Combined Mathematical Modeling of Different Transport Networks, Considerations and Complex Analysis

Tamás Péter[1] and Krisztián Szabó[2]

[1] Department of Control for Transportation and Vehicle Systems, Budapest University of Technology and Economics; Stoczek u. 2, H-1111 Budapest, Hungary; peter.tamas@mail.bme.hu

[2] Institute for Computer Science and Control, Hungarian Academy of Sciences, Kende u. 13-17, H-1111 Budapest, Hungary; szabo.krisztian@sztaki.mta.hu

Abstract: On the grounds of our traffic network model developments an exact mathematical model can be created based on the union of different types of transport networks. The initial network is the vehicle traffic network that integrates the public transport network as well. During this study the union of road and pedestrian networks is initially created and then we further examine the possibility of generalization of complex networks thereafter.

Keywords: transport processes; network union model; analysis of movements on mixed trajectories; smart city

1   Introduction

The traffic simulations based on traditional methods start out from trip recording (I) or from classical traffic counting methods (II).

(I) With the first type of modeling we talk generally about traffic generation, traffic distribution and traffic sharing, which is expressed in the form of "from/to matrices". This assessment/forecasting method requires very careful circumspection, as it must be representative, while performing an adequate number of survey according to the representativity can be extremely expensive.

In urban districts modeling many models are known (Lill's law of travel, Stouffer's hypothesis, Detroit method, Fratar method, Furness method, Voorhees model, Intervening Opportunity Model, Competing Opportunities Model, multiple regression model, travel cost model, electrostatic model, other synthetic models) which operate with difficult or impossible to measure growth factors, indexes, empirical exponents, etc.

The input parameters of modeling, therefore, can only be regarded as a benchmark in many cases, thus the product of the simulation can only be considered with the respective significance.

(II) The second type of models are based on junction or cross-sectional traffic counts which employ well-defined methods fixed in road standards and are generally considered to be recognized during transport planning as well. The traffic count is performed according to vehicle types which are expressed in unit vehicle and weighted by multiplication factors.

All these should be carried out at different times of the day in a seasonal manner upon which the daily vehicle traffic, design hour traffic volume, day and night traffic can be calculated [1, 3]. The advantage of simulation software operating with this type of traffic counts is that the input values determined with reasonable certainty may be accompanied by results of appropriate quality which are well understood by the profession [5]. However, the drawback systems using traffic counts is that the relationship between traffic volume, average speed and vehicle density will not give a clear assignment, e.g. several average speed values can be assigned to a certain traffic volume. **The system thus carries uncertainties since, traditional cross-sectional or junction traffic counting methods involve information loss.**

2 The Approach and Methodology used in our Traffic Network Model Developments

A passenger often travels along a complex trajectory during a trip [21]. This condition occurs when the route consists of trajectories from multiple traffic network parts, e.g. in urban traffic the employed bicycle, pedestrian, road and rail traffic trajectories can alter freely according to the needs in the most appropriate way. Of course, this is not only typical of urban transport. Each compound trip takes place on a complex trajectory, which can include, in addition to the above, water and air transport as well.

The choice of a complex trajectory is always a solution of a conditional optimum problem of certain quality, considering the possibilities and travelers' goals. The mixed trajectories play an important role in the modeling of optimal arrivals and in the load and reliability analysis of the complex traffic networks as, well.

The different dynamic traffic network systems operate in parallel following mainly the laws of their own and their control. Their operations are also, effected by the external environmental and meteorological processes [7, 16, 23].

However, the state parameters of these systems have a mutual impact on each other and influence each other's traffic processes as well.

All of these complex relationships justify the examination of the union of different dynamic traffic network systems and the actuality of the topic.

The dynamic network and any of its trajectories in our study has two basic characteristics: the map graph is the *geometrical characteristics* and the type of the flowing "material" is the *dynamic characteristics*. If for two dynamic networks – in the case of this study two transport networks – the two graphs are identical, but the type of the flowing "material" is different we consider them different disjoint dynamic networks. This could be for example the case of a shared pedestrian-bicycle path that can be used by pedestrians and bicycles together, or a public traffic lane used by trams and buses together. In our modeling, therefore, all types of dynamic models describe their own flow. If another flow is present either on the sections, or even as a crossing movement, its state parameters are taken into account in the influence, inhibition and transfer prohibition functions of the model.

The unified network graph is analyzed for different modes of transport. Because of the novelty of the approach it is particularly interesting from the pedestrian and vehicle traffic point of view. The pedestrian traffic graph is also divided into sectors according to our vehicle traffic network model development, on which the passenger facilities of public transport, the intersections with the vehicle network, taxi and shared mobility (car-sharing, bike-sharing) stations create additional disposition points.

The road traffic network model in this study includes the surface public transport network as well. Their joint operation takes place according to the laws of a large macroscopic model. The pedestrian traffic, also, uses this in the unified model "assembly line".

The vehicle traffic network models are to some extent autonomous dynamic macroscopic systems in the large network, based on their own speed-density, control laws. At the same time, the integration with the pedestrian system results in the entry of new dynamics and conditions in its operation as well, e.g. during rush hours, the time spent at stops by public transport vehicles and the dynamics of the vehicles changes.

The pedestrian traffic network model provides more advanced features as well. In case of pedestrians the laws of speed-density functions show up relatively rarely along their own route, but at congestion points they should be considered.

Based on the developed model [6], the optimal network paths, trajectories can be determined. It can be transmitted also to the discussed complex network model.

2.1　Motivations

One of the biggest challenges in the modeling is the alignment of various modes of transport and the mapping of processes of method and vehicle changing points.

The developed new approach was motivated by several basic issues that can no longer be ignored while using ITS network models. The research gives a new direction of targeted basic research in the field of transportation in connection with priority industries [21].

The use of traditional modeling approach raises several unanswered questions and constantly struggles with size problems.

All subnetwork is different, there is a wide variety of specific rules, in this context, studying any part of the network by itself is only a very small part of the whole and it can be only an example taken from the large network.

If only software algorithmic models are used, they are not suited to give a wider range of exact mathematical conclusions and results.

In traditional models the role of the parking places is a problem [22], since they are different types of actors than road sections, so-called foreign elements with "storage properties".

The complexity of the transport processes requires the application of a high-level of automation and intelligent transport systems (ITS), the common basis of which are the transport models [25, 26, 27, 28].

Our new vehicle traffic network model is macroscopic and map graph invariant, it can be described with a special hyper-matrix structure [17]. Its main strengths are the uniformization of the very complex network and the high computing speed.

2.2　The Applied New Model

There are four kinds of connections between the examined internal and the surrounding external network sectors. The dynamic operation (internal and external) of the entire network is systematized by the connection hyper-matrix. The connection hyper-matrix provides information on any sector in terms of with which other sectors are they connected by what kind of dynamic transfer. The system of differential equations containing the connection hyper-matrix describes the operation of all sectors of the network that means the operation of the entire network [18, 20].

Our macroscopic model, which describes the transport processes on large-scale road networks, belongs to the class of positive nonlinear systems. The model is suitable for simulation test of large-scale road networks, planning and traffic control systems [19].

$$\begin{bmatrix} \dot{x} \\ \dot{s} \end{bmatrix} = \begin{bmatrix} \langle L \rangle^{-1} & \\ & \langle P \rangle^{-1} \end{bmatrix} \begin{bmatrix} K_{11}(x,s) & K_{12}(x,s) \\ K_{21}(x,s) & K_{22}(x,s) \end{bmatrix} \begin{bmatrix} x \\ s \end{bmatrix} \qquad (1)$$

In our model we apply $0 \le x(t) \le 1$ and $0 \le s(t) \le 1$ normalized vehicle density state parameters.

$x \in \Re^n$ is the state parameter vector of internal sectors,

$s \in \Re^m$ is the state parameter vector of external sectors,

$\dot{x} \in \Re^n$ is the time derivative of the state parameter vector of internal sectors,

$\dot{s} \in \Re^m$ is the time derivative of the state parameter vector of external sectors,

$\langle L \rangle$ and $\langle P \rangle$ are diagonal matrices containing the length of internal and external sectors, respectively:

$$\langle L \rangle = \langle l_1, l_2, ..., l_n \rangle , \ \langle P \rangle = \langle p_1, p_2, ..., p_m \rangle$$

The submatrices of K connection hyper-matrix are:

$K_{11} \in \Re^{nxn}$, $K_{12} \in \Re^{nxm}$, $K_{21} \in \Re^{mxn}$, $K_{22} \in \Re^{mxm}$ és $x \in \Re^n$, $s \in \Re^m$.

The elements of the matrices are v_{ij} connection speeds which are created as the product of the following seven connection functions:

$$v_{ij} = \alpha_{ij} \ \beta_{ij} \ \gamma_{ij} \ u_{ij}(t) \ S \ (x_i(t)) V_{ij} \ (x_i(t), x_j(t), \underline{e}_i, \underline{e}_j) \ E(x_j(t))$$

In case of the above product at $j \circledR i$ connection $\alpha_{ij} = \alpha_{ij}(\underline{x}(t), t)$ means distribution, $\beta_{ij} = \beta_{ij}(\underline{x}(t), t)$ is a factor obstructing or supporting the connection, $\gamma_{ij} = \gamma_{ij}$ (t) denotes the intensity of the connection and $u_{ij}(t)$ the traffic light signal, which are non-negative dimensionless values.

S and E are internal prohibiting automation functions acting at the connections.

$S(x)$ is an automatic internal self-control function that takes the value 1 or 0. The connection is permitted if the density of the receiving section x is less than 1, otherwise it is prohibited.

$E(x)$ is an automatic internal self-control function that takes the value 1 or 0. The connection is prohibited if the density of the transmitting section x is less than or equals 0, otherwise it is permitted.

V_{ij} is the transfer speed [m/s], which at time t depends on the state parameters of the connected sections and on $\underline{e}_i$, $\underline{e}_j$ environmental parameter vectors of sections i and j.

Based on this the elements of the connection matrix are speed values [m/s].

The values in the main diagonal of K_{11} and K_{22} are 0 or negative, all other elements take non-negative values. All elements of K_{12} and K_{21} take non-negative values. These matrices are, therefore, Metzler matrices, consequently the K connection hyper-matrix defined by them that describes the entire connection system is a Metzler matrix as well [2, 8].

Our macroscopic model of transport processes on large-scale road networks belongs to the class of positive nonlinear systems.

2.3   Mathematical Model of the Narrowed Traffic Network

The narrowed network model consists of an internal network of n sectors bounded by an arbitrary closed curve „G" and m external sectors with densities s_1, s_2, …, s_m, which have direct connections with an internal sector, the state of which is considered as known based on measurements. In this model among the matrices forming the connection hyper-matrix only K_{11} and K_{12} matrices play a role because they represent every transfer with relation to the internal sectors. (The external-external connections are not considered in this case). The system of differential equations of the model is the following:

$$\dot{x} = <L>^{-1} [K_{11}(x,s)x + K_{12}(x,s)s] \qquad (2)$$

Where: $x \in \Re^n$, $\forall$ $x_i \in [0,1]$, (i=1,2,…,n), $\dot{x} \in \Re^n$, $s \in \Re^m$, $\forall$ $s_i \in [0,1]$, (i=1,2,…,m), $L = \mathrm{diag}\{l_1, \dots , l_n\}$, l_i the length of internal sections in the main diagonal ($\forall l_i > 0$, i=1,2,…,n), $K_{11} \in \Re^{n \times n}$, $K_{12} \in \Re^{n \times m}$.

3 The Union of Vehicle and Pedestrian Traffic Network Models

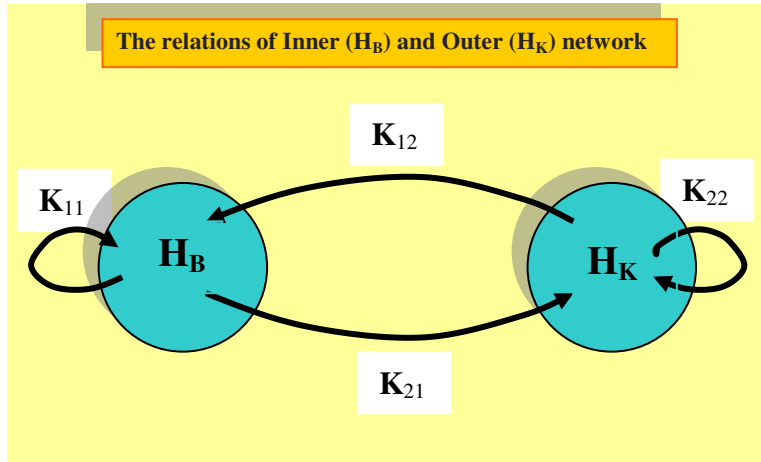The relations of Inner (H_B) and Outer (H_K) network

K_{12}

K_{11}

H_B

K_{22}

H_K

K_{21}

Figure 1

Connections of internal and external networks in case of road traffic

The relations of Inner ($\tilde{H}_B$) and Outer ($\tilde{H}_K$) network

$\tilde{K}_{12}$

$\tilde{K}_{11}$

$\tilde{H}_B$

$\tilde{K}_{22}$

$\tilde{H}_K$

$\tilde{K}_{21}$
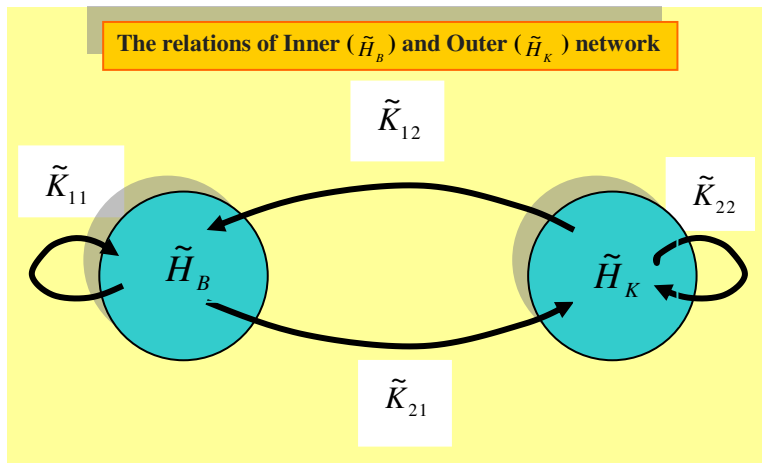
Figure 2

Connections of internal and external networks in case of pedestrian traffic

The developed and used road traffic model examined primarily the vehicle processes. This model is a macroscopic model with arbitrary network dimension.

We show that the model can be extended with appropriate considerations to pedestrian traffic network modeling and analysis of movements and speed processes of pedestrians on the network as well [21].

It must be emphasized that although the organizational structure of pedestrian traffic network model and road traffic model is identical, but the two models form a disjoint network.
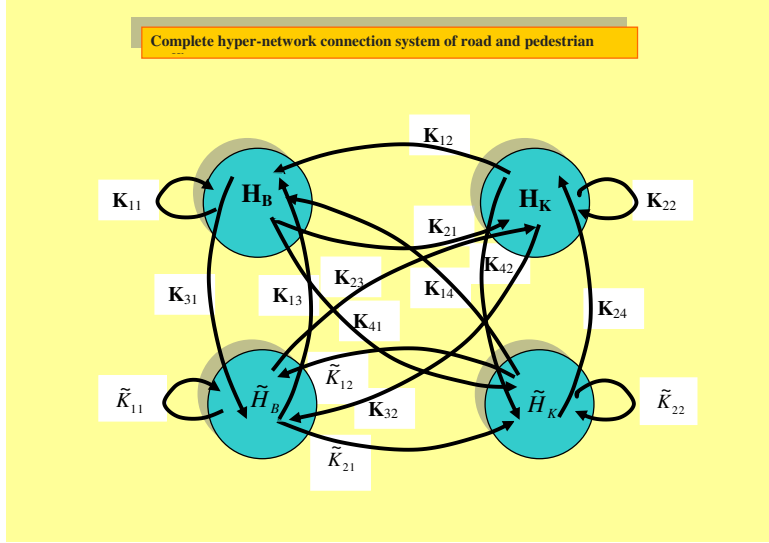

Figure 3

Hyper-network connection system of road and pedestrian traffic

The hyper-matrix describing the entire connection system of the union of vehicle and pedestrian traffic network models is the following:
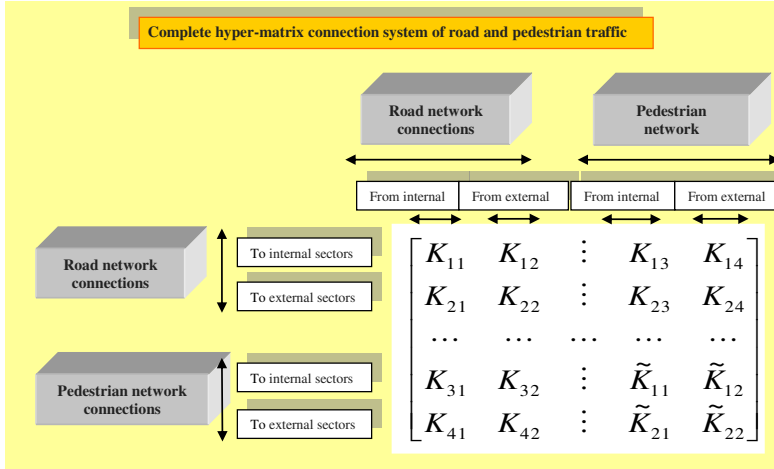

Figure 4

The hyper-matrix describing the entire connection system of road and pedestrian traffic

In the union of the two networks the mathematical modeling applies the following hyper-matrix in the system of differential equations:

$$K = \begin{bmatrix} K_{11} & K_{12} & \vdots & K_{13} & K_{14} \\ K_{21} & K_{22} & \vdots & K_{23} & K_{24} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ K_{31} & K_{32} & \vdots & \tilde{K}_{11} & \tilde{K}_{12} \\ K_{41} & K_{42} & \vdots & \tilde{K}_{21} & \tilde{K}_{22} \end{bmatrix} \qquad (3)$$

During the modeling a closed domain is analyzed (the so called narrowed model is used) and traffic on the edges are considered on the basis of measurement results, so all external-external relations,

e.g. K_{22}, $\tilde{K}_{22}$ are not needed to be analyzed.

