. Usually, robotic aircraft approach the target to be annihilated by flying along the surface (e.g. with TERCOM navigation) during the whole period of the operation, or in the beginning they fly along a ballistic orbit then return to the airspace flying along the surface.

Robotic airplanes – within the class of the tactical rockets, together with pilotless airplanes and air-to-surface rockets – are classified as so-called "tactical aerodynamic rockets". As for the development of robotic aircraft, besides the application of more and more modern technology, we can also observe a trend according to which it is not the vehicles of high technology that are being developed, but similarly to the "weapon of the poor people", the ballistic missile. First and foremost with the transformation of anti-ship vehicles and with the use of various GPS-based devices available off the shelf, cheap weapon systems of limited warfare capability are created.

The limited ability of accuracy and precision concerns the probability of reaching the target. However, the importance of the weapon system is given by the fact that in possession of this type of relatively cheap vehicle the power relations of the region can be changed, not to mention the fact that robotic aircraft are even able to carry weapons of mass destruction and have a survival ability of great degree.

As for the size, the form, the mode of control or the types of weapons carried by them there is a very wide variety of manoeuvring robotic aircraft in use, and their range is from 30 km up to 1,000–3,000 km. After the short-range anti-ship versions several countries are developing manoeuvring robotic aircraft capable of attacking ground targets, and in their control GPS-based satellite navigation technology can also be applied. The aim of the development of the future technologies of manoeuvring robotic aircraft is to increase the range and improve the accuracy of these vehicles.

Usually the cruise missiles are planned to be applied in the destruction of non-mobile targets that do not have strategic importance, and similarly to ballistic missiles they play a serious psychological role and can maintain permanent threats in crisis areas. [6: 2–3]

Unmanned Air Vehicle (Pilotless Aircraft)

We classify the unmanned air vehicles of remote control and the programmed unpiloted vehicles, the so-called drones, as pilotless aircraft. These vehicles can be characterised by the fact that they are easy and cheap to acquire and produce, and can be used in a number of ways. They can be applied for reconnaissance purposes, electronic warfare, for strikes

against ground (water) targets, and they can participate in the neutralisation of air defence systems and can be involved in artillery fire control and deception.

It is a great advantage of pilotless aircraft that it demands much less crew training than aircraft, and its application increases the morality of warfare by the fact that the missions do not risk the life of the pilots.

Due to its small size, its effective radar wave reverberating design, its low flying profile, it is a very difficult assignment for air defence systems to annihilate pilotless aircraft. [6: 2–5]

Air-to-Surface Missiles

Tactical air-to-surface missiles – similarly to robotic aircraft and pilotless aircraft – belong to tactical aerodynamic missiles. According to their control systems they can be radio-guided, self-guided against radars, half-active laser guided, electro-optical and TV-guided. Due to their large velocity and small effective radar cross section these targets are also very difficult to locate, follow and annihilate.

We have to especially deal with the anti-radar missiles to be applied against radiating radio-technological devices (radars), which are the most important equipment for the neutralisation of air defence systems, because it means a very serious threat, since with its application the enemy can raid our reconnaissance and fire control radars from a relatively great distance and from outside the fire zone. Anti-radar missiles can have a very destructive effect if they are applied together with deceiving pilotless aircraft used for the activation of radars. [6: 2–4]

Concluding Remarks

As we have already written in the beginning of our article, it is very difficult to "predict", but as Kossuth Lajos said: "The past is the mirror for the future." We cannot ignore and neglect the elaboration of longer-term prognoses. We must consider our possibilities and abilities, and deal with the prospective tendencies and the problems indicated by them. We have to find the adequate answers to three questions in order that prospective development should not cause drawbacks to air defence experts and troops realising it. These are proliferation issues, the quick development of air warfare vehicles, and the widening airspace threats represented by civilian vehicles.

As for the proliferation of mass destruction vehicles, the problem first and foremost requires a non-military solution. It is a political, diplomatic and economic effort that is able to hinder and slow down this very dangerous and harmful process, even if it is not able to stop it. The attention of state and economic decision-makers must be turned to this circle of problems, and we should participate in the hindering of proliferation.

The search for the "air space answer" to the quick development of air warfare vehicles is nowadays on the beam. The building of early reconnaissance and forecasting systems provides a suitable basis for the maintenance of active and effective ground- and air-based air defence systems. At the moment it has limits of expenditures rather than technological limits. The extension of the local air defence systems into regional systems in time and space and via the increase of security level can be effective. The alarm services and situations of suitable quantity and quality can provide sufficient defence, even against surprise-like aggression.