The union of these two models means that the two networks establish close dynamic connections with each other and the existing ones are complemented with new internal-internal, external-internal, and internal-external connections.
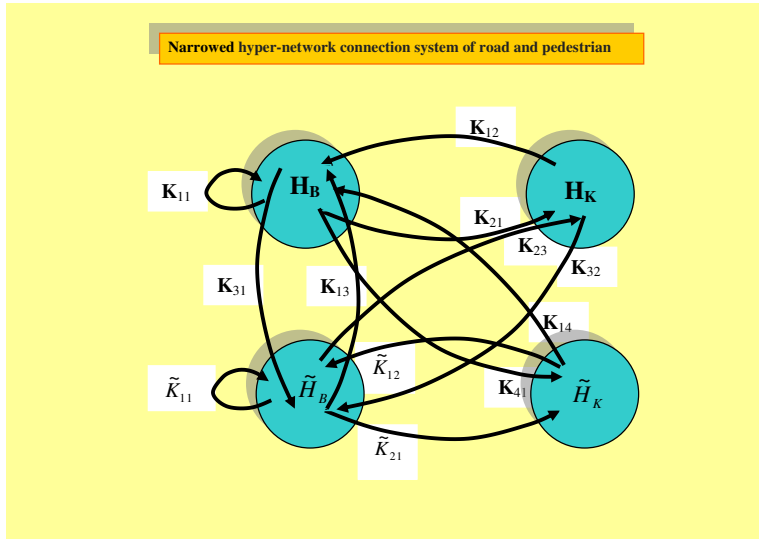

Figure 5
Narrowed connection system of road and pedestrian traffic hyper-network

The hyper-matrix describing the narrowed connection system of the union of vehicle and pedestrian traffic network models is the following:

$$
K = \begin{bmatrix}
K_{11} & K_{12} & \vdots & K_{13} & K_{14} \\
K_{21} & 0 & \vdots & K_{23} & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots \\
K_{31} & K_{32} & \vdots & \tilde{K}_{11} & \tilde{K}_{12} \\
K_{41} & 0 & \vdots & \tilde{K}_{21} & 0
\end{bmatrix}
\tag{4}
$$

3.1 Remarks on the Union of Network Models

1) The union operation is mathematically closed because the same macroscopic approach is applied to pedestrian traffic and road vehicle traffic. The mathematical model of both dynamic networks is a Euler model belonging to the positive nonlinear system class. *Further generalization is the union of several physically different - but according to the above - mathematically identical network models. Since they are disjoint, we can perform the extension of base networks practically in a serial recursive manner, and thus the base network is replaced by the extended one.*

The structure of the model in this case is the following: the top left corner of the main diagonal of the new connection hyper-matrix contains the base network while the sub matrix in bottom right corner of that represents the additional network (both with the entire internal-internal, input and output connections). The transport connections from the new network towards the base network are entered in the top right corner of the antidiagonal of the new connection hyper-matrix, while those from the base network towards the new network are in the bottom left corner of the antidiagonal.

So technically the extension is always the filling of the antidiagonal, which considers the connections between the two networks.

Any changes that can be transferred into these new network parts corresponding to the road model features must be defined at the elements of the connection matrices.

The union related to public transport network in case of buses and trolleys can be handled easier than the one related to pedestrian and cycling network. The macroscopic flow modeling of tram, subway, suburban railway and railway traffic is well manageable with the scheduled stopping of the flow at stops, similarly to the traffic light control method. The validation of the model can be performed with the analysis of speed processes and travel times.

2) An important characteristic of the union is that it is created through a dynamic co-operation of physically separate networks. *It can be concluded that some networks keep their internal autonomy, while at the same time they have a high level of co-operation with each other.*

3) The pedestrian traffic network has two types of edge connection, on the one hand the inputs and outputs originating from the **external pedestrian network** connections and on the other hand those from the **road network** element connections by the union.

4) The input-output connections of the road network are not changed by the union, they invariably receive and transfer vehicle flow. **The time delays in the vehicle traffic network, however, can be greatly affected** by the state parameter vector $\mathbf{x}^\sim$ of the pedestrian density that occurs at pedestrian crossings and in the case of passenger exchange stop times of community vehicles which are taken into account in the model by the $\beta_{ii}(\underline{\mathbf{x}}^\sim)$ pedestrian state-based obstruction factors. In some overload cases the speed reductions due to the change in acceleration capabilities of the vehicles emerge at the speed-density law, taking into account the pedestrian state parameter vector $V(\underline{\mathbf{x}}, \underline{\mathbf{x}}^\sim, \underline{\mathbf{e}})$.

5) A common property of both networks is that state and time-dependent **flow distribution is created** at each junction. This property can be recorded and measured with cameras. **At the macroscopic modeling the personal expectations and preferences of the travelers prevail through the distributions. However, another, better traffic control may change the distributions in the same time of the day.**

6) **The transfer processes between the section elements are disturbed by traffic lights and obstructed or supported by the state-dependent transfer factors, the transfer speeds are affected by the state and environmental factors.**

7) **The "system control" in respect of "traveler influence" can be studied optimizing the passes through the union of two networks (speed/time, cost, environmental load)** [12]**.** The analysis of the passes of pedestrians between selected pairs of points on the network on mixed trajectories is performed in "traveler stream" manner (with mixed-use of pedestrian road sections and community vehicles) it is practical to give strategies for the determination of the selected pairs of points. *The control strategy should be formulated as a model predictive control **MPC**-based one with a model computing time orders of magnitude faster than the real time, of which our mathematical model is capable.*

If the optimal arrival is analyzed on multiple trajectories, task requires the solution of a variational computational problem. Along all trajectories the X length of the road travelled to t time leads to an X(t) route function that gives *T* travel time upon arrival at point "B" and this mapping provides the real function J:

J: $X(t) \rightarrow T$ (5)

The model describing the large-scale transport networks thus can be applied in real time route recommendation taking into account the traffic parameters as well.

4　Example for the Union of Vehicle and Pedestrian Traffic Models

4.1　Description of the Example Model

The solid lines indicate the road network, the dashed lines indicate the pedestrian network. The input and output connections of both are denoted by dual-line arrows.

The tilde-marked section numbering and input-output points relate to pedestrian traffic. The standard section numbering and input-output points relate to road traffic. $\mathbf{T}_1$ and $\mathbf{T}_2$ transits between pedestrian and road traffic.
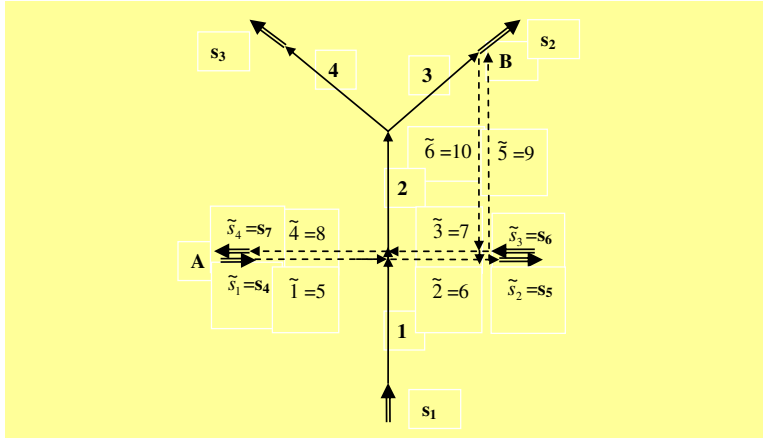

Figure 6

Simple example of road and pedestrian network connections

The relationship between the two networks is characterized by the following:

Road traffic influencing observed:

The road traffic between 1 and 2 is obstructed by the pedestrian traffic of sections $\tilde{2}$ and $\tilde{4}$ passing across the junction through the marked pedestrian crossing, which is considered with the pedestrian density.

The distributions acting at pedestrian traffic:

$\tilde{3}$ to $\tilde{4}$ and $\tilde{3}$ to 2.

$\tilde{1}$ to $\tilde{2}$ and $\tilde{1}$ to 2.

$\tilde{6}$ to $\tilde{3}$ and $\tilde{6}$ to $\tilde{s}_2$.

$\tilde{2}$ to $\tilde{5}$ and $\tilde{2}$ to $\tilde{s}_2$.

$\tilde{s}_3$ to $\tilde{5}$ and $\tilde{s}_3$ to $\tilde{3}$.

The distributions acting at road traffic:

2 ℰ to 3 and 2 ℰ to 4.

3 ℰ to s_2 and 3 ℰ to 6˜.

4.2 The Computer-Algebraic Mathematical Model of the Example Model

The analysis of the union of vehicle and pedestrian traffic networks. Let us consider the previously discussed connection hyper-matrix describing the entire system of connections: Figure 4.

In the example model the following matrices are zero matrices:

- Since there are no external-external connections within the own network $K_{22} = 0$, $K\tilde{}_{22} = 0$.

- Since there are no connections from external pedestrian networks to road networks $K_{14} = 0$, $K_{24} = 0$.

- Since there are no connections from external road networks to pedestrian networks $K_{32} = 0$, $K_{42} = 0$.

- Since there are no connections from internal road network to external pedestrian networks $K_{41} = 0$.

Based on the above, the following sub-matrices of the connection hyper-matrix describe the connection system of the network union:

$$K = \begin{bmatrix} K_{11} & K_{12} & \vdots & K_{13} & 0 \\ K_{21} & 0 & \vdots & K_{23} & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ K_{31} & 0 & \vdots & \tilde{K}_{11} & \tilde{K}_{12} \\ K_{41} & 0 & \vdots & \tilde{K}_{21} & 0 \end{bmatrix} \tag{6}$$

Using the narrowed network model described based on the system of differential equation (2) the construction of submatrices was carried out using a computer-algebraic model. The symbols used in the following are found in section 2.2, at the description of the mathematical model of universal vehicle traffic network.

For the description of the road network model the following matrices were determined:

K_{11}:

$$
\begin{bmatrix}
-\beta_{2,1}\, S(x_2)\, V(x_2,x_1)\, E(x_1) & 0 & 0 & 0 \\
\beta_{2,1}\, S(x_2)\, V(x_2,x_1)\, E(x_1) & -\alpha_{3,2}\, S(x_3)\, V(x_3,x_2)\, E(x_2) - (1-\alpha_{3,2})\, S(x_4)\, V(x_4,x_2)\, E(x_2) & 0 & 0 \\
0 & \alpha_{3,2}\, S(x_3)\, V(x_3,x_2)\, E(x_2) & -\beta_{outp,\,2,3}\, S(s_2)\, V(s_2,x_3)\, E(x_3) & 0 \\
0 & (1-\alpha_{3,2})\, S(x_4)\, V(x_4,x_2)\, E(x_2) & 0 & -S(s_3)\, V(s_3,x_4)\, E(x_4)
\end{bmatrix}
$$

In the example model the road traffic obstruction by the pedestrian traffic occurs at the marked pedestrian crossing at T_1 taking into account the pedestrian density with the function $\beta_{2,1}(\tilde{x}_2, \tilde{x}_4)$.

$\mathbf{K}_{12}$:

$$
\begin{bmatrix}
\beta_{inp,\,1,1}\, S(x_1)\, V(x_1,s_1)\, E(s_1) & 0 & 0 \\
0 & 0 & 0 \\
0 & 0 & 0 \\
0 & 0 & 0
\end{bmatrix}
$$

For the description of the pedestrian network model the following matrices are applied:

$\mathbf{\tilde{K}}_{11}$:

$$
\begin{bmatrix}
-\beta_{2,1}\, S(x_2)\, V(x_2,x_1)\, E(x_1) & 0 & 0 & 0 & 0 & 0 \\
\beta_{2,1}\, S(x_2)\, V(x_2,x_1)\, E(x_1) & -\alpha_{5,2}\, S(x_5)\, V(x_5,x_2)\, E(x_2) - \alpha_{outp,\,2,2}\, S(s_2)\, V(s_2,x_2)\, E(x_2) & 0 & 0 & 0 & \alpha_{2,6}\, S(x_2)\, V(x_2,x_6)\, E(x_6) \\
0 & 0 & -\alpha_{4,3}\, S(x_4)\, V(x_4,x_3)\, E(x_3) & 0 & 0 & (1-\alpha_{2,6})\, S(x_3)\, V(x_3,x_6)\, E(x_6) \\
0 & 0 & \alpha_{4,3}\, S(x_4)\, V(x_4,x_3)\, E(x_3) & -S(x_4)\, V(s_4,x_4)\, E(x_4) & 0 & 0 \\
0 & \alpha_{5,2}\, S(x_5)\, V(x_5,x_2)\, E(x_2) & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & -\alpha_{2,6}\, S(x_2)\, V(x_2,x_6)\, E(x_6) - (1-\alpha_{2,6})\, S(x_3)\, V(x_3,x_6)\, E(x_6)
\end{bmatrix}
$$

$\mathbf{\tilde{K}}_{12}$:

$$
\begin{bmatrix}
S(x_1)\, V(x_1,s_1)\, E(s_1) & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & \alpha_{inp,\,3,3}\, S(x_3)\, V(x_3,s_3)\, E(s_3) & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & (1-\alpha_{inp,\,3,3})\, S(x_5)\, V(x_5,s_3)\, E(s_3) & 0 \\
0 & 0 & 0 & 0
\end{bmatrix}
$$

The amounts to be transferred from the pedestrian traffic to the vehicles, e.g. the one at point $\mathbf{T_1}$ from sections $\tilde{1}$ and $\tilde{3}$ obviously only reduces the density of pedestrian traffic, but does not increase the vehicle density. Similarly, the passengers getting off at point $\mathbf{T_2}$ do not reduce the vehicle density, but increase the pedestrian density on section $\tilde{6}$. In the case of section $\tilde{5}$ one can see that there is a pedestrian flow to an external road network section.

However, the pedestrians on vehicles ("temporary" passengers) are periodically removed from pedestrian traffic, in this way they play the same role as the parking places on the road traffic network. It is special in the way that they are staying on moving "parking places".

The pedestrian traffic getting on at point T_1 and likewise getting on or off at point T_2 can be controlled by the schedule-adjusted $\gamma_i(t)$ intensity functions.

5   Generalization of the Problem

The generalization of the problem in the case of surface traffic networks is the determination of the union of n different types of disjoint traffic networks. This results in a joint system model, which considers co-operations existing in the real world. The general approach can be illustrated well with n=4 different types of disjoint traffic networks.

Let us consider the following network types:

H_K: road network

H_G: pedestrian network

H_P: rail network

H_B: bicycle network

According to the above in addition to the previously considered road vehicle and pedestrian processes rail and bicycle network processes are taken into account as well.

In the creation of the union of various disjoint traffic network systems a more complex case can occur than the model type used in practice (the cases allowing multiple connections), because in this case internal and external networks (which generate the input and output processes) belongs to each network type (the state parameters of which are studied).

The main diagonal (the colored matrices) of the connection hyper-matrix of the combined network (the union) contains the connection hyper-matrices of the different sub-networks.

The mathematical model contains n^2 - n connection hyper-matrices outside the main diagonal, since in principle any internal element of a network can be in connection with internal or external elements of a different network. Similarly, any external element of a network can be in connection with internal or external elements of a different network. These are the connections outside the main diagonal of the hyper-matrix. In the example these are illustrated by the arrows.

		H_K		H_G		H_P		H_B	
		Inner	**Outer**	**Inner**	**Outer**	**Inner**	**Outer**	**Inner**	**Outer**
H_K	**Inner**								
	Outer								
H_G	**Inner**								
	Outer								
H_P	**Inner**								
	Outer								
H_B	**Inner**								
	Outer								

Figure 7

The connection hyper-matrix of the union in the case of four different types of disjoint traffic networks

In reality, the physical properties of each network systems are those which, in principle, permit or prohibit the potential cross-connections. The properties related to the cross-connections are presented below for the case of a simpler structure with four different types of global disjoint traffic networks.

The creation of the union of global disjoint traffic network systems is the simpler case, because each network type has only internal networks (the state parameters of which is examined on the entire surface of the Earth) and the external networks are empty networks.

	H_K	H_G	H_P	H_B
H_K		$K_{K,G}$	O	$K_{K,B}$
H_G	$K_{G,K}$		$K_{G,P}$	$K_{G,B}$
H_P	$K_{P,K}$	$K_{P,G}$		$K_{P,B}$
H_B	O	$K_{B,G}$	O	

Figure 8

The connection hyper-matrix of the union in the case of four different types of global disjoint traffic networks

The properties of the sub-matrices of the connection hyper-matrix are the following:

$K_{K,G}$: transfers passengers from pedestrian traffic to road traffic, but does not increase the vehicle density of road traffic

$K_{G,K}$: transfers passengers from road traffic to pedestrian traffic, but does not decrease the vehicle density of road traffic

$K_{P,G}$: transfers passengers from pedestrian traffic to rail traffic, but does not increase the vehicle density of rail traffic

$K_{G,P}$: transfers passengers from rail traffic to pedestrian traffic, but does not decrease the vehicle density of rail traffic

0: rail traffic does not transfer vehicles to road traffic

$K_{P,K}$: road traffic can transfer vehicles to rail traffic, but it does not increase the vehicle density of rail traffic (e.g. rail transportation of road vehicles)

0: road traffic does not transfer vehicles to bicycle traffic

$K_{K,B}$: bicycle traffic can transfer vehicles to road traffic. There are two cases: 1) the bicycle effectively takes part in the traffic and increases the density, 2) it is placed on a vehicle without increasing the density. The bicycle network density is decreasing in both cases.

$K_{B,G}$: transfers passengers from pedestrian traffic to bicycle traffic. The pedestrian density is decreasing; the bicycle density is increasing from the parking bicycles.

$K_{G,B}$: transfers passengers from bicycle traffic to pedestrian traffic. The density of parking bicycles and the passenger density is increasing as well.

0: rail traffic does not transfer vehicles to bicycle traffic

$K_{P,B}$: transfers passengers from bicycle traffic to rail traffic, but does not increase the vehicle density of rail traffic (e.g. transportation of bicycles). The density of bicycle traffic is decreasing.

Conclusion

We presented a new macroscopic mathematical modeling technique based on the union of the models. The model belongs to very new and modern area in the field of mathematical research, as it is located in the non-linear positive system class. The special hyper-matrix structure defines the cooperation of network elements and the system of differential equations describing the connections. These models provide excellent opportunities for the application of the Lyapunov function method in various areas of the domain-level non-linear control as well. The application possibilities in the domain-level control in big cities are significant. The model has excellent capabilities in the optimization of movements along the trajectories and in the reduction of traffic-oriented environmental pollution [11, 24]. The model is particularly suitable for describing the transfer/vehicle exchange processes that are critical elements of transportation chains.

The exploration and the better understanding of processes occurring in large-scale complex traffic networks led us to a new principled modeling and to the introduction of flexible optimal control methods. The purpose of the control can change flexibly based on the different states of the network in different domains. The network ITS is a variable network that evaluates the optimal satisfaction of direct demands related to traffic during the operation, such as optimal arrival, environmental load, safety and power-saving operation [13, 14, 15]. The projects aiming at creating urban traffic models serve for reducing both congestion and environmental impact [9, 10].

The additional software development related to the research aims to provide a tool for real-time industrial control tasks in case of large-scale intelligent road networks. Notably, in every case the preparation of the traffic system plan of the smart city [4] is the primary compass for the implementation of an ITS network. The synergy and the potential for further exploitation of the results are very positive. New accelerated methods can arise for the analysis of complex movements (transportation chains) along the trajectories, the environmental load and safety as well.

References

[1]   Georgia Aifandopoulou, Iraklis Stamos, Glikeria Myrovali, Maria Morfoulaki (2015) Mapping the Opportunities of Rail Transport; a Data-Driven Methodology for Capturing Rail Competitiveness at European Level, Period. Polytech. Transp. Eng., Vol. 43, No. 1 (2015) pp. 48-54, DOI: 10.3311/PPtr.7739

[2]   Caccetta and Rumchev, (2000) A Survey of Reachability and Controllability for Positive Linear Systems,' Annals of Operations Research, Vol. 98, pp. 101-122, 2000

[3]   Yi Cao, Zhongyi Zuo, Huizhi Xu (2014) Analysis of Traffic Conflict Characteristic at Temporary Reversible Lane, Period. Polytech. Transp. Eng., Vol. 42, No. 1 (2014) pp. 73-76, *DOI:* 10.3311/PPtr.7068

[4]   Bálint Csonka, Csaba Csiszár (2016) Service Quality Analysis and Assessment Method for European Carsharing Systems, Period. Polytech. Transp. Eng., Vol. 44, No. 2 (2016) pp. 80-88, DOI: 10.3311/PPtr.8559

[5]   Thierry Brenac, Christophe Perrin, Bastien Canu, Joël Magnin, Adrien Canu (2015) Influence of Travelling Speed on the Risk of Injury Accident: a Matched Case-Control Study, Period. Polytech. Transp. Eng., Vol. 43, No. 3 (2015) pp. 129-137, DOI: 10.3311/PPtr.7520

[6]   Ákos Dömötörfi, Tamás Péter, Krisztián Szabó (2016) Mathematical Modeling of Automotive Supply Chain Networks, Period. Polytech. Transp. Eng., Vol. 44, No. 3 (2016) pp. 181-186, DOI: 10.3311/PPtr.9544

[7]   Domokos Esztergár-Kiss, Zoltán Koppányi, Tamás Lovas (2016) Mobility Mapping Based on a Survey from the City of Berlin, Period. Polytech. Transp. Eng., Vol. 44, No. 1 (2016) pp. 35-41, DOI: 10.3311/PPtr.7587

[8]   Farina, L. and Rinaldi, S, (2000) *Positive Linear Systems Theory and Applications.* John Wiley &; Sons, Inc.

[9]   István Lakatos; Dániel Pup; Dávid Czeglédi: Determination of Power and Torque Curves of Electric Driven Vehicles based on Diagnostic Methods, in ASME, ASME 2015 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference: Mechatronics for Electrical Vehicular Systems. 31 p., Boston, 2015.08.02-2015.08.05. New York: American Socety of Mechanical Engineers

(ASME), 2015. Paper DETC2015-46724; V009T07A069. 11 p. 9, (ISBN: 978-0-7918-5719-9)

[10]   Adam Titrik; Istvan Lakatos; David Czegledi: Saturation Optimization of Selective Waste Collection Vehicles Based on Real-Time Info-Communication System, In: ASME (szerk.), ASME 2015 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference: Mechatronics for Electrical Vehicular Systems. 31 p., Boston, 2015.08.02-2015.08.05. New York: American Socety of Mechanical Engineers (ASME) 2015, Paper DETC2015-46720; V009T07A068. 7 p., (ISBN:978-0-7918-5719-9)

[11]   István Lakatos; Péter Kőrös; Viktor Nagy: Operation and Applicability Issues of Powertrain Models in Electric Vehicle Development, In: Anon (szerk.), 10th IEEE/ASME International Conference on Mechatronic and Embedded Systems and Applications, Senigallia, 2014.09.10-2014.09.12. (IEEE) New York: IEEE, 2014, Paper CD (ISBN:978-1-4799-2772-2)

[12]   Lakatos István: Diagnostic Measurement for the Effective Performance of Motor Vehicles, Acta Polytechnica Hungarica (ISSN: 1785-8860) 10: (3) pp. 239-249 (2013) Engineering (miscellaneous)

[13]   István Lakatos; Ferenc Szauter; Dávid Czeglédi: Investigating the Appliance of Aluminum as a Winding Material with High Efficiency Electric Motor, 12th IEEE/ASME International Conference on Mechatronic and Embedded Systems and Applications, Auckland, New-Zealand, 2016.08.29-2016.08.31. New York: IEEE, 2016 Paper 43, DOI: 10.1109/MESA.2016.7587119

[14]   Ernő Horváth, Péter Kőrös, István Lakatos, Péter Dely: Development of Individual Information Technology Systems of Experimental Vehicles, in Szakál Anikó (szerk.) 10th Jubilee IEEE International Symposium on Applied Computational Intelligence and Informatics (SACI 2015), Timisoara, Romania, 2015.05.21-2015.05.23. Budapest: Óbudai Egyetem, 2015, pp. 489-493 (ISBN:978-1-4799-9910-1) IEEE Xplore

[15]   István Lakatos, Peter Dely: In-Cylinder Pressure Indication of Internal Combustion Engines for Diagnostic Purposes, In: Anikó Szakál (szerk.) SAMI 2015 • IEEE 13th International Symposium on Applied Machine Intelligence and Informatics, Herlany, Slovakia, 2015.01.22-2015.01.24. Herlany: IEEE Hungary Section, 2015, pp. 37-40 (ISBN:978-1-4799-8220-2; 978-1-4799-8221-9)

[16]   Mohammad Maghrour Zefreh, Ádám Török (2016) Improving Traffic Flow Characteristics by Suppressing Shared Taxis Maneuvers, Period. Polytech. Transp. Eng., Vol. 44, No. 2 (2016) pp. 69-74, *DOI:* 10.3311/PPtr.8226

[17]   Péter, T. (2012) Modeling Nonlinear Road Traffic Networks for Junction Control, International Journal of Applied Mathematics and Computer Science (AMCS), 2012, Vol. 22, No. 3, pp. 723-732, DOI: 10.2478/v1006-012-0054-1

[18]   Peter, T. and Szabo, K. (2012) A New Network Model for the Analysis of
Air Traffic Networks. Periodica Polytechnica- Transportation Engineeering
40/1 (2012) 39-45 DOI: 10.3311/pp.tr.2012-1.07

[19]   Peter, T. Bokor, J. and Strobl, A (2013) Model for the Analysis of Traffic
Networks and Traffic Modelling of Győr, pp 167-172, Doi: 0023, IFAC
Workshop on Advances in Control and Automation Theory for
Transportation Applications (ACATTA 2013) which is to be held in
Istanbul, Turkey, 16-17 September 2013, http://www.acatta13.itu.edu.tr/

[20]   Péter, T. and Fazekas, S. (2014) Determination of Vehicle Density of Inputs
and Outputs and Model Validation for the Analysis of Network Traffic
Processes, Periodica Polytechnica, Transportation Engineering Vol. 42, No
1, 2014, pp. 53-61

[21]   Péter T., Csiszár Cs., Mándoki P.: Különböző felszíni közlekedési
hálózatok forgalmának együttes modellezése és komplex analízise.
Közlekedéstudományi Szemle, LXVII. 1, 2017, pp. 16-34

[22]   Domokos Esztergár-Kiss, Csaba Csiszár (2016) Multicriteria Analysis of
Hungarian Journey Planners, Period. Polytech. Transp. Eng., Vol. 44, No. 2
(2016) pp. 97-104, *DOI:* 10.3311/PPtr.8570

[23]   Renata Stasiak-Betlejewska (2015) Value Engineering Application in the
American Transportation Industry, Period. Polytech. Transp. Eng., Vol. 43,
No. 4 (2015) pp. 206-210, *DOI:* 10.3311/PPtr.8448

[24]   Szauter, F., Péter, T. and Lakatos, I. (2014) Examinations of Complex
Traffic Dynamic Systems and New Analysis, Modeling and Simulation of
Electrical Vehicular Systems, 10th IEEE/ASME International Conference
on Mechatronic and Embedded Systems and Applications. Senigallia, 10-
12 Sept. 2014, INSPEC Accession Number: 14701078, pp. 1-5, DOI:
10.1109/MESA.2014.6935613, Print ISBN: 978-1-4799-2772-2

[25]   T. Vadvári and P. Várlaki (2015) Identification of Supply Chains Based on
Input-Output Data Period. Polytech. Transp. Eng., Vol. 43, No. 3 (2015)
pp. 162-167, *DOI: 10.3311/PPtr.7931* Online published: 28-04-2015

[26]   T. Vadvári, P. Várlaki, Queuing Models and Subspace Identification in
Logistics, *Acta Technica Jaurinensis,* Vol. 8, No. 1, pp. 63-76, 2015, *DOI:
10.14513/*actatechjaur. v8.nl. 352

[27]   P. Várlaki, T. Vadvári, A. Rövid, Identification of Input-Output Models For
Supply Chains - Analytical and Intelligent Approaches, in *Global Conference on
Operations and Supply Chain Management,* Bandung, Indonesia, pp. 88-98,
March 12-14, 2012, ISBN:978-967-5705-06-9

[28]   Máté Zöldy, Ádám Török (2015) Road Transport Liquid Fuel Today and
Tomorrow: Literature Overview, Period. Polytech. Transp. Eng., Vol. 43,
No. 4 (2015) pp. 172-176, *DOI:* 10.3311/PPtr.8095

Evolving Fuzzy Models for Automated Translation

Alina Ţenescu[1], Radu-Emil Precup[2], Nicuşor Minculete[3]

[1]Faculty of Letters, University of Craiova, Str. A. I. Cuza 13, 200585 Craiova, Romania; E-mail: tenescu.alina@ucv.ro

[3]Department of Automation and Applied Informatics, Politehnica University of Timisoara, Bd. V. Parvan 2, 300223 Timisoara, Romania
E-mail: radu.precup@upt.ro

[3]Faculty of Mathematics and Informatics, Transilvania University of Brasov, Str. Iuliu Maniu 50, 500091 Brasov, Romania; E-mail: minculete.nicusor@unitbv.ro

Abstract: This paper targets two goals. First, it analyzes the most common errors in automated translation from French to English and from English to French performed by a statistical and a hybrid translation engine with the help of the evaluation metric SAE J2450. The test of concordance is applied in order to study the agreement between the original text and its retroversion within the same translation memory software. Seven categories of primary errors are considered, which cover the fields of terminology, semantics, structure, orthography, punctuation and completeness. Second, evolving fuzzy models are developed, which give the overall paragraph score using the seven categories of primary errors as inputs. The fuzzy models permit the users to establish the accuracy of translation and to grade the quality of translations resulted from the reintroduction of the result of translation in the same software application. They also allow the comparison of two popular translation memory programs, namely Google Translate (GT) and Systran, in the framework of the issues of concordance in translation and artificial learning.

Keywords: accuracy; automated translation; concordance; errors; evolving fuzzy systems

1 Introduction

Nowadays, we are witness to an unquestionable reality: the extent and dimensions of informative and formative space are increasingly influenced by the Internet, by computerization and automation which pervade the users' cognitive sphere, directing the contents and orienting the global organization of their ideas. The Internet and computer systems are becoming determining factors of the new dimensions of educational area – revealing the intentions, the content, the means

and the instruments of education – and provide access to a wide range of contents, information and current trainings, which prove themselves necessary all the more so as we all participate to the multidimensional process of globalization of education. In order to implement a new teaching program (for example, in foreign languages or translation) which makes use of traditional methods, but also of the original means made available by new technologies, the teaching staff (teaching to students in translation and interpreting) need training and specialization in three domains: specialization in the domain of pedagogy and education, specialization in the scientific domain linked to the informational content of teaching and specialization in the domains of informatics and programming. We also have to take into account the fact that the new generation of students, "the digital natives", answers more easily and effectively to a novel way of learning, combining the classic classroom sessions with independent learning and with computer-based training (CBT) and web-based training (WBT). The education program based on blended learning helps students in foreign languages and translation to develop digital skills and increases the level of knowledge and handling of online tools of translation engines, in parallel with the improvement of their textual competence and the enhancement of information and technology literacy.

The first goal of this paper is to analyze the most frequent errors in automatic translation. These errors can be identified by students in career training (translation and interpreting) and by professional translators as they carry out translation of a text from French to English ("theme") and from English to French ("version"), following the automatic processing of a paragraph of academic text by two translation memory software applications, namely Google Translate (GT) and Systran. We will then apply the concordance test ("goodness of fit" test, the "adjustment" test) on texts which are translated from French to English and from English to French by each translation software mentioned above, in order to establish the degree of concordance between the original text T1, which is translated, from English to French ("version") and the text T3, which represents the result obtained after the translation into French of T1 is reintroduced within the same translation memory software and retranslated into English.

In the context of the first goal, the text T1 in English will be translated into French with the help of an online translation memory software application (for example, GT or Systran). We obtain the text T2 in French. In a second phase, we introduce the text T2 in the same application and we translate it into English. We note the result with T3 and we compare T1 to T3, establishing the degree of concordance between them with the aid of a statistical test. We will use a statistical test to study the agreement or the concordance of the translation to the original text and to compare T1 with T3.

The analysis of the most frequent errors in automatic translation leads to seven categories of primary errors, i.e. those which cover the fields of terminology, semantics, structure, orthography, punctuation and completeness. The second goal of this paper is to develop fuzzy models, which give the Overall Paragraph Score

(OPS) using these seven categories of primary errors as inputs. The evolving fuzzy models are built upon the results obtained by the Process Control group of the Politehnica University of Timisoara, Romania, and reported in [1-6]. Four evolving fuzzy models are developed for the translation from English to French of the data obtained using GT and Systran in terms of employing either the Recursive Least Squares (RLS) algorithm or the weighted Recursive Least Squares (wRLS) algorithm to update the rule consequents of the fuzzy models. The basic version of the incremental online identification algorithm is implemented according to [6] using the software support specific to eFS Lab [7, 8]. Evolving fuzzy models for the translation from French to English are not developed as the translation is accurate enough to produce low errors that create numerical problems in the identification algorithm.

The results of the new ideas expressed as the two goals are important as our final goal is to propose a course or a path making use of an original method and of a pedagogical software tool that makes possible for students in translation and professional translators to extend and detail research on the existing differences between the degrees of accuracy and quality of translations furnished by different language translation software applications, as well as on the recurrent errors with each translation software. Since translated texts or paragraphs are far from being correct or perfect, the process of automated translation as imperfect model obliges students not only to discriminate the wrong elements found in translation, but also to critically analyze and improve their translator skills by constantly having to refine, revise and proofread an instant automated translation. The use of nonlinear models as fuzzy ones can be beneficial with this regard.

This paper is organized as follows: the analysis of translation results with focus on GT and Systran is carried out in the next section. The metamorphosis of a text through translation (quality and evaluation) is first treated, and a discussion on the choice of automatic translation engines and evaluation metric for translation quality is included. Section 3 is dedicated to the development of the new evolving fuzzy models. The results are validated on real data. The conclusions are highlighted in Section 4.

2   Analysis of Translation Results

Nowadays, a translator is not only he who deciphers the message issued by a transmitter, but also he who recodes it and renders it comprehensible to the receiver. This is though not possible without touching and altering a little bit the content of the message, sometimes at the expense of its quality, value and concordance by comparison to the source. As shown in [9], Kumarajiva asserted that translation resembles, in fact, "to already masticated food that will be given to him who cannot chew it himself". Automated translation would then be

tantamount to victuals chewed by the mouth of an artificial organism whose aim would be to guide the human operator in his work as "validator", as verifier and checker of linguistic solutions generated by the implementation of mathematical algorithms, or, according to Toudic et al. [10], as "post-editor" or as proof-reader who edits, revises, corrects and removes errors from translations made by an automated engine, an engine with translation memory, so as to achieve a certain degree of acceptability and quality. Nevertheless, we can easily become aware of the fact that such food does not have the original taste and aroma anymore.

Bowker and Ehgoetz [11] insist on the fact that automated or machine translation (MT) entails the shift from the concept of "absolute quality" to that of threshold of acceptability of translation by the final beneficiary. It is obvious that this questions the notion of relativity of quality, defined by relationship to the level of acceptability of the product for this final recipient.

Today, the emergence of new media and the development of automated translation and of translation memory tools have changed the perspective on the notion of quality in translation. Green et al. argue that this is not anymore focused on how faithful or unfaithful the translation is by relationship to the source-text [12]. On the contrary, it is re-centered to take into account the final usage or the particular use of the translated text.

While, as stated in [10], the tools of computer-assisted translation and translation memories diminish the final responsibility of the translator as far as the quality of the resulting translation is regarded, the quality of the translated text is established by the quality of translation machines, of pre-translated segments, of terminological databases, as well as by the quality of the revision work lead by tools of evaluation of translation quality and by the human post-editor, contends Koehn [13]. Bar-Hillel draws attention on this partnership between the translation automaton and the post-editor, partnership whose major problem is, according to him, "the region of optimality in the continuum of possible divisions of labor" [14].

Though the evaluation of automated translation systems and the evaluation of translation quality prove to be objects of infinite debates and difficult issues to solve, our objective is to evaluate the quality and the concordance (with the source) of the translation produced by systems using fully automatic translation, which does not involve any human participation or revision. Aside from these automated systems, there are numerous tools and instruments supporting translation, such as electronic dictionaries, whose evaluation is achieved, according to Ryan [15], with the participation of individuals, and which relies upon the evaluation measures of the human-machine interface and on the techniques of evaluation of usability, as argue Hartley and Popescu-Belis [16].

This section is focused on the manner to assess, measure and compare the quality of translations provided by two automatic translation systems (GT and Systran), translations (from English to French and from French to English) deemed

completed and used as such by different users, or by other machine translation systems or software applications. We will first identify the most recurrent errors in the translation from English to French and from French to English of a paragraph of academic text processed by two professional machine translation systems (GT and Systran), using the translation quality metric SAE J2450. In a second stage, we will compare, via the test of concordance ("goodness of fit" test), the quality of translations resulted from the reintroduction of the result of translation in the same software application as well as their concordance with the source-text.

Online translators are based on several approaches of automated translation. These approaches are grouped in four main categories: 1) rule-based machine translation (RBMT) which operates in a more complex way than word-to-word translation and applies linguistic rules in three stages (analysis, transfer, generation), allowing different words to be placed in different places, according to the context, as demonstrate Nirenburg [17] and Costa-Jussa et al. [18]; 2) statistical machine translation (SMT) which is a complex form of word translation, word-based or, more recently, phrase-based, using statistical weights in order to choose the most likely translation of a given word. SMT uses probability and adopts a learning algorithm of the translation produced by humans, show Brown et al. [19]; 3) example and analogy-based machine translation which makes use of bilingual corpora with parallel texts as its main knowledge source and is thus based on previously seen examples in these parallel corpora, according to Nagao [20] and Turcato et al. [21]; 4) hybrid machine translation which blends fundamental elements of the rule-based machine translation system with those of the statistical-based system and whose result depends on the quality of the alignment of candidate translations to the source-sentence and will bring the literal sense over into the statistical result, show Boretz [22] and Hunsicker et al. [23], thus taking advantage of the fluidity of statistical machine translation and of the accuracy of rule-based machine translation, asserts Drexler [24]. Hybrid translation is performed in two stages: first, the linguistic analysis of the text in the source language, contends Rubino [25]; secondly, we pass to the target-language, through non-linguistic approaches, by the act of translating sub-sentential segments, while lexical selection in the target-language is achieved, in the second phase, through a language model.

As each approach of automated translation has a different principle and mode algorithm, the results of translation will be different. Given the opposed character of approaches to automated translation, we have decided to undertake a comparative study of translations provided through the automated translation by a statistical machine translation software (GT) and by a hybrid automated translation software (Systran). We have chosen the statistical automated translator GT since it is one of the most popular online translation software applications, and all the more so as it continuously increases the list of language pairs for which it establishes parameters of translation memory, as well as for its easy access, usability and applicability. The second automated translator, Systran, is built on a

hybrid translation software application, one of the most well-known using hybrid machine translation. The reason behind its choice lies in the old kinship with GT, namely GT initially used a rule-based machine translation of Systran, since Systran was one of the first companies to develop and use RBMT systems and is, nowadays, one of the most well-known to use hybrid MT.

Green et al. [26] and Madsen [27] have argued that even though with automatic translation it is impossible to achieve precise translation and though automated translators commit errors, their relevance and utility in our lives are undeniable. Therefore, we should focus on how these automated translators face issues raised by different language pairs, language grammars and certain grammatical structures, on how they handle and tackle differences in linguistic typologies, translation of idioms and colloquialisms, and on how they draw out responses to language change across time and to linguistic, cultural and social barriers and uncertainties.

If one of the objectives of this section is to comprehend the level of performance of the two approaches (SMT and HMT) in automated translation of academic texts, we also want to reveal the weaknesses and the strengths that can affect the performance of the two online translation software applications. A second objective would be to study and analyze the typical errors in the translation of academic texts from English to French and from French to English, using SAE J2450 translation quality standard, in order to discover which of the two approaches is more accurate and precise. Using the categories established by the SAE J2450 evaluation metric for quality of translation, we manually evaluated the results of translations provided by the two software applications and we calculated the score, expressed as OPS, indicating the performance of HMT and SMT. The score represents a specific case of translation.

If we compare the results of translations from French to English and from English to French, as well as the results of translations resulted from the reintroduction of the result of translation in the same software application, we highlight the advantages and disadvantages of the two approaches, taking into account indicators such as the accuracy of translation at several levels, the performance, the flexibility and the adaptation to changes. We also show the relationship between the conclusions of the manual evaluation of results of automated translation with each of the two software applications and the technical procedures of each of the two approaches.

Although, according to Papineni et al [28], the evaluation of automated translation is sometimes considered a sort of "black art" or occult science, we have nevertheless chosen the SAE J2450 translation quality metric because it has the advantage of being easy to use and because it can be applied irrespective of the source-language or target-language. Another advantage is represented by its capacity to try and solve the issue of ambiguities by way of two meta-rules: 1) "When an error is ambiguous, always choose the earliest primary category",

2) "When in doubt, always choose 'serious' over 'minor'" (SAE J2450, [29]), choice which would ensure, according to Secară [30] a greater coherence and "consistency of classification of errors across evaluators".

The evaluation metric comprises four sub-divisions: 1) seven categories of primary errors which cover the fields of terminology, semantics, structure, orthography, punctuation, completeness; 2) two secondary sub-categories: serious error, minor error; 3) two meta-rules which help the evaluators in their decision, in the case of ambiguity; 4) numeric weights for each primary category and sub-category. For the first sub-division, there are recognized seven categories of errors taken into account in the evaluation of translation quality. The overall aim is to generalize, which renders a scheme of general categories of errors with SAE J2450. The errors are first weighed and afterwards classified as 'serious' or 'minor' (SAE, 2001) [29]. The weights of the seven categories of primary errors categorized as serious errors/minor errors are [29]: Wrong Term (WT): 5/2, Syntactic Error (SE): 4/2, OMission (OM): 4/2, Word Structure and Agreement error (SA): 4/2, MisSPelling (SP): 3/1, Punctuation Error (PE): 2/1, and Miscellaneous Error (ME): 3/1.

In this section, in order to obtain the final note for the quality of translation in the target-language, we calculated, in the same manner as the one established by SAE standard, the sum of numerical values of the totality of committed errors and then we divided it by the total number of words in the source-text. Hartley and Popescu-Belis [16] assert that we have thus the "occasion to calibrate in order to favor" either the correct grammatical form or the terminological correctness.

The analysis allowed us to identify the typical errors in the translation from English to French and from French to English performed by the two translation memory programs GT and Systran. We tested not only the performance of each software application, but also the precision of translations provided by these automated translation applications. We chose as corpus of study two paragraphs of recent academic texts:

T1 [31]: "Information literacy has been a tremendous "win" for academic libraries. But it risks becoming, looking back, also a symbol of a great loss. If we do not refocus our efforts on the educational, cultural, and technological shifts in which "information literacy" per se becomes a somewhat arbitrary label for the very stuff of learning and information discovery in today's academic (and larger) world, we will have won the battle but lost the campaign".

T4 [32]: "Une caractéristique de ces différentes approches de l'utilisabilité consiste à étendre le champ d'application de l'ergonomie aux produits et aux technologies interactives de la vie courante. Non seulement parce que ceci constitue un nouveau champ de débouchés, mais surtout parce qu'avec le renouvellement rapide des produits, il existe une relation étroite entre les manières de les consommer et les manières de les produire. À l'expression "comprendre le travail pour le transformer", on pourrait rajouter "comprendre les modes de

relations aux produits pour transformer le travail" et "comprendre les relations aux produits pour agir sur l'expérience".

As far as the paragraphs are concerned, the one in English is an excerpt from an article on information literacy [31], while the one in French is an excerpt from a study on the acceptability of new technologies [32]. We focused on the comparison between the source-paragraph and the translated result. In the choice of paragraphs selected from academic texts, we used as criterion of selection their provenance from a credible source and their relatively recent publication (during the last six years). Some results of the translation are given as follows:

T2 (translated result using GT): maîtrise de l'information a été un "gagnant" énorme pour les bibliothèques universitaires. Mais il risque de devenir, en regardant en arrière, aussi un symbole d'une grande perte. Si nous ne recentrer nos efforts sur les changements éducatifs, culturels et technologiques dans lesquels "information literacy" en soi devient une étiquette quelque peu arbitraire pour la substance même de l'apprentissage et de découverte de l'information dans le monde académique (et plus) d'aujourd'hui, nous aurons gagné la bataille, mais a perdu la campagne.

T3 (reintroduction of the translated result in the same software application, GT): information literacy has been a "win" huge for academic libraries. But it may become, looking back, as a symbol of a great loss. If we focus our efforts on educational change, in which cultural and technological "information literacy" itself becomes a somewhat arbitrary label for the substance of learning and discovery of information in the academic world (and more) today, we have won the battle but lost the campaign.

T2' (translated result using Systran): L'instruction de l'information a été une "victoire" énorme pour les bibliothèques scolaires. Mais elle risque devenir, regardant en arrière, aussi un symbole d'une grande perte. Si nous ne refocalisons pas nos efforts sur les décalages éducatifs, culturels, et technologiques dans lesquels la "instruction de l'information" devient en soi un label quelque peu arbitraire pour la substance même de l'étude et de la découverte de l'information en monde scolaire (et plus grand) d'aujourd'hui, nous aurons gagné la bataille mais aurons perdu la campagne.

T3' (reintroduction of the translated result in the same software application, Systran): The instruction of information was an enourmous "victory" for the school licraries. But it risks to become, looking at behind, also a symbol of a great loss. If we do not refocalisons our efforts on the shifts educational, cultural, and technological in which "instruction of information" becomes in oneself a somewhat arbitrary label for the substance even of the study and of the discovery of information in world school (and larger) of today, we will have gained the battle but will have lost the countryside.

T5 (translated result using GT): A characteristic of these different approaches to usability is to extend the scope of ergonomics products and interactive technologies in everyday life. Not only because this is a new field of opportunities, but mainly because with the rapid turnover of products, there is a close relationship between ways of consuming and ways of producing them. In the term "understand the work to transform the" one might add "understand the products to relationship patterns to transform the work" and understand the relationships to products to act on experience.

T6 (reintroduction of the translated result in the same software application, GT): Une caractéristique de ces différentes approches de la facilité d'utilisation est d'étendre le champ d'application des produits d'ergonomie et technologies interactives dans la vie quotidienne. Non seulement parce que cela est un nouveau champ de possibilités, mais surtout parce que le renouvellement rapide des produits, il existe une relation étroite entre les modes de consommation et les moyens de les produire. Dans le terme "comprendre le travail de transformer" le on pourrait ajouter "comprendre les produits à des modèles relationnels pour transformer le travail" et comprendre les relations aux produits d'agir sur l'expérience.

T5' (translated result using Systran): A characteristic of these various approaches of the utilisability consists in extending the field of application of ergonomics to the products and interactive technologies of the everyday life. Not only because this constitutes a new field of outlets, but especially because with the fast renewal of the products, there exists a close relationship between the manners of consuming them and manners of producing them. With the expression "to include work to transform it", one could add "to understand the modes of relations to the products to transform work" and to understand the relations with the products to act on the experiment.

T6'(reintroduction of the translated result in the same software application, Systran): Une caractéristique de ces diverses approches de l'utilisability consiste en prolongeant le champ de l'application de l'ergonomie aux produits et aux technologies interactives de la vie quotidienne. Non seulement parce que ceci constitue un nouveau champ des débouchés, mais particulièrement parce qu'avec le renouvellement rapide des produits, là existe une relation étroite entre les façons de les consommer et les façons de les produire. Avec l'expression "pour inclure le travail pour le transformer", on a pu s'ajouter "pour comprendre les modes des relations aux produits pour transformer le travail" et pour comprendre les relations avec les produits pour agir sur l'expérience.

For the evaluation of quality of translation provided by the software applications we identified and classified the errors according to SAE J2450 translation quality standard and QA model and we associated to each found error the corresponding weight according to this metric. In the process of observation, classification and comparative analysis of the quality of translations performed by the two online translation memory software applications, we employed the current version of GT

and Systran, that of June 2016. After having gathered data, we analyzed them using the method for error calculation established by SAE J2450 QA model. The total score of each translation counts the errors of each category found in the translation of each paragraph by the two software applications, while the overall paragraph score is calculated by dividing the sum of numerical weights of the totality of errors committed in the paragraph by the total number of words in the paragraph excerpt from the source-text. We then compared the results obtained following the translations performed by the two online memory translators. We also applied the test of concordance so as to analyze the degree of accuracy of translation as the result of initial translation (from T1 in English to T2 in French/from T4 in French to T5 in English) was reintroduced in the same software application (we retranslated T2 in the same language as that of the source-text T1) and we compared T1 with T3 and T4 with T6, using the same evaluation metric for translation quality.

Figure 1 shows us that when the source-text is in English, a higher degree of accuracy in translation is achieved by the translation memory program GT. This is also true for the test of concordance, when in the second phase, we introduce the text T2 in the same software application (GT) and we retranslate it into English. Comparing T1 with T3 and T1 with T3', we notice a higher degree of concordance between T1 and T3 (GT) than between T1 and T3' (Systran). OPS will be used as the output of the fuzzy models in the next section.
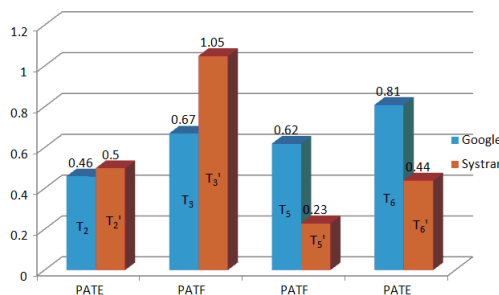
Figure 1
Comparison of Overall Paragraph Scores (OPS), where: PATE – the paragraph of academic text in English and PATF – the paragraph of academic text in French

When the source-language is French, we obtain opposite results: a higher degree of accuracy is achieved by the translation memory software application Systran. The same result was reached when we applied the concordance test to compare T4 with T6 and T6'.

If we make a comparison of scores associated to wrong terms, we draw the following conclusions: when the source-paragraph is in English, we encounter many more errors committed by Systran than by GT. When the source-paragraph is in French, there are more errors committed by GT than by Systran (as shown in Figure 2). WT is the first input to the fuzzy models developed in the next section.
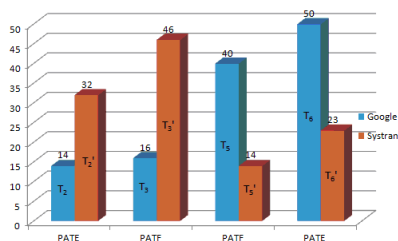
Figure 2
Comparison of Wrong Term (WT) scores

When we reintroduce the result of the paragraph translated from English to French in the same software application so as to compare the new result with the source-text (T1 in English), we remark fewer wrong terms with GT than with Systran. When we reintroduce the result of the paragraph translated from French to English in the same software application so as to compare the new result with the source-text (T4 in French), we note less wrong terms with Systran than with GT.

As regards the comparison related to the number of syntactic errors, when the source-paragraph is in English, Systran commits fewer errors than GT (according to Figure 3). When we reintroduce the paragraph T2 in the same MT application (GT) and T2' in the same software application (Systran) and we retranslate them into English, we obtain a higher accuracy with GT than with Systran. SE is the second input to the fuzzy models developed in the next section.
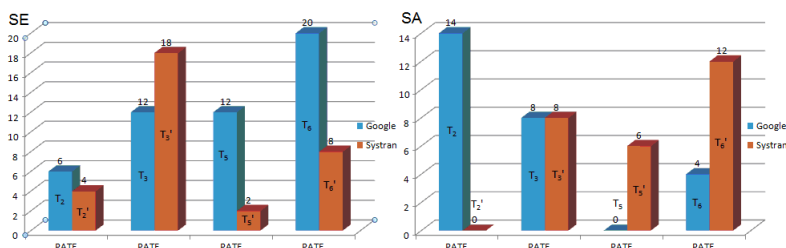
Figure 3
Comparison of Syntactic Error (SE) and Word Structure and Agreement error (SA) scores

When the source-paragraph is in English, the scores measuring the weight of SA errors are higher with GT than with Systran. For the test of concordance, when we reintroduce the result of the paragraph translated from English to French in the same software application in order to compare the new result with the source-text (T1 in English), we get similar scores registering the weight of morphological errors for GT and for Systran.

When the source-paragraph is in French, the scores measuring the weight of SA errors are higher with Systran than with GT. For the concordance test, when we reintroduce the result of the paragraph translated from French to English in the same automated translation application in order to compare the new result with the

source-text (T4 in French), we note higher scores for SA errors with Systran than with GT. The obvious result is that while we translate from English to French, we achieve a higher precision with GT and that, when we translate from French to English, a higher degree of accuracy is reached by Systran (as illustrated in Figure 3). SA is the third input to the fuzzy models developed in the next section.

The weakness of automated translation provided by Systran consists in the fact that it records more wrong terms than GT, inclusively at the test of concordance. By comparison with Systran, GT counts higher scores while it weighs the burden of syntactic errors.

The category of WT is the most important and in the translation performed by Systran, the existence of a higher score recorded for WT is sometimes due to the lack of entry in the dictionary of the automated translator (this is the case of the translation of the English term "utilisability" into French: the automated translator copies the English term as it is while it provides the translation into French). Syntactic errors represent a second major category of errors found with the two software applications whose presence might be justified by the gap between two languages of different roots: Latin root and Germanic root. As regards SA errors, with the two automated translators there is a weakness in the recognition of (noun) determiners and a lack of correspondence and agreement between tenses and verbal moods, a lack of sequence of tenses, as well as erroneous translations of moods and tenses (for example, in the translation T2: "nous ne recentrer nos efforts " and "nous aurons gagné la bataille, mais a perdu la campagne").

As far as the category ME is concerned, a mistake of this kind is characterized as such when it does not befit the other six types of error acknowledged by the metric and the mistake which is most often included in the seventh category is either an awkward choice of preposition or a preposition that misses in the target language. After the comparative study of the behavior of automated translators, we encounter miscellaneous mistakes in translations performed by both GT and Systran, while we can also reveal the fact that GT counts a higher omission error score, as in the translation from English to French the online translation memory application omits terms it does not know. On the other hand, the hybrid translation engine of Systran records a higher error score whilst performing the translation of abbreviations (for example, abbreviations of EU units or institutions). Punctuation errors are common with both automated engines, yet GT repeats the common mistake of not adapting the English inverted commas quotation marks to "guillemets" in the translation from English to French; it is also highly possible that with GT a complex sentence begins with a lower case letter instead of an upper case letter. While the statistical engine favors the choice of synthetic or Saxon Genitive, the hybrid machine translator prefers to choose the analytical or periphrastic Genitive in the translation of texts from French to English. If both engines record mistakes in the translation of idioms, the hybrid application achieves a better performance in the translation of idiomatic expressions from French to English than the statistical translation engine, as it also registers a better

choice of tenses and moods after certain conjunctions (of subordination, such as "bien que" which compulsorily requires a Subjunctive and not an Indicative). Last but not least, we should emphasize that GT is designed to provide a higher translation speed in both translations (from English to French and from French to English) as compared to the hybrid engine Systran which is not only a bit slower, but it also sometimes gets blocked.

3 Evolving Fuzzy Systems Modeling

The flowchart of the online identification algorithm that produces evolving fuzzy systems models is presented in Figure 4, where k is the data sample index, $\mathbf{p}_k$ is the current data point:

$$\mathbf{p}_k = [p_k^1 \quad p_k^2 \ldots p_k^{n+1}]^T, \quad \mathbf{p} = [\mathbf{z}^T \quad y]^T = [z_1 \quad z_2 \ldots z_n \quad y]^T$$
$$= [p^1 \quad p^2 \ldots p^n \quad p^{n+1}]^T \in \Re^{n+1}, \tag{1}$$
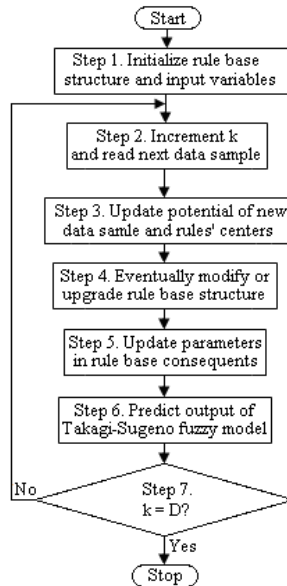
Figure 4
Flowchart of online identification algorithm [6]

T stands for matrix transposition, and the input-output data set is:

$$\{\mathbf{p}_k \mid k = 1 \ldots D\} \subset \Re^{n+1}, \tag{2}$$

where D is the number of input-output data points, also called data points or (data) samples.

The rule base of fuzzy models of Takagi-Sugeno type is:

$$\text{Rule } i : \text{IF } z_1 \text{ IS } LT_{i1} \text{ AND} ... \text{AND } z_n \text{ IS } LT_{in}$$
$$\text{THEN } y_i = a_{i0} + a_{i1}z_1 + ... + a_{in}z_n, i = 1...n_R, \tag{3}$$

where z_j, $j = 1...n$, are the input variables, n is the number of input variables, $LT_{ij}, i = 1...n_R, j = 1...n$, are the input linguistic terms, y_i is the output of the local model in the rule consequent of the rule index $i, i = 1...n_R$, n_R is the number of rules, and $a_{il}, i = 1...n_R, l = 0...n$, are the parameters in the rule consequents. The fuzzy model structure includes the algebraic product t-norm as an AND operator and the weighted average defuzzification method.

As specified in the previous sections, the fuzzy models developed in this paper are characterized by a number of $n = 7$ input variables, which are $z_1 = \text{WT}$, $z_2 = \text{SE}$, $z_3 = \text{OM}$, $z_4 = \text{SA}$, $z_5 = \text{SP}$, $z_6 = \text{PE}$ and $z_7 = \text{ME}$, and the output variable is $y = \text{OPS}$. The system inputs are presented in Figure 5, which illustrates the input data for both training and validation (testing).
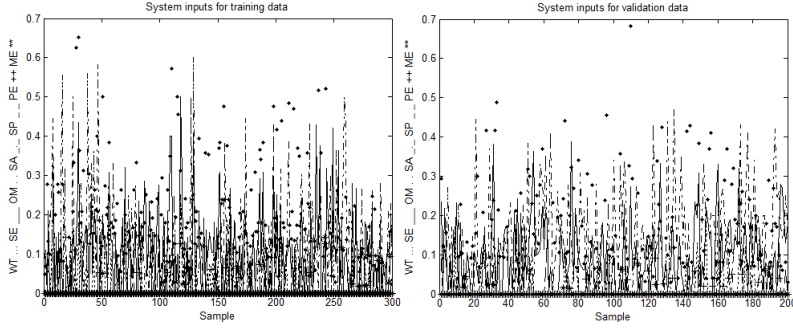
Figure 5

System inputs versus data sample for training data and validation (testing) data

The root mean square error (RMSE) is used as a global performance index in order to compare the four fuzzy models. Its definition is:

$$\text{RMSE} = \sqrt{(1/D)\sum_{k=1}^{D}(y_k - y_{d,k})^2}, \tag{4}$$

where y_k is the model output and $y_{d,k}$ is the real system output, i.e., the actual value of OPS obtained by the application of GT or Systran in the translation from English to French.

The fuzzy model 1, which corresponds to GT using RLS, has evolved to $n_R = 6$ rules, it consists of 132 parameters and its performance on the validation data is $\text{RMSE} = 0.25924$. The fuzzy model 2, which corresponds to GT using wRLS, has

evolved to $n_R = 6$ rules, it consists of 132 parameters and its performance on the validation data is $\mathrm{RMSE} = 0.25923$. The fuzzy model 3, which corresponds to Systran using RLS, has evolved to $n_R = 7$ rules, it consists of 154 parameters and its performance on the validation data is $\mathrm{RMSE} = 0.29652$. The fuzzy model 4, which corresponds to Systran using wRLS, has evolved to $n_R = 7$ rules, it consists of 154 parameters and its performance on the validation data is $\mathrm{RMSE} = 0.29653$.

These results show that for the data set considered in this paper the best performance is obtained by the fuzzy models that model the GT-based translation, namely the fuzzy models 1 and 2. The use of either the RLS algorithm or the wRLS the algorithm to update the rule consequents of the fuzzy models in the online identification algorithm does not affect the results.

The outputs of the fuzzy model 2 and of the real system output are illustrated in Figure 6 for both training data and validation (testing) data.
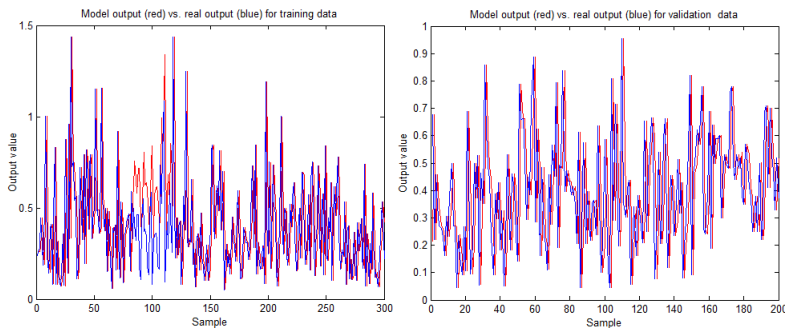
Figure 6
System output (OPS) versus data sample of evolving fuzzy model and real system for training data and validation (testing) data

The results presented in Figure 6 are encouraging. However, different conclusions are expected to be obtained for other applications [33-40].

Conclusions

This paper has proposed evolving fuzzy models, which give the Overall Paragraph Score in translation using these seven categories of primary errors as inputs. Four fuzzy models have been developed using an incremental online identification algorithm applied to the translation from English to French. The data sets have been obtained by processing the results obtained by the application of GT and Systran.

By using the excerpts from academic texts and two source-languages (English and French), it is not easy to discern that on the occasion of evaluation of translation provided by the hybrid translation software application, carried out with the help of SAE J2450 standard, we record a higher score than that of the evaluation of

translation made by the statistical software application for all seven categories of errors. It is true that statistical software performs better than the hybrid software in the translation of terms, but it is also obvious that in the translation from French to English and on applying the test of concordance between the original paragraph in French and the translation resulted from the reintroduction of the result translated from French to English in the same software application (Systran), we obtain a higher degree of accuracy and precision for the hybrid software than for the statistical software.

All these operations of translation and the passages from English to French and from French to English in translation performed by automated engines, as well as the process of evaluation of translation quality with the help of SAE J2450 metric, of the test of concordance and of the assessment made with the statistics software Excel, provide a reference guide for automated translation analysis which could prove its utility and efficiency for students in translation and for professional translators, since it gives them the chance to enhance their analytic thinking skills and to expand their researches on the existing differences between degrees of accuracy of translation engines chosen in their work, on the weak and strong points of each of these automated translators, on the recurrent errors of each software application, all the more so as the error is always found in the center of their perception of translation learning.

The future research will be focused on the development of neural network models and other appropriate applications with various optimization approaches [41-50] included in the identification algorithm. The development of fuzzy models for the translation from French to English will require modifications in the structure of the identification algorithms.

Acknowledgement

References

[1]   R.-E. Precup, H.-I. Filip, M.-B. Radac, C. Pozna, C.-A. Dragos, S. Preitl: Experimental Results of Evolving Takagi-Sugeno Fuzzy Models for a Nonlinear Benchmark, Proceedings of 2012 IEEE 3rd International Conference on Cognitive Infocommunications, Kosice, Slovakia, 2012, pp. 567-572

[2]   R.-E. Precup, H.-I. Filip, M.-B. Radac, E. M. Petriu, S. Preitl, C.-A. Dragos: Online Identification of Evolving Takagi-Sugeno-Kang Fuzzy Models for Crane Systems, Applied Soft Computing, Vol. 24, 2014, pp. 1155-1163

[3]   R.-E. Precup, E.-I. Voisan, E. M. Petriu, M.-B. Radac, L.-O. Fedorovici: Implementation of Evolving Fuzzy models of a Nonlinear Process, Proceedings of 12th International Conference on Informatics in Control, Automation and Robotics, Colmar, France, 2015, Vol. 1, pp. 5-14

[4]   R.-E. Precup, E.-I. Voisan, E. M. Petriu, M.-B. Radac, L.-O. Fedorovici: Gravitational Search Algorithm-Based Evolving Fuzzy Models of a Nonlinear Process, Informatics in Control, Automation and Robotics, J. Filipe, K. Madani, O. Gusikhin, and J. Sasiadek, Eds., Lecture Notes in Electrical Engineering, Springer International Publishing, Cham, Vol. 383, 2016, pp. 51-62

[5]   R.-E. Precup, T.-A. Teban, T. E. Alves de Oliveira, E. M. Petriu: Evolving Fuzzy Models for Myoelectric-Based Control of a Prosthetic Hand, Proceedings of 2016 IEEE International Conference on Fuzzy Systems, Vancouver, Canada, 2016, pp. 72-77

[6]   R.-E. Precup, M.-B. Radac, E. M. Petriu, R.-C. Roman, T.-A. Teban, A.-I. Szedlak-Stinean: Evolving Fuzzy Models for the Position Control of Twin Rotor Aerodynamic Systems, Proceedings of 2016 IEEE 14th International Conference on Industrial Informatics, Poitiers, France, 2016, pp. 237-242

[7]   J. V. Ramos, A. Dourado: On Line Interpretability by Rule Base Simplification and Reduction, Proceedings of European Symposium on Intelligent Technologies, Hybrid Systems and Their Implementation on Smart Adaptive Systems EUNITE 2004, Aachen, Germany, 2004, pp. 1-6

[8]   L. Aires, J. Araújo, A. Dourado: Industrial Monitoring by Evolving Fuzzy Systems, Proceedings of Joint 2009 IFSA World Congress and 2009 EUSFLAT Conference, Lisbon, Portugal, 2009, pp. 1358-1363

[9]   A. Bantaş, E. Croitoru: Didactica Traducerii, Editura Teora, Bucharest, 1999 (in Romanian)

[10]   D. Toudic, K. Hernandez Morin, F. Moreau, F. Barbin, G. Phuey: Du Contexte Didactique aux Pratiques Professionnelles: Proposition d'une Grille Multicritères pour l'Évaluation de la Qualité en Traduction Spécialisée, ILCEA Online, No. 19, 2014, pp. 1-19

[11]   L. Bowker, M. Ehgoetz: Exploring User Acceptance of Machine Translation Output: A Recipient Evaluation, Across Boundaries: International Perspectives on Translation, D. Kenny, K. Ryou, Eds., Cambridge Scholars Publishing, Newcastle upon Tyne, 2007, pp. 209-224

[12]   S. Green, J. Heer, C. D. Manning: The Efficacy of Human Post-Editing for Language Translation, Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Paris, France, 2013, pp. 439-448

[13]   P. Koehn: 20 Years of Statistical Machine Translation, University of Edinburgh Press, Edinburgh, 2009

[14]   Y. Bar-Hillel: The Present Status of Automatic Translation of Languages, Advances in Computers, No. 1, 1960, pp. 91-163

[15]   R. Ryan: Traduction Technique: les Langues Contrôlées au Service de l'Ergonomie Documentaire, ILCEA Online, No. 14, 2011, pp. 1-18

[16]   A. Hartley, A. Popescu-Belis: Évaluation des Systèmes de Traduction automatique, in Évaluation des Systèmes de Traitement de l'Information, S. Chaudiron, Ed., Éditions Hermès, 2004, pp. 311-335

[17]   S. Nirenburg: Knowledge-Based Machine Translation: Parameter Estimation, Machine Translation, No. 4, 1989, pp. 5-24

[18]   M. Costa-Jussà, M. Farrús, J. Mariñom, J. Fonollosa: Study and Comparison of Rule-Based and Statistical Catalan-Spanish Machine Translation Systems, Computing and Informatics, Vol. 31, 2012, pp. 245-270

[19]   P. Brown, S. Della Pietra, V. Della Pietra, R. Mercer: The Mathematics of Statistical Machine Translation: Parameter Estimation, Computational Linguistics, Vol. 19, No. 2, 1991, pp. 263-311

[20]   M. Nagao: A Framework of a Mechanical Translation between Japanese and English by Analogy Principle, Artificial and Human Intelligence, R. Banerji, Ed., Elsevier Science Publishers, 1984, pp. 173-180

[21]   D. Turcato, P. McFetridge, F. Popowich, J. Toole: A Unified Example-Based and Lexicalist Approach to Machine Translation, Proceedings of 8th International Conference on Theoretical and Methodological Issues in Machine Translation, Chester, UK, 1999, pp. 33-43

[22]   A. Boretz: AppTech Launches Hybrid Machine Translation Software, SpeechTechMag.com published on the 2nd of March 2009, http://www.speechtechmag.com/Articles/News/Speech-Technology-News-Features/AppTek-Launches-Hybrid-Machine-Translation-Software-52871.aspx

[23]   S. Hunsicker, C. Yu, C. Federmann: Machine Learning for Hybrid Machine Translation, Proceedings of 7th Workshop on Statistical Machine Translation, Association for Computational Linguistics, Montreal, Canada, 2012, pp. 312-316

[24]   V. Drexler: "I Don't Translate, I Create!" - An On-line Survey on Uniformity versus Creativity in Professional Translations, Anchor Academic Publishing, Hamburg, 2016

[25]   R. Rubino: Traduction Automatique Statistique et Adaptation à un Domaine Spécialisé, PhD Thesis, University of Avignon and the Vaucluse, Montreal, Canada, 2011

[26] S. Green, S. Wang, J. Chuang, J. Heer, C. Manning: Human Effort and Machine Learnability in Computer-Aided Translation, EMNLP, 2014, pp. 1225-1236

[27] M. W. Madsen: The Limits of Machine Translation, MSc Thesis, University of Copenhagen, Copenhagen, Denmark, 2009

[28] K. Papineni, S. Roukos, D. Ward: Corpus-based Comprehensive and Diagnostic MT Evaluation: Initial Arabic, Chinese, French and Spanish Results, Proceedings of HLT, Second International Conference on Human Language Technology Research, M. Markus, Ed., Morgan Kaufmann, 2002, pp. 132-137

[29] SAE: Surface Vehicle Recommended Practice, http://www.apex-translations.com/documents/sae_j2450.pdf, 2001

[30] A. Secară: Translation Evaluation – A State of the Art Survey, Proceedings of the EColorMe/Mellange Workshop, Leeds, UK, 2005, pp. 39-44

[31] S. M. Cowan: Information Literacy: The Battle We Won That We Lost?, Portal: Libraries and the academy, Vol. 14, No. 1, 2012, pp. 23-32

[32] J. Barcenilla, J.-M.-C. Bastien: L'Acceptabilité des Nouvelles Technologies: Qquelles Relations avec l'Ergonomie, l'Utilisabilité et l'Expérience Utilisateur ?, Le Travail Humain, Vol. 72, No. 4, 2009, pp. 311-331

[33] R.-E. Precup, S. Preitl, M. Balas, V. Balas: Fuzzy Controllers for Tire Slip Control in Anti-lock Braking Systems, Proceedings of IEEE International Conference on Fuzzy Systems, Budapest, Hungary, 2004, Vol. 3, pp. 1317-1322

[34] R.-E. Precup, S. Preitl: PI-Fuzzy Controllers for Integral Plants to Ensure Robust Stability, Information Sciences, Vol. No. 20, 2007, pp. 4410-4429

[35] R.-E. Precup, M.-B. Radac, M. L. Tomescu, E. M. Petriu, S. Preitl: Stable and Convergent Iterative Feedback Tuning of Fuzzy Controllers for Discrete-Time SISO Systems, Expert Systems with Applications, Vol. 40, No. 1, 2013, pp. 188-199

[36] A. L. Kazakov, A. A. Lempert: On mathematical Models for Optimization Problem of Logistics Infrastructure, International Journal of Artificial Intelligence, Vol. 13, No. 1, 2015, pp. 200-210

[37] Á. Takács, D. Á. Nagy, I. J. Rudas, T. Haidegger: Origins of Surgical Robotics: From Space to the Operating Room, Acta Polytechnica Hungarica, Vol. 13, No. 1, 2016, pp. 13-30

[38] I. J. Rudas, J. Gáti, A. Szakál, K. Némethy: From the Smart Hands to Tele-Operations, Acta Polytechnica Hungarica, Vol. 13, No. 1, 2016, pp. 43-60

[39]   B. Takács, R. Dóczi, B. Sütő, J. Kalló, T. A. Várkonyi, T. Haidegger, M. Kozlovszky: Extending AUV Response Robot Capabilities to Solve Standardized Test Methods, Acta Polytechnica Hungarica, Vol. 13, No. 1, 2016, pp. 157-170

[40]   S. Anbarasi, S. Muralidharan: Enhancing the Transient Performances and Stability of AVR System with BFOA Tuned PID Controller, Control Engineering and Applied Informatics, Vol. 18, No. 1, 2016, pp. 20-29

[41]   I. Škrjanc, S. Blažič, O. E. Agamennoni: Identification of Dynamical Systems with a Robust Interval Fuzzy Model, Automatica, Vol. 41, No. 2, 2005, pp. 327-332

[42]   F. G. Filip: Decision Support and Control for Large-scale Complex Systems, Annual Reviews in Control, Vol. 32, No. 1, 2008, pp. 61-70

[43]   A. Gajate, R. E. Haber, P. I. Vega, J. R. Alique: A Transductive Neuro-fuzzy Controller: Application to a Drilling Process, IEEE Transactions on Neural Networks, Vol. 21, No. 7, 2010, pp. 1158-1167

[44]   J. Vaščák, K. Hirota: Integrated Decision-Making System for Robot Soccer, Journal of Advanced Computational Intelligence and Intelligent Informatics, Vol. 15, No. 2, 2011, pp. 156-163

[45]   Zs. Cs. Johanyák: Fuzzy Modeling of Thermoplastic Composites' Melt Volume Rate, Computing and Informatics, Vol. 32, No. 4, 2013, pp. 845-857

[46]   P. Baranyi, Y. Yam, P. Varlaki: TP Model Transformation in Polytopic Model-Based Control, Taylor & Francis, Boca Raton, FL, 2013

[47]   O. Castillo, E. Lizárraga, J. Soria, P. Melin, F. Valdez: New Approach Using Ant Colony Optimization with Ant Set Partition for Fuzzy Control Design Applied to the Ball and Beam System, Information Sciences, Vol. 294, 2015, pp. 203-215

[48]   O. Arsene, I. Dumitrache, I. Mihu: Expert System for Medicine Diagnosis Using Software Agents, Expert Systems with Applications, Vol. 42, No. 4, 2015, pp. 1825-1834

[49]   A. Khmelev, Y. Kochetov: A Hybrid Local Search for the Split Delivery Vehicle Routing Problem, International Journal of Artificial Intelligence, Vol. 13, no. 1, 2015, pp. 147-164

[50]   L. Zhu, C.-X. Fan, Z.-G. Wen, H.-R. Wu: Coverage Optimization Strategy for WSN based on Energy-aware, International Journal of Computers, Communication and Control, Vol. 11, No. 6, 2016, pp. 877-888

Numerical Determination of Orthotropic Material Properties of Textile Composite Layers and their Validation by Measurement

Gergely Bojtár[1], Béla M. Csizmadia[2], János Égert[1]

[1] Széchenyi István University, Department of Applied Mechanics, Egyetem tér 1, H-9026 Győr, Hungary, bojtar@sze.hu, egert@sze.hu

[2] Szent István University, Institute of Mechanics and Machinery, Páter K. u. 1, H-2103 Gödöllő, Hungary, csizmadia.bela@gek.szie.hu

Abstract: The knowledge of orthotropic material properties of composite layers is very important for mechanical design of machine structures from composite material. This information cannot be found in related professional literature therefore it is necessary to determine by measurement, by basic experiments of mechanics of materials. The paper presents a new 3D finite element layer model-cell based on the meso-structure of a textile composite layer, with which one shall be able to determine the material constants of an orthotropic layer. The applicability of the numerical layer model-cell and the accuracy of the numerical results are confirmed by experimental results. The numerically determined material properties of the layers are used at mechanical modeling and computation of complex, layered composite structures.

Keywords: orthotropic material properties; textile composite; composite layer; finite element method, layer model-cell; validation with experiments

1 Introduction

Textile composites are plastics reinforced by glass, carbon, aramid, etc. fabrics. The laminated textile composite plates consist of one 2D fabric in each layer. The reinforcing fibers are arranged into flat roving, that run parallel to each other in the textile. The thickness of the textile and a composite layer can be measured in tenth of mm-s. The mechanical analysis of a textile composite structure is a complicated problem because every single layer and the roving in it have anisotropic feature generally. With the majority of engineering problems the behavior of the material can be described by homogeneous macroscopic modeling with satisfying accuracy from engineering point of view. There is a verified computation method for determination of the material properties of a multilayered complete plate from

given layer material properties. The aim of this research is to provide a new, improved numerical method, a layer model-cell for computation of material properties of a single layer.

Different kinds of fabrics are used in textile composites and the majority of them have two main roving directions, perpendicular to each other. These main or principal fiber directions make it possible that the behavior of the majority of textile composite materials can be handled as an orthotropic material from macroscopic point of view with sufficient accuracy.

In Figure 1 it is seen the $x_1 x_2$ material coordinate system of the textile composite layer. x_1 is the chain direction (first principal material direction) of the textile whereas x_2 is the weft direction (second principal material direction).
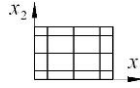
Figure 1
Material coordinate system of a textile composite layer

From macroscopic point of view a textile composite layer can be considered as an orthotropic material in the plane stress state with good approximation. The linear elastic, orthotropic material law in case of the plane stress state, in the coordinate system of principal material directions of the layer, has the following form [4], [9]:

$$\underline{\overline{\sigma}} = \underline{\underline{C}}\,\underline{\overline{\varepsilon}}\;,$$

$$\begin{bmatrix} \overline{\sigma}_1 \\ \overline{\sigma}_2 \\ \overline{\tau}_{12} \end{bmatrix} = \frac{1}{1-\nu_{12}\nu_{21}} \begin{bmatrix} E_1 & \nu_{21}E_1 & 0 \\ \nu_{12}E_2 & E_2 & 0 \\ 0 & 0 & (1-\nu_{12}\nu_{21})G_{12} \end{bmatrix} \begin{bmatrix} \overline{\varepsilon}_1 \\ \overline{\varepsilon}_2 \\ \overline{\gamma}_{12} \end{bmatrix}. \tag{1}$$

Material law (1) of the textile composite layer contains five macroscopic material constants: the E_1, E_2 moduli of elasticity in principal material directions, the ν_{12}, ν_{21} Poisson's ratios and the G_{12} shear modulus of elasticity. Four of them are independent because the matrix $\underline{\underline{C}}$ of material constants is symmetric, therefore the relationship between the Young's moduli and Poisson's ratios is [4]:

$$\frac{\nu_{12}}{E_1} = \frac{\nu_{21}}{E_2}\;. \tag{2}$$

The average strain and stress of a textile composite layer with Volume *V* can be defined as follows [11], [12], [14]:

$$\overline{\varepsilon}_{ij} = \frac{1}{V} \int\limits_{(V)} \varepsilon_{ij}\, dV\;, \quad \overline{\sigma}_{ij} = \frac{1}{V} \int\limits_{(V)} \sigma_{ij}\, dV\;. \tag{3}$$

There are no standards for building of the structure of a textile composite, for the fiber/matrix proportion and coupling and for the weaving method therefore there are a number of ways for the composite material variations and engineering applications. Therefore, the material properties of a composite structure usually cannot be found in the literature, in manuals or in standards. For determination of the material constants of a textile composite there are usually two procedures:

- In the first one, the given material is produced and the material properties are determined by measurement on test specimens cut out from the material.

- In the other one, the material constants of a single composite layer are determined by a so-called mixture rule knowing the material constants and volume fraction of the fiber and the matrix in the composite layer.

The aim of our research is to provide a numerical method for determination of the material constants in the material law (1). For this purpose, we developed a finite element 3D layer model-cell. This model-cell is suitable for the determination of the macroscopic material constants of a textile composite layer of known geometry and material parts. We have proved by measurements on eight-layer composite test specimens that the macroscopic material constants computed by the layer model-cell can reach the necessary accuracy for the modeling of the structure. The paper describes the necessary steps for building of the layer model-cell and the prescription of the loading and the periodic boundary conditions for the given loading cases in general. It shows how to calculate the material constants from the numerical results.

For determination of material properties of the layer there are several numerical methods in the relevant literature which use the material properties of the textile and matrix of the layer [10], [15], [16], [17]. Compared to the models that can be found in the literature we specified the kinematic boundary conditions in a different way in this paper. In addition, we defined the average stress of the finite element model-cell also in a different way. From these modifications, we hope the improvement and growth of the accuracy of the layer model-cell.

2   The Roving Model-Cell

A roving in the textile includes several thousands of μm diameter reinforcing fibers which are impregnated by the matrix material. There is also matrix material between the single fibers. We do not model these single fibers separately in the layer model-cell of the textile composite. We consider the roving as one "fiber" if we model the layer reinforced by textile woven from homogeneous roving's.

The cross section and the $x_{r1}x_{r2}x_{r3}$ roving material coordinate system of the investigated roving, which is impregnated by matrix, can be seen in Figure 2. In

figures the length dimensions are given in mm. The direction parallel to the rein-forcing fibers is x_{r1}. We determined the orthotropic, macroscopic material proper-ties of the roving by the roving model-cell described in the paper [3]. For the structure of the roving model cell we assume that the reinforcing fibers are posi-tioned in a regular hexagonal shape everywhere in the complete roving. In a hex-agonal layout (Figure 3) one fiber is surrounded by six others in a way that those are in equal distance from each other. The cross section of the roving model-cell is marked by a thick line in Figure 3.
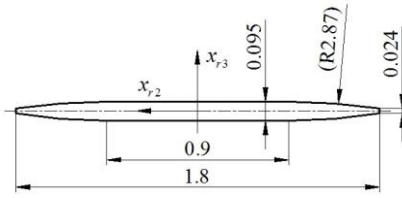
Figure 2

The cross section and

material coordinate system of a flat roving
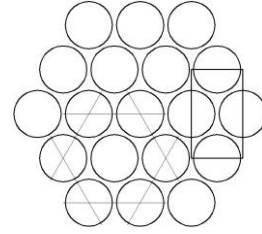
Figure 3

Hexagonal layout of the

fibers in the roving

In the studied case the matrix is always isotropic, but the carbon fibers are trans-versely isotropic and are in a hexagonal layout in the roving model-cell, therefore, also the roving-model cell has transversely isotropic behavior. In the roving the plane of isotropy is the $x_{r2}x_{r3}$ plane. For the investigated case the macroscopic material properties determined by the roving model-cell are [3]:

$E_{r1} = 177\,236$ MPa, $\qquad \nu_{r12} = \nu_{r13} = 0.202$, $\qquad G_{r12} = G_{r13} = 4\,115$ MPa,

$E_{r2} = E_{r3} = 10\,352$ MPa, $\qquad \nu_{r23} = 0.430$, $\qquad G_{r23} = 3\,620$ MPa.

We validate the results of the roving model-cell also here because these material properties are the input data for the textile composite layer model-cell. The experiment validates both the results of the roving model-cell and the textile composite layer model-cell together.

3 The Layer Model-Cell

3.1 The Geometry and Finite Element Mesh of Layer Model-Cell

A textile composite layer can also be considered as periodic parts, volume elements, cells. The layer can be built up from these periodic domains, from so called model-cells.

We show the building up of the layer model-cell, and validate the numerical results by experiment. We produced the eight layer, 2 mm thick composite specimen reinforced with plain weave carbon textile used for the experiments by manual lamination (Figure 4). The matrix material was polyester resin. In order to prevent that the production process would influence the material properties of the test specimens we hardened each sheet with identical technological parameters in the autoclave: $T = 65 - 68\,°\text{C}$, $p = 6\,\text{bar}$, $t = 2\,\text{h}$.

Figure 4

Eight-layer textile composite plate

The thickness of the textile and the waviness of the roving we determined after lamination by measurement. The cross section of the textile can be seen in Fig. 5 after lamination.
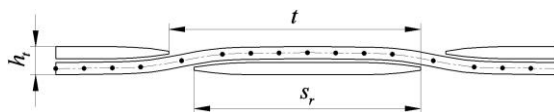
Figure 5

Cross section of the investigated textile after lamination

During the building up of the layer model with thickness h_t we assumed that the textile with h_t thickness is located in the middle of the composite layer (Fig. 6):

$$v = \frac{h - h_t}{2}.$$ (4)
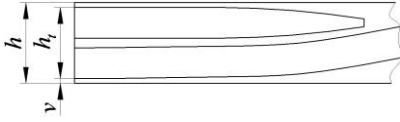
Figure 6

Textile in the composite layer
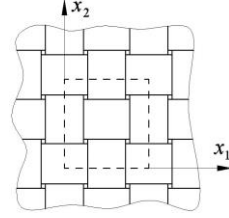
Figure 7

One unit in the textile

In order to be able to provide more simple boundary conditions we apply the side surfaces of the model cell parallel to the planes of the material principal directions [8]. The dimensions in x_1 and x_2 direction of the layer model-cell we determined in a way that the cell should contain one unit both in chain- and weft directions. In the case of a plain weave textile this means 2-2 complete ($1 + 2 \times 0.5 = 2$) roving (Figure 7, dashed line square). E.g. at a 2/2 twill weave textile one unit consists of 4-4 roving.

The x_1, x_2, x_3 material principle directions of the model-cell and the x_{r1}, x_{r2}, x_{r3} principal direction of the roving can be seen in Figure 8. In the case of the roving $x_{r1} x_{r3}$ is the symmetry plane. We applied the model-cell's side surfaces perpendicular to x_1 and x_2 axes in a way that those should coincide with the $x_{r1} x_{r3}$ and $x_{r2} x_{r3}$ material principal direction planes of the roving. In this case the four side surfaces at tension remain in plane and they displace parallel with the $x_1 x_3$ and $x_2 x_3$ planes of the model-cell. At shear in $x_1 x_2$ plane the points of the side surfaces will displace in the same way in direction of the shear. At shear, there is no deformation on the side surfaces perpendicular to x_1 axis in x_2 direction and on the surfaces perpendicular to x_2 axis in x_1 direction.
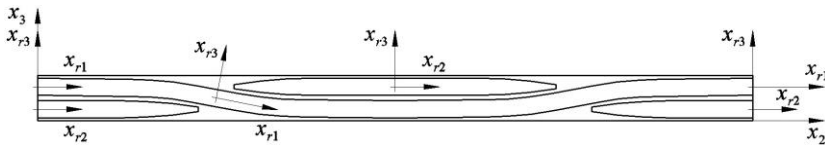
Figure 8

x_1, x_2, x_3 material principal direction of the layer model-cell and

x_{r1}, x_{r2}, x_{r3} material principal direction of the roving

The dimensions of the layer model-cell and the identity signs of the six side surfaces are seen in Figure 9. We marked the side lengths with *a* and *b* because we provided a general formulation for the composite layer finite element modeling. This formulation is also valid for such a weave where $a \neq b$, e.g. the chain- and

the weft direction roving densities are not identical. The dimensions of the investigated composite layer model-cell in Figure 9 are:

$$h = 0.25 \text{ mm}, \quad h_t = 0.22 \text{ mm}, \quad v = 0.015 \text{ mm},$$

$$a = b = 4 \text{ mm}, \quad t_1 = t_2 = 2 \text{ mm}, \quad s_{r1} = s_{r2} = 1.8 \text{ mm},$$

$$A_{A+} = A_{A-} = 1 \text{ mm}^2, \quad A_{B+} = A_{B-} = 1 \text{ mm}^2, \quad A_{C+} = A_{C-} = 16 \text{ mm}^2.$$

We did the finite element computation with the NX I-deas 6.1 program code. We applied a mesh of parabolic tetrahedron elements for the layer model-cell. Due to the periodicity, we need to apply the so called periodical boundary conditions on the layer model-cell. Therefore, we need to generate the finite element mesh in a way that, on the opposite side surfaces there should be nodes opposite of each other and by this way node pairs will be created on two side surfaces of the model-cell. On the side surfaces perpendicular to x_1 and x_2 axes it was necessary to specify surface connections for the two-two surface pairs belonging to the matrix and located opposite to each other, and for the surface pair perpendicular to the x_3 axis.
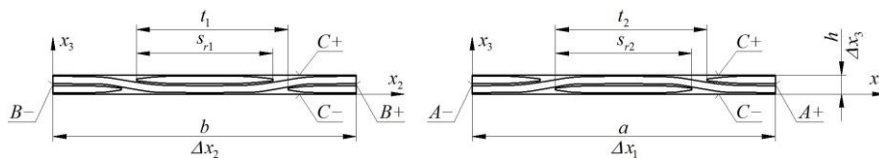
Figure 9
Dimensions of the layer model cell
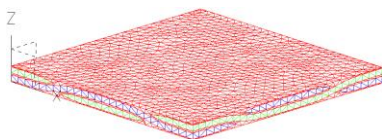
Figure 10
Finite element mesh of the layer model-cell
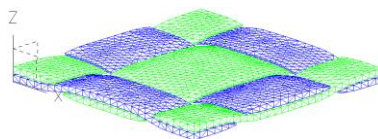
Figure 11
Finite element mesh of the textile layer

We generated the finite element mesh so that we generated second order triangle shaped shell elements for the surfaces first then by using those we generated second order tetrahedron shaped (3D) elements inside the layer model-cell. In Figure 10 we can see the finite element mesh of the layer model cell, which includes 38 974 elements and 56 587 nodes. Figure 11 shows the finite element mesh of the textile. *XYZ* coordinate system shown in Figures 10 and 11 corresponds to the $x_1 x_2 x_3$ coordinate system of Figures 8 and 9.

3.2 Loading of the Layer Model-Cell

We simulated the basic experiments of mechanics of materials with the layer model-cell in a way that we specified a kinematic loading and node displacement field for the side surfaces of the model-cell. For the nodes at the corners of the model-cell and at the center points of the side surfaces we referred with node numbers in Figure 12.
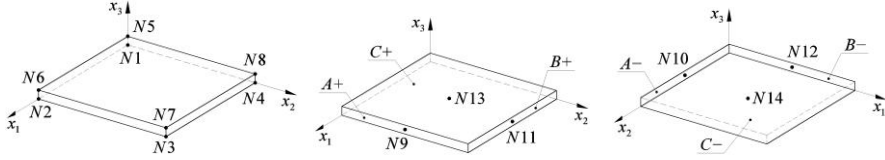

Figure 12

Marking the nodes at the eight corners of the layer model cell and at the center points of the side
surfaces

A node displacement vector is in the model-cell in the $x_1 x_2 x_3$ coordinate system:

$$\vec{u} = u\vec{e}_1 + v\vec{e}_2 + w\vec{e}_3.$$ (5)

For the layer model-cell (Figure 9) the following general periodic boundary conditions can be applied [11]:

$$u_i^{j+}\left(X_1, X_2, X_3\right) - u_i^{j-}\left(X_1, X_2, X_3\right) = \Delta x_j \bar{\varepsilon}_{ij} = c_i^j.$$ (6)

u_i^{j+} and u_i^{j-} are the displacement in X_i direction on the surface pair perpendicular to the X_j axis in the $X_1 X_2 X_3$ global coordinate system. The $j+$ index means the positive X_j axis and $j-$ marks the negative X_j direction in the coordinate system of the model-cell (Figure 9). Δx_j is the side length of the model-cell. c_i^j ($i = j = 1, 2, 3$) is the change of distance of the side surfaces (tension or compression) in X_j directions of the model-cell, whereas $c_i^j = c_j^i$ ($i \neq j = 1, 2, 3$) is the displacement in the plane of the side surface resulting from the shear of the side surfaces. The (6) boundary condition ensures the periodicity and the continuity of the displacement field in the composite material layer. The (6) equation provides, for the appropriate points, the difference of the displacements on both opposite surfaces. u_i^{j+} and u_i^{j-} displacements are functions of X_1, X_2, X_3 coordinates. Therefore, these surfaces remain not necessarily in plane during the deformation.

With the layer model-cell we simulated the basic experiments (tension-pressure and shear) of mechanics of materials and we determined the orthotropic material properties from the model-cell computed stress and strain state. We realised the simulation of these basic experiments with separate, kinematic loading cases.

During providing the kinematic loading, on $A+$, $A-$, $B+$, $B-$ side surfaces we gave displacements not for the individual nodes but the complete side surfaces in a way that we prescribed the displacement only for the middle node and we connected all nodes on the surface in the given direction. This prescription ensures that the side surface can displace in the prescribed direction as a rigid body. *From the loading, we got reaction forces. These reaction forces were reduced to the center node of the surface. Since, the side surfaces of the layer model-cell coincide with the planes of the material principal directions of the roving (Figure 8), the following average stress appears on the side surfaces of the model-cell:*

$$\bar{\sigma}_{ij} = \frac{1}{A_j} \int_{(A_j)} \sigma_{ij} \, dA = \frac{F_i}{A_j} \, . \tag{7}$$

F_i is a reaction force in x_i direction on the surface perpendicular to x_j axis and, A_j is the area of the side surface.

3.3   Tension Test Simulation with the Layer Model-Cell

For the computation of the E_1 modulus of elasticity and ν_{12} Poisson's ratio of the textile composite layer we need to simulate an x_1 directional tension for E_2 and for ν_{21} we need to simulate an x_2 directional tension with the layer model-cell. At tension the strains are non-zero ($\bar{\varepsilon}_1$, $\bar{\varepsilon}_2 \neq 0$) in the strain tensor in the (1) constitutive equation. The (6) general equation for tension is the following:

$$c_i^j = \Delta x_j \bar{\varepsilon}_j \, , \qquad \qquad \bar{\varepsilon}_j = \frac{c_i^j}{\Delta x_j} \, , \qquad \qquad \left(i = j = 1, \, 2\right). \tag{8}$$

At tension tests the side surfaces of the model-cell perpendicular to x_1 and x_2 axes remain in plane and displace parallel with the planes of principal directions. We ensured the parallelism requirement in the case of the tension loading by connecting all the nodes on the side surface in the direction perpendicular to the surface. These prescriptions are summarized by the relationships (9) – (10). According to equation (9) we connected all nodes on $A+$ surface in x_1 direction. The independent node is $N2$, meaning that the displacement of all nodes on $A+$ surface are identical with displacement of the $N2$ node in x_1 direction.

$$u_{A-} = u\left(0; x_2; x_3\right) = u_{N1} \, , \qquad \qquad u_{A+} = u\left(a; x_2; x_3\right) = u_{N2} \, , \tag{9}$$

$$v_{B-} = v\left(x_1; 0; x_3\right) = v_{N5} \, , \qquad \qquad v_{B+} = v\left(x_1; b; x_3\right) = v_{N8} \, . \tag{10}$$

In relationships (9) – (10):

$$0 \leq x_1 \leq a \, , \qquad 0 \leq x_2 \leq b \, , \qquad 0 \leq x_3 \leq h \, . \tag{11}$$

The *A– and B–* side surfaces were clamped only in the central node in the direction perpendicular to the surface at the x_1 and x_2 directional tensions:

$$u_{A-} = u_{N10} = 0, \qquad v_{B-} = v_{N12} = 0. \tag{12}$$

As a result of the tension the waviness (Figure 8) in the roving decreases in direction of the tension and increases perpendicularly to the tension. Therefore, *C+* and *C–* side surfaces do not remain in plane. In order to restrain the rigid body displacement and rotation of the *C–* side surface we have to prescribe zero displacement for the *N1* node in x_3 direction:

$$w_{N1} = 0. \tag{13}$$

The kinematic loading belonging to the x_1 and x_2 directional tension test of the layer model-cell were determined as follows. At the x_i directional tension we prescribed the $\bar{\varepsilon}_i$ specific strain for the model-cell and computed the x_i directional displacement of the side surface perpendicular to the x_i direction from the (8) equation. In this case, we did not prescribe the cross contraction.

In the case of the x_1 directional tension $\bar{\varepsilon}_1$ is given, consequently:

$$u_{A+} = u_{N9} = a\,\bar{\varepsilon}_1. \tag{14}$$

In the case of the x_2 directional tension $\bar{\varepsilon}_2$ is prescribed, out of which the displacement of the *B+* side surface is:

$$v_{B+} = v_{N11} = b\,\bar{\varepsilon}_2. \tag{15}$$

In the case of x_1 and x_2 directional tensions we need to specify periodic boundary condition in addition so that the opposite nodes should displace on the same way on the side surface. For fulfilment of this requirement the node pairs must be connected on the opposite $A-/A+$ and $B-/B+$ side surfaces, except the edges of the model-cell:

$$A-/A+: \quad v\big(0;x_2;x_3\big) = v\big(a;x_2;x_3\big), \qquad w\big(0;x_2;x_3\big) = w\big(a;x_2;x_3\big), \tag{16}$$

$$B-/B+: \quad u\big(x_1;0;x_3\big) = u\big(x_1;b;x_3\big), \qquad w\big(x_1;0;x_3\big) = w\big(x_1;b;x_3\big). \tag{17}$$

In relationships (16) – (17) and (19) – (26):

$$0 < x_1 < a, \qquad 0 < x_2 < b, \qquad 0 < x_3 < h. \tag{18}$$

On the layer model cell we prescribed periodic boundary condition for the $C-/C+$ side surface in the plane of the side surface.

$$C-/C+: \quad u\big(x_1;x_2;0\big) = u\big(x_1;x_2;h\big), \qquad v\big(x_1;x_2;0\big) = v\big(x_1;x_2;h\big). \tag{19}$$

With this boundary condition we ensured that bending will not occur at the tension of the layer due to the decrease or increase of waviness of the roving. The straight line connecting the node pairs so remained perpendicular to the middle surface before deformation.

As we have already written earlier, we intend to use this layer model-cell for the determination of material properties of a given layer of a multilayer textile composite plate. Therefore, we cannot prescribe periodicity in x_3 direction for the $C-/C+$ side surface while modeling a layer of such a composite plate where every layer is the same (textile, layer thickness, fiber direction). Namely there is a very small likelihood that the textile reinforcement is positioned in the same way in the layers following each other in the composite plate. For this case a $[0°/0°/0°/0°]$ layer composition textile composite plate is shown as an example in Figure 13. The dashed lines mark the boundaries of the layers.
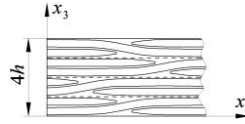
Figure 13
Positioning of the textile in a four layer composite plate

We need to connect the node pairs at the opposite edges, except the corner points of the model-cell, according to the (20) – (26) relationships:

On edges in x_1 direction: $u(x_1;0;0)=u(x_1;b;0)$, $w(x_1;0;0)=w(x_1;b;0)$, (20)

$$u(x_1;0;h)=u(x_1;b;h), \quad w(x_1;0;h)=w(x_1;b;h), \quad (21)$$

$$u(x_1;b;0)=u(x_1;b;h). \quad (22)$$

On edges in x_2 direction: $v(0;x_2;0)=v(a;x_2;0)$, $w(0;x_2;0)=w(a;x_2;0)$, (23)

$$v(0;x_2;h)=v(a;x_2;h), \quad w(0;x_2;h)=w(a;x_2;h), \quad (24)$$

$$v(a;x_2;0)=v(a;x_2;h). \quad (25)$$

On edges in x_3 direction: $w(0;0;x_3)=w(a;0;x_3)=w(a;b;x_3)=w(0;b;x_3)$. (26)

On the $C-$ side surface every corner has the same displacement in x_3 direction. Therefore, we connect the nodes at the corners in x_3 direction. The independent node is $N3$:

$$w_{N1}=w_{N2}=w_{N4}=w_{N3}. \quad (27)$$

On the *C+* side surface we also connect the nodes at the corners in x_3 direction.
The independent node is *N7*:

$$w_{N5} = w_{N6} = w_{N8} = w_{N7}. \tag{28}$$

Figure 14 shows the connection between the nodes at the opposite side surfaces
and edges. It is necessary to skip the edges at the opposite sides and the corners at
the opposite edges so that the model-cell will not be over-determined.

Figure 14
Connecting the nodes for ensuring the periodicity

At the tension in x_1 direction we numerically determined the following reaction
forces and the cross contraction:

$$\vec{F}_{A+} = -\vec{F}_{A-}, \ \vec{F}_{B-} = \vec{F}_{N1} = \vec{0}, \ v_{B+}.$$

We calculate the $\bar{\sigma}_1$ average normal stress as the quotient of the reaction force
appearing on the *A+* side surface and the area of the side surface according to (7).
From that we determine E_1 Young's modulus in the usual way. v_{12} Poisson's ratio
can be calculated from the numerically determined v_{B+} displacement and from the
prescribed u_{A+} displacement. These calculations can also be done for x_2 direction
in the similar way. At the tension in x_2 direction we numerically determined the
following reaction forces and the cross contraction:

$$\vec{F}_{B+} = -\vec{F}_{B-}, \ \vec{F}_{A-} = \vec{F}_{N1} = \vec{0}, \ u_{A+}.$$

Table 1 summarises the prescribed kinematic loading of the investigated textile
composite model cell as well as the numerically determined reaction forces, the
cross contraction and the calculated average stresses and material properties for
both loading cases.

Table 1
The given and determined quantities of the investigated textile composite model-cell

Tension in x_1 direction	
Kinematic loading	$\bar{\varepsilon}_1 = 2 \cdot 10^{-3}$, $u_{A+} = u_{N9} = 8 \cdot 10^{-3}$ mm
Numerically determined quantities	$\vec{F}_{A+} = \left(101.4\vec{e}_1\right)\mathrm{N}$, $\vec{F}_{A-} = \left(-101.4\vec{e}_1\right)\mathrm{N}$
	$v_{B+} = -8.271 \cdot 10^{-4}$ mm

Average stress	$\bar{\sigma}_1 = 101.4$ MPa
Material properties	$E_1 = 50\,700$ MPa , $\nu_{12} = 0.103$
Tension in x_2 direction	
Kinematic loading	$\bar{\varepsilon}_2 = 2 \cdot 10^{-3}$, $v_{B+} = v_{N11} = 8 \cdot 10^{-3}$ mm
Numerically determined quantities	$\vec{F}_{B+} = (101.4\vec{e}_2)\text{N}$, $\vec{F}_{B-} = (-101.4\vec{e}_2)\text{N}$
	$u_{A+} = -8.271 \cdot 10^{-4}$ mm
Average stress	$\bar{\sigma}_2 = 101.4$ MPa
Material properties	$E_2 = 50\,700$ MPa , $\nu_{21} = 0.103$

3.4   Shear Test Simulation with the Layer Model-Cell

Formula (6) of the pure shear looks as follows:

$$c_i^j = \Delta x_j \frac{1}{2}\bar{\gamma}_{ij}, \qquad c_j^i = \Delta x_i \frac{1}{2}\bar{\gamma}_{ji}, \qquad (i \neq j = 1,\,2). \tag{29}$$

For the calculation of G_{12} shear modulus of elasticity of the textile composite layer we simulated a pure shear on the $x_1 x_2$ surface. In this case the $\frac{1}{2}\bar{\gamma}_{12} = \frac{1}{2}\bar{\gamma}_{21}$ average shearing strain differs from zero in the strain tensor (1).

$$\frac{1}{2}\bar{\gamma}_{12} = \frac{1}{2}\bar{\gamma}_{21}, \qquad \bar{\gamma}_{12} = \frac{1}{2}\bar{\gamma}_{12} + \frac{1}{2}\bar{\gamma}_{21}. \tag{30}$$

We modeled the pure shear by a prescribed displacement in x_2 direction of the side surfaces perpendicular to the x_1 axis or by a prescribed displacement in x_1 direction of the side surfaces perpendicular to the x_2 axis of the layer model-cell. We realized the above kinematic prescriptions/conditions at the center points of the $A+$, $A-$, $B+$, $B-$ side surfaces by using the following values:

$$v_{A-} = v_{N10} = 0, \qquad v_{A+} = v_{N9} = a\frac{1}{2}\bar{\gamma}_{21}, \tag{31}$$

$$u_{B-} = u_{N12} = 0, \qquad u_{B+} = u_{N11} = b\frac{1}{2}\bar{\gamma}_{12}. \tag{32}$$

For the modeling of the shear test it was necessary to connect the nodes on the side surfaces perpendicular to x_1 and x_2 axes:

$$v_{A-} = v(0; x_2; x_3) = v_{N1}, \qquad v_{A+} = v(a; x_2; x_3) = v_{N2}, \tag{33}$$

$$u_{B-} = u\left(x_1; 0; x_3\right) = u_{N5}, \qquad\qquad u_{B+} = u\left(x_1; b; x_3\right) = u_{N8}. \qquad (34)$$

In relationships (33) – (34):

$$0 \le x_1 \le a, \qquad 0 \le x_2 \le b, \qquad 0 \le x_3 \le h. \qquad (35)$$

In the case of the $x_1 x_2$ plane shear the $C+$ and $C-$ side surfaces do not remain in plane. For prevention of the rigid body displacement and rotation we restrained the $N1$ node at the center of $C-$ side surface in x_3 direction:

$$w_{N1} = 0. \qquad (36)$$

In the case of a pure shear the node pairs must be connected at the opposite side surfaces, with the exception of the edges of the model-cell, according to (37) – (39). In the plane of $C-/C+$ side surface, similarly to the tension test simulation, we also provided periodic boundary condition.