Air defence is applied, not element by element, but strictly as an integrated system. Together with all necessary components it can give a suitable answer to current military challenges and threats.

Nowadays the applicability of civilian aircraft as air raid vehicles represents the most dif- ficult problem. There is no perfect, effective solution against terrorist operation in every sit- uation. The possibility of acquiring small civilian aircraft, as mentioned above in the article, and their configuration into weapons is unfortunately too simple and not expensive at all. The restriction or ad absurdum the complete prohibition of private aviation is not the best solution in the same way as the building of the continuous and direct air defence of the threatened targets and objects. The first solution would generate nearly unsolvable legal problems, while the second would mean unendurable expenditures. In these fields, the air defence seems to be the "loser". The available military vehicles and possibilities do not and cannot grant a satisfactory solution, mainly due to the problems caused by the very narrow reaction time. In the same way as at the dangers caused by proliferation, in this field there is also state inter- ference and regulation necessary, completing it with strictly regulated military contribution, and as a final step, with the annihilation of suspicious or obvious enemy air targets as soon as possible. Undoubtedly, 11th September, 2001, redefined many things for possible activity against "civilian" air vehicles.

To sum it up, air defence as defence activity is always at a disadvantage against the threats arriving from the air, but we must do our best in order that this disadvantage decrease.

References:

[1] *Magyarország Nemzeti Katonai Stratégiája*. www.kormany.hu/download/a/40/00000/ nemzeti_katonai_strategia.pdf (downloaded: 13 04 2015)

[2] JOBBÁGY Z.: Clausewitz and the Gestalt of War. *AARMS*, 12 1 (2013), 117–128.

[3] NOVÁKY E.: *A jövőkutatás módszertani megújulása*. https://miau.gau.hu/miau/05/jkut.html (downloaded: 13 04 2015)

[4] JOBBÁGY Z.: On the Shifting Balance of Command and Control: A Tale of the Genes and Xenophon. *Science and Military*, 6 1 (2011), 28–34.

[5] IFPA: Air, Space, and Cyberspace Power in the 21st Century. *38th IFPA-fletcher Conference on National Security Strategy and Policy*, Washington, D.C.: Ronald Reagan Building and International Trade Center, January 20–21, 2010. www.ifpa.org/pdf/USAFreportweb.pdf (downloaded: 13 04 2015)

[6] TÓTH S.: A képesség tervezés nemzetközi és nemzeti aspektusa. Kihívások, fenyegetések, az erőszak kezelése, erőszak alkalmazása. Képesség alapú légierő és a képességfejlesztés nemzeti és szövetségi feladatainak összehangolása, *Repüléstudományi konferencia*, Szolnok, 2005.

[7] *FM 3–01 (FM 44–100) U.S. Army Air and Missile Defense Operations.* https://info.publicintelligence.net/USArmy-AirMissileDefense.pdf (downloaded: 13 04 2015)

[8] *FM–44–100 Air Defense Artillery Reference Handbook, 2000*. Final Draft.

[9] KRAJNC Z.: A légierő alkalmazásának alapkérdései. *Új Honvédségi Szemle*, 7 (2005), 87–98.

[10] KRAJNC Z.: *A légierő megváltozott szerepe a XXI. század hadviselésében.* Budapest: Geopolitikai Tanács, 2006. (kutatási jelentés)

AARMS Vol. 14, No. 2 (2015) 259–264. **10.32565/aarms.2015.2.11**

Advantages of Water Fog Use as a Fire Extinguisher

KUTI Rajmund[1]

Due to the consequences of global climate change and complex impacts of civilization, the Earth's drinking water reserves have started to decrease or become so seriously contaminated in some areas that it is inadequate for human consumption. International research groups and organizations are calling the countries to greater environmental and security awareness. Severity of environmental protection regulations made the usage of environmental friendly fire-fighting materials (water, among others) more important again. In Hungary, fire brigades use drinking water in approximately 90% of their fire-fighting operations. Some recently developed water fog fire extinguisher apparatuses have made special utilization of water available, as their use is friendly to the environment, effective and economical and in comparison with the conventional fire-fighting techniques consume a considerably less amount of water. In my paper I am going to demonstrate and summarize the advantages of water fog fire extinction based on the physical and chemical properties of water.

Keywords: *water, water fog, fire extinguishing effects, effectiveness, economy, environmental protection*

Preface

Environmental protection aspects and environmental awareness have become more and more important in the last few years even in the field of fire-fighting. After the elimination of the halons[2] as an extinguishing agent, research and development of new fire-fighting technologies and their practical implementation have become necessary. The use of water has come to forefront again, but not in the conventional way but in water fog form. Water fog extinguishing systems make the use of water available in a special way making it more effective and non-dangerous to humans and the environment.

Fire fighters recognized during the development of fire extinguishing that the heat draining capability of water can be increased with the use of a cloud of small water particles instead of a closed water mass on the burning material. With the use of improved fire extinguishers, pumps and fire hose directors it is possible to get vaporized water at a appropriate pressure to extinguish fires more effectively.

It has been well known how to generate gentle fog-like water spray using different nozzles for a long time now, but in most cases its kinetic energy is not enough, despite the proper

1 kutir@mtk.nyme.hu

2 Halon: Haloalkan, or halogenoalkane, a group of chemical compounds consisting of alkanes with linked halogens. Halon is a liquefied, compressed gas that very effectively stops the spread of fires by chemically disrupting combustion. Production of halons was internationally ceased from 1994 due to their degrading effect on the ozone layer.

particle sizes, so effective injection of the water spray into flames is not an easy task. Smaller water particles can be easily driven by the hot, upstream gas flow or vaporized by the heat radiation too early at the periphery of the flames, so their fire extinguishing effect is not generated.

Physical and Chemical Properties of Water Fog and Its Fire Extinguishing Effects

The key to fire extinguishing effectiveness of water fog lies within the appropriate kinetic energy of its extremely small particles. "Water fog" is the name of the aerosol, if it contains particles with a diameter of 1,000 mm or bigger less than 1% of the water, measured in 1 meter distance from the spray nozzle. Sizes of water particles should be checked with a laser measurement instrument at least in 24 different points in the aerosol. [1]

Fire extinguishing effect means formation of circumstances that could restrain or inhibit combustion or its formation. Water fog is an extremely useful extinguishing agent that can be used in a wide range of applications, except some special cases, where it is forbidden to use water. Its effectiveness comes from its fire extinguishing capability and from more different aspects. A series of experiments prove that different fire extinguishing effects appear at the same time using water fog.

When using water fog for fire extinguishing, water particles in a combustion zone should be as small as possible, this way maximizing the surface of the extinguishing agent, so the cooling effect is greater and the necessary amount of water is much less compared to traditional fire extinguishing methods using water flush.

It is possible to generate water fog in several pressure ranges using different spray nozzles, but there will be differences in the sizes of water particles in the aerosols. And particle size affects the active cooling surface. [2] Sizes, numbers of water drops and an active cooling surface related to 1 liter of water can be seen in Table 1.

Table 1. Active surface is growing as the particle size decreases. [5: 1–7]

Size of water drops (mm)	Number of water drops	Active surface related to 1 liter of water (m^2)
10	1,900	0.6
1	1,900,000	6
0.1	1,900,000,000	60
0.01	1.9E12	600
Getting down a magnitude order		
x0.1	**x1,000**	**x10**

Cooling Effect of Water Fog

During the combustion process the majority of the liberated heat energy is passed from the combustion zone with heat conduction, heat radiation and heat convection. The remaining energy is necessary for further combustion that is 8–16 % of the original. If we can decrease the temperature below the burning point (or flash point, in case of a flammable liquid) delivering an extinguishing agent into the flame zone, there will not be enough energy in the combustion area to maintain burning. The most successful way to do this is to drain heat by evaporation. In this case, part of the water fog becomes steam in the combustion zone. Due to the upthrust generated by the evaporation the convective stream of water vapor leaves the combustion zone and becomes liquid again emanating the transported heat in a distance. Size of water particles is very important in this case, because processes of heating till boiling point and evaporation happen on the surface.

Small water drops that remain in the combustion zone in liquid form have a further cooling effect due to heat drain. In this case efficiency of cooling is worse than by evaporation. So it really matters what the ratio of evaporation and simple cooling is during fire extinguishing. [2] Very small water particles mean large specific surface but are also useless, because larger fires have such intensive gas flow that the extinguishing agent cannot enter the combustion zone deep enough, so fire extinguishing is not sufficient.