$$A-/A+: \quad u\left(0; x_2; x_3\right) = u\left(a; x_2; x_3\right), \qquad w\left(0; x_2; x_3\right) = w\left(a; x_2; x_3\right). \qquad (37)$$

$$B-/B+: \quad v\left(x_1; 0; x_3\right) = v\left(x_1; b; x_3\right), \qquad w\left(x_1; 0; x_3\right) = w\left(x_1; b; x_3\right). \qquad (38)$$

$$C-/C+: \quad u\left(x_1; x_2; 0\right) = u\left(x_1; x_2; h\right), \qquad v\left(x_1; x_2; 0\right) = v\left(x_1; x_2; h\right). \qquad (39)$$

In (37) – (39) and (41) – (47) relationships:

$$0 < x_1 < a, \qquad 0 < x_2 < b, \qquad 0 < x_3 < h. \qquad (40)$$

In the case of the opposite edges, except of the corner points of the model-cell, we need to connect the node pairs as follows:

On edges in x_1 direction: $v\left(x_1; 0; 0\right) = v\left(x_1; b; 0\right)$, $w\left(x_1; 0; 0\right) = w\left(x_1; b; 0\right)$, (41)

$$v\left(x_1; 0; h\right) = v\left(x_1; b; h\right), \; w\left(x_1; 0; h\right) = w\left(x_1; b; h\right), \quad (42)$$

$$v\left(x_1; b; 0\right) = v\left(x_1; b; h\right). \qquad (43)$$

On edges in x_2 direction: $u\left(0; x_2; 0\right) = u\left(a; x_2; 0\right)$, $w\left(0; x_2; 0\right) = w\left(a; x_2; 0\right)$, (44)

$$u\left(0; x_2; h\right) = u\left(a; x_2; h\right), \; w\left(0; x_2; h\right) = w\left(a; x_2; h\right), \quad (45)$$

$$u\left(a; x_2; 0\right) = u\left(a; x_2; h\right). \qquad (46)$$

On edges in x_3 direction: $w\left(0; 0; x_3\right) = w\left(a; 0; x_3\right) = w\left(a; b; x_3\right) = w\left(0; b; x_3\right).$ (47)

For taking into consideration the periodicity we connect the nodes at the corners in x_3 direction on $C-$ side surface, the independent node is $N3$:

$$w_{N1} = w_{N2} = w_{N4} = w_{N3}. \qquad (48)$$

On the $C+$ side surface we also connect the nodes at the corner points in x_3 direction, the independent node is the $N7$:

$$w_{N5} = w_{N6} = w_{N8} = w_{N7}. \tag{49}$$

The numerically determined reaction forces at $x_1 x_2$ plane shear are the following:

$$\vec{F}_{A+} = -\vec{F}_{A-}, \quad \vec{F}_{B+} = -\vec{F}_{B-}, \quad \vec{F}_{N1} = \vec{0}.$$

Due to the duality of τ stresses the average shear stress must be identical on the side surfaces perpendicular to the x_1, x_2 axes:

$$\bar{\tau}_{12} = \frac{F_{1B+}}{A_{B+}}, \qquad \bar{\tau}_{21} = \frac{F_{2A+}}{A_{A+}}, \qquad \bar{\tau}_{12} = \bar{\tau}_{21}. \tag{50}$$

The stress distribution is never homogenous in the layer model-cell. Figure 15 shows the τ_{12} stress distribution, the deformation is illustrated with a zoom of $200\times$.
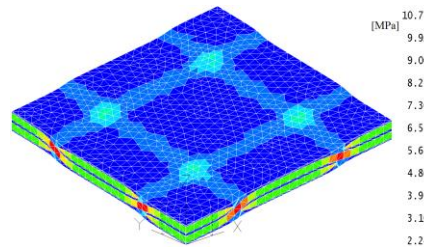
Figure 15
τ_{12} stress distribution in the layer model-cell at shear test in $x_1 x_2$ plane

Shear modulus in $x_1 x_2$ plane:

$$G_{12} = \frac{\bar{\tau}_{12}}{\bar{\gamma}_{12}}. \tag{51}$$

Table 2 summarises the given kinematic loading at the plane shear test of the textile composite model-cell as well as the numerically determined reaction forces, the average stresses and the shear modulus.

Table 2
The given and determined quantities of the investigated textile composite model-cell

Pure shear in the $x_1 x_2$ plane	
Kinematic loading	$\bar{\gamma}_{12} = 2 \cdot 10^{-3}$, $\frac{1}{2}\bar{\gamma}_{12} = \frac{1}{2}\bar{\gamma}_{21} = 10^{-3}$
	$v_{A+} = v_{N9} = 4 \cdot 10^{-3}$ mm, $u_{B+} = u_{N11} = 4 \cdot 10^{-3}$ mm

Reaction forces	$\vec{F}_{A+} = \left(5.8\vec{e}_2\right)\mathrm{N}$, $\vec{F}_{A-} = \left(-5.8\vec{e}_2\right)\mathrm{N}$
	$\vec{F}_{B+} = \left(5.8\vec{e}_1\right)\mathrm{N}$, $\vec{F}_{B-} = \left(-5.8\vec{e}_1\right)\mathrm{N}$
Average stresses	$\bar{\tau}_{12} = 5.8\ \mathrm{MPa}$, $\bar{\tau}_{21} = 5.8\ \mathrm{MPa}$
Shear modulus	$G_{12} = 2\,900\ \mathrm{MPa}$

3.5 Macroscopic Material Properties of the Composite Layer Determined by the Layer Model-Cell – Summary of Numerical Results

According to Sections 3.3 and 3.4, by carrying out simulations on the layer model-cell five orthotropic material constants can be defined, four of them are independent due to the requirements of (2) equation. However, in this case the studied composite layer has only three independent material constants because the reinforcement in the textile is the same in x_1 and x_2 directions (Figure 9):

$$E_1 = E_2 = 50\,700\ \mathrm{MPa}\,, \qquad \nu_{12} = 0.103\,, \qquad G_{12} = 2\,900\ \mathrm{MPa}\,. \qquad (52)$$

3.6 Characteristic Data of the Textile Composite Layer Used in the Experiment

Figure 16
Plain weave textile
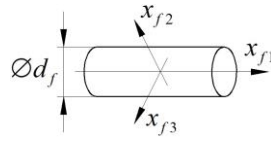
Figure 17
Material coordinate system of a carbon fiber

The composite is reinforced with a SIGRATEX KDL 8003 type plain weave carbon tissue. In Figure 16 we can see the plain weave textile before impregnation. The elementary fibers are ordered into a flat, untwisted roving. The type of the roving is Torayca T300-3K. 3K means that there are $n_f = 3000$ carbon fibers with $d_f = 7\ \mu\mathrm{m}$ in the roving.

Material properties of applied carbon fiber are known from the [7], [13] literature, in the $x_{f1}x_{f2}x_{f3}$ material coordinate system of Figure 17:

$E_{f1} = 230\,000\ \mathrm{MPa}$ $\nu_{f12} = \nu_{f13} = 0.166$ $G_{f12} = G_{f13} = 6\,432\ \mathrm{MPa}$

$E_{f2} = E_{f3} = 15\,000\ \mathrm{MPa}$ $\nu_{f23} = 0.400$ $G_{f23} = 5\,357\ \mathrm{MPa}$.

The matrix material is an AROPOL M105TB type polyester resin, which is linear elastic, isotropic material. The E_m Young's modulus and the v_m Poisson's ratio of the matrix were determined by measurement according to ASTM D638-10 standard [1]. The G_m modulus of elasticity was determined by the (53) relationship, that is valid for the isotropic materials:

$$E_m = 3\,677 \ \text{MPa}\,, \quad v_m = 0.346\,, \quad G_m = \frac{E_m}{2\left(1+v_m\right)} = 1\,365.9 \ \text{MPa}\,. \tag{53}$$

We produced layered composite sheets also from the polyester resin, from which we cut out the test specimens which we needed for the measuring of the material properties of the matrix material. The parameters of production process of the matrix material were identical to the ones applied at the production of the textile composite plates ($T = 65-68\ °\text{C}$, $p = 6\ \text{bar}$, $t = 2\ \text{h}$), therefore the material properties measured on these test specimen must be identical to the material properties of the matrix material of the investigated composite.

4   Validation of Results of the Layer Model-Cell by Measurement

The macroscopic material constants of a layer of an eight-layer textile composite laminate was determined by finite element simulation with the layer model-cell in Section 3. For validation of results by measurement we can only use an eight-layer composite laminate.

In the measurement investigated laminate all layers has the same orientation: $\left[0°/0°/0°/0°/0°/0°/0°/0°\right]$. The $x_1 x_2 x_3$ material coordinate system of the composite laminate is identical to the material coordinate system of a single layer. The material constants of the eight-layer textile composite plate are identical to the material constants of the single layers (52), therefore:

$$E_1^{8l} = E_2^{8l} = E_1 = 50\,700 \ \text{MPa}\,, \quad v_{12}^{8l} = v_{12} = 0.103\,, \quad G_{12}^{8l} = G_{12} = 2\,900 \ \text{MPa}\,.$$

The E_1 modulus of elasticity and the v_{12} Poisson's ratio of the eight-layer textile composite were determined by measurement according to EN ISO 527-4 standard [6]. The experimental examinations were performed with the test specimen that can be seen in Figure 18. In the experiment tensions took place in the material principal directions, $\vec{e}_1 = \vec{e}_x$ and $\vec{e}_2 = \vec{e}_y$. In the tension test the ε_1 longitudinal and the ε_2 transversal strain, the F tension force and the change of L length of specimen were measured with $t = 0.02\ \text{sec}$ sampling.
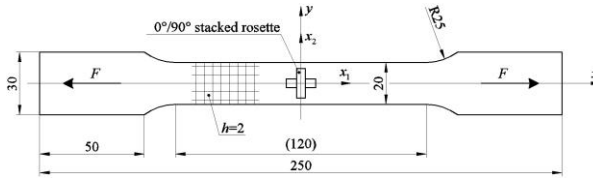
Figure 18

Tension test specimen
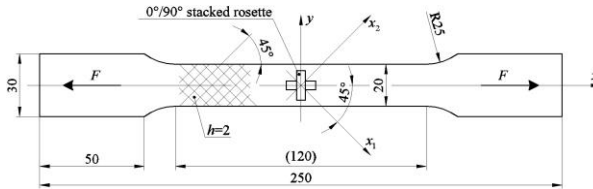
Figure 19

Shear test specimen

Figure 20

Test specimen in the
tension test machine

For determination of E_1 modulus of elasticity a $\sigma_1(\varepsilon_1)$ linear function was fitted on the $(\varepsilon_1; \sigma_1)$ measured point set by the least squares method. The slope of this linear function is the investigated E_1. For the determination of ν_{12} Poisson's ratio a $\varepsilon_2(\varepsilon_1)$ linear function was fitted for the $(\varepsilon_1; \varepsilon_2)$ measured point set by the least squares method. ν_{12} is equal to the -1 times the slope of the obtained linear function. We processed the measurement data by the Microsoft Excel program. We carried out the tension test by using six tension test specimens. One of these can be seen in Figure 20 in the tension test machine.

The G_{12} shear modulus was measured according to regulations of ASTM D3518 / D3518M – 13 standard [2]. The shape and dimension (Figure 19) of the shear test specimen was the same as the tension test specimen (Figure 18), however in the shear test specimen the reinforcing textile is positioned in every layer in 45° compared to the longitudinal axis of the test specimen. It means that the x_1 material principal direction had a –45° angle with the axis of the tension. At the shear test the ε_x, ε_y strains, the F tension force and the change of L length of specimen were measured with $t = 0.02$ sec sampling.

The γ_{12} shearing strain value from the ε_x, ε_y strains (coordinate transformation) and the τ_{12} shear stress from the σ_x normal stress could be computed. The τ_{12} stress is a linear function of the γ_{12} shearing strain. For determination of G_{12} shear modulus of elasticity a $\tau_{12}(\gamma_{12})$ linear function was fitted on the $(\gamma_{12}; \tau_{12})$ measured point set by the least squares method. The slope of this linear function is

the investigated G_{12}. The shear test was also carried out by using six test specimens.

The averages of the material constants were calculated from the six tension and six shear results. We determined the variance and variance square of the measured values as well as the deviation from the average. If the deviation was too large at any of the test specimens, we studied whether the given measurement has a mathematically or statistically detectable gross error. If we noted a gross error, we ignored the measured results belonging to the given experiment. Then we used Student's distribution to calculate the error range of the determined material constants [5].

Table 3 contains the computed and measured results for the E_1 modulus of elasticity, the ν_{12} Poisson's ratio and the G_{12} shear modulus of elasticity. We can observe that the material properties created by the layer model-cell are within the margin of error of the measurement. The deviation of computed results is below 4% compared to the average of the measurement. So, the layer model-cell fulfils the necessary accuracy required for engineering modeling. This fact confirms that the model-cell approach for determination of layer properties is an efficient and well applicable method for determination of the orthotropic material constants of textile composite layers.

Table 3
Comparing the measurement results to the layer model-cell results

Finite element layer model-cell	Measurement	Deviation of the results of model-cell from the average value of measurement	
$E_1 = 50\,700$ MPa	$E_1 = 50\,094 \pm 1\,480$ MPa	$+606$ MPa	$+1.21\,\%$
$\nu_{12} = 0.103$	$\nu_{12} = 0.102 \pm 0.016$	$+0.001$	$+0.98\,\%$
$G_{12} = 2\,900$ MPa	$G_{12} = 3\,018 \pm 254$ MPa	-118 MPa	$-3.91\,\%$

Conclusions

The paper presents a layer model-cell for the estimation and determination of the macroscopic, orthotropic material properties of a single layer of a multilayered textile composite material. It introduces the building up of the finite element layer model-cell and the prescription of boundary conditions applied to the finite element computations. It shows an example for determination of macroscopic material properties of the given textile composite layer with the numerical modeling of a unidirectional tension and plane shear by the layer model-cell.

Beside the determination of the material properties by the layer model cell the paper shows results for the same material by experimental way. The comparison of the results obtained from simulation and experiments confirms and proves the applicability of the layer model-cell and the roving model-cell in engineering.

References

[1]   ASTM D638-10, Standard Test Method for Tensile Properties of Plastics, 2010

[2]   ASTM D3518 / D3518M – 13 Standard Test Method for In-Plane Shear Response of Polymer Matrix Composite Materials by Tensile Test of a ±45° Laminate, 2013

[3]   Bojtár G., M. Csizmadia B., Égert J.: Numerical Estimation Method of Orthotropic Material Properties of a Roving for Reinforcement of Composite Materials, Acta Polytechnica Hungarica, Vol. 13, No. 6, p. 163-182, ISSN 1785-8860, 2016

[4]   Bunsell, A. R.; Renard, J.: Fundamentals of Fibre Reinforced Composite Materials, Series in Material Science and Engineering, Institute of Physics Publishing, 398 p., 2005

[5]   M. Csizmadia B.: Planning Experiments, SZIE Egyetemi Kiadó, Gödöllő, 2016 (In Hungarian)

[6]   EN ISO 527-4: Plastics – Determination of tensile properties – Part 4: Test conditions for isotropic and orthotropic fibre-reinforced plastic composites, 1997

[7]   Goda T.: Wear Mechanism of Composite-Steel Sliding Pairs, PhD Thesis, 2002, Budapest University of Technology and Economics, Faculty of Mechanical Engineering, Supervisor: Váradi K. (In Hungarian)

[8]   Hallal, A., Fardoun, F., Younes, R., Hage Chehade, F.: Evaluation of Longitudinal and Transversal Young's Moduli for Unidirectional Composite Material with Long Fibers, Advanced Materials Research, Vol. 324, p. 189-192, 2011

[9]   Kollár L. P.; Springer, G.: Mechanics of Composite Structures, Cambridge University Press, 480 p., 2003

[10]   Kroupa, T., Janda, P., Zemčík, R.: Linear Two-Scale Model for Determining the Mechanical Properties of a Textile Composite Material, Materials and Technology, Vol. 46, No. 2, p. 97-101, 2012

[11]   Leijan H.: Determining Micro- and Macro- Geometry of Fabric and Fabric Reinforced Composites, Dissertation, 2013, Department of Mechanical and Nuclear Engineering College of Engineering, Kansas State University, Manhattan, Kansas, Advisor: Youqi W.

[12]   Siva Bhaskara Rao Devireddy; Sandhyarani Biswas: Effect of Fiber Geometry and Representative Volume Element on Elastic and Thermal Properties of Unidirectional Fiber-Reinforced Composites, Hindawi Publishing Corporation Journal of Composites, Article ID 629175, 12 pages, 2014

[13]   Torayca T300 Factory catalogue (No. CFA-001) Toray Carbon Fibers America, Inc. (CFA), 2012

[14]   Xia, Z., Zhang, Y., Ellyin, F.: A Unified Periodical Boundary Conditions for Representative Volume Elements of Composites and Applications, International Journal of Solids and Structures, Vol. 40, No. 8, pp. 1907-1921, 2003

[15]   Zako, M., Uetsuji, Y., Kurashiki, T.: Finite Element Analysis of Damaged Woven Fabric Composite Materials, Composites Science and Technology, Vol. 63, Issues 3-4, p. 507-516, 2003

[16]   Zhang C.: Multi-Scale Characterization and Failure Modeling of Carbon/Epoxy Triaxially Braided Composite, Dissertation, 2013, The Graduate Faculty of The University of Akron, Advisor: Binienda, W. K.

[17]   Zhuo C.; Fan Y.; Meguid, S. A.: Multi-Level Modeling of Woven Glass/Epoxy Composite for Multilayer Printed Circuit Board Applications, International Journal of Solids and Structures, Vol. 51, Issues 21-22, pp. 3679-3688, 2014

Security Issues and Future Challenges of Cloud Service Authentication

Shu Yun Lim[1], M. L Mat Kiah[2], Tan Fong Ang[2]

[1]Faculty of Business Technology and Accounting, Unitar International University, 47301 Selangor Darul Ehsan, Malaysia; lim_sy@unitar.my

[2]Faculty of Computer Science & Information Technology, University of Malaya, 50603 Kuala Lumpur, Malaysia; misslaiha@um.edu.my, angtf