Investigation of cooling effect revealed that vaporization of water at high pressure results in a much better cooling surface compared with the use of traditional, so called sprinkler devices. Intensive cooling means an advantage not only in fire extinguishing, but also by draining heat this way and protecting persons and assets on site from dangerous warmth.

Oxygen Squeeze out Effect of Water Fog

Small water particles in fire quickly become steam. This procedure happens in the flame zone on high temperatures. On lower temperatures, for example on evacuation routes, steam does not form. It is a great advantage during the use of water fog, because presence of oxygen is essential for a safe escape of personnel without SCBA.[3]

Water drops coming from traditional sprinklers evaporate much slower. Ideally small particles of water fog can result in a 300 times quicker evaporation speed. Evaporation and steam formation happens only in high temperature areas and the volume of water can increase up to 1,750 times higher than in liquid form, this way squeezing out oxygen from the combustion zone. Evaporation during combustion can help the entering of appropriate water fog particles into the flames. To stop fire it is necessary to decrease oxygen concentration on ambient pressure below 16%. In this case water fog works very similarly to inert extinguishing gases, nitrogen or argon (inertisation effect). [3]

With the use of water fog, burning material can be covered with fog first, than with a steam cloud, blocking the entrance of oxygen into the combustion zone. This extinguishing effect becomes more obvious with the use of water fog than with traditional water use.

3 SCBA – self-contained breathing apparatus

Inhibition Extinguishing Effect of Water Fog

The procedure of burning at high temperatures is a very complex phenomenon. In addition to oxidation, it is a chain reaction with the key presence of activated, free ions and radicals with a very short lifetime. In case of vaporization at a high pressure, ions are also forming from water molecules. These ions can recombine with other positive and negative ions and free radicals present in the combustion zone. This recombination can suspend the chain reaction of burning. In this case the process is called *homogeneous inhibition*.

New research and practical experiments showed that there is also a *heterogeneous inhibition* using water fog. This means that water particles entering the combustion zone are acting like a wall simply blocking the chain reaction of burning to pass on, so the fire stops. The inhibition extinguishing effect depends on the result of vaporization.

Hitting Effect of Water Fog

Universal fire hose directors or fire extinguisher lances of mobile water fog fire extinguishers working with combustion engines are suitable to create concentrated water fog jets. Water fog coming with great kinetic energy can rip the flames off the burning material this way disrupting the surface of the fire. A concentrated water jet is uninterrupted, has a small diameter and high speed even with the use of water fog, so it has a great hitting power and medium range. Using a concentrated water fog jet it can pass the flame zone with lower efficiency, but still can be used to disrupt smaller fires. Nevertheless, its complex extinguishing effects can not show up totally, because its surface in contact with the fire is too small and the contact time is too short.

Possible Fire Extinguishing Applications of Water Fog

Basically, water fog can be used in case of fires of all flammable materials, where water is allowed, but with special attention during wintertime to the danger of freezing.

It was confirmed by experiments that even fires of charged electric devices can be extinguished with the use of water fog with appropriate high pressure. Lately, built-in water fog extinguishers (as substitutes for halons) were designed and installed into electric stations, command rooms, digital server centers and telephone exchanges. [4]

A disadvantage of water fog use for fire extinguishing is that some chemicals can react with water and an explosion can be a result of the reaction. Alkali metals and earth metals are in this group such as sodium, potassium or magnesium, and also their carbides and hydrides.

Other sources of danger during the use of water fog can be that at a high temperature, for example in case of extinguishing metal fires, thermal dissociation of water can happen. High temperature causes degradation of water to its components, gaseous hydrogen and oxygen forming detonating gas ($2H_2 + O_2$, oxyhydrogen) this way, which can fuse to become water again causing an explosion. Of course, this cannot happen with the use of built-in water fog extinguishers, because it is forbidden to install them in such an environment. In case of mobile water fog extinguishers, they must not be used to extinguish metal fires.

Basically, high surface tension value of water is not advantageous for fire extinguishing. This problem arises in case of water fog also, but only marginally. Compared with the traditional utilization, using water fog means a greater surface for the same amount of water, so surface tension is a smaller problem. [5]

A well-known phenomenon is that liquids in gases or in other liquids, which are immiscible with them, try to be in spherical form because it has the minimal surface for a given volume. This way surface tension is a kind of resistance against the force that tries to enlarge the surface of the liquid. So surface tension is a kind of force that affects within the length of the surface and tries to decrease its area. Its dimension is Newton/meter.

During fire extinguishing, water drops try to form spheres to minimize their surfaces, that is why they contact burning material's surfaces minimally, so they are not wetting the surface of the burning material adequately. That is why surface tension of water must be decreased by adding some wetting or moisturizing agents. These agents change the ratios in the cohesion forces, so the force between burning surfaces and water drops get bigger than among the water drops themselves. This way water drops can stick to burning surfaces more easily, water can even impregnate porous bodies, increasing evaporation and cooling. During fire extinguishing it is usually realized by mixing a foaming material in 1% of the water amount flowing thru the pump.

In Table 2 you can see the utilization possibilities of normal water jets, water fog and different extinguishing materials.

Table 2. Different fire extinguishing materials and their utilization possibilities. [2]

	A	**B**	**C**	**D**	**E***	**Notice**	**Versatility**
Extinguishing materials							
Concentrated water jet	+3	–2	0	–2	–2	Damages caused by water	1
Diffused water jet	+2	+1/0	+1/0	–1	–1	Smaller damages caused by water	1
Water fog	+2/+1	+2/+1	+2/+1	+1	+1	No damages	4
Fire extinguishing foam	+2	+3	0	–1	–2	Environmental damages	2
Flame extinguishing powder	+1/0	+2	+3	0/–1	+1	Contamination, panic because of dust	2
Embers extinguishing powder	+2	+2	+3	+1	–1	Contamination, panic because of dust	3
Metal extinguishing powder	0	0	0	+2	0	Contamination	1
Carbon dioxide gas	0	0	+3	0	+3	Risk of choking	2
Carbon dioxide snow	0	+2	+1	0	+3	Cold shock	2
Halons	+1	+2	+2	–2	+3	Damage to ozone layer	3

Legend:

- **In presence of electric current,*
- *+3 excellent fire extinguishing,*
- *+2 good fire extinguishing,*
- *+1 limited use only,*
- *0 not adequate,*
- *–1 utilization is worth considering,*
- *–2 utilization is dangerous,*

After the comparison of the above data, it can be seen that water fog is a useful, very versatile fire extinguishing agent and at the same time it is environmentally friendly. It should be noticed that in the presence of an electric current only the built-in water fog fire extinguishers can be used with the maximum observance of all security precautions. [6]

Conclusion

Formation of water fog and use of water fog fire extinguishers provide new solutions in fire protection with the combination of positive properties of traditional gaseous and sprinkler extinguishers. The successfulness of fire extinguishing with water fog aerosols based on their water drop size reduction, increased kinetic energy and distribution of drops, and besides the cooling effect on the burning surface the sudden evaporation of water drops results in reduced oxygen concentration in the vicinity of the fire. What is more, in parallel with these, both homogenous and heterogeneous inhibition take place, resulting in a break in the chain reaction of burning. So this environmental friendly and very effective way of fire extinguishing is harmless for humans and causes no water damage faces a great future in fire protection.

References

[1] NAGY Gy.: Vízköddel oltó berendezések, *Védelem*, 1 (2000).

[2] KUTI R.: *Vízköddel oltó berendezések speciális felhasználási lehetőségei és hatékonyságuk vizsgálata a tűzoltás és kárfelszámolás területén*. Budapest: ZMNE, 2009. (Ph.D. Dissertation)

[3] FOGTEC: *The Smarter Way of Firefighting*. Köln: FOGTEC Brandschutz GmbH and Co. KG, 2006.

[4] NÁDOR A.: Vízködös oltórendszerek – nem árt ismerni, mit miért választunk. In. *Védelem Online – Tanulmányok*, 51, 2008. www.vedelem.hu/letoltes/tanulmany/tan51.pdf (downloaded: 05 03 2015)

[5] KUTI R.: Miben rejlik a vízköd tűzoltási hatékonysága? In. *Védelem Online – Tanulmányok*, 501, 2014. www.vedelem.hu/letoltes/tanulmany/tan501.pdf (downloaded: 07 03 2015)

[6] KUTI R., FŐLDI L.: A beépített vízköddel oltó rendszerek újabb alkalmazási lehetőségeinek feltárása. *Hadmérnök*, III 2 (2008), 60–66. www.hadmernok.hu/archivum/2008/2/2008_2_kuti.pdf (downloaded: 07 05 2015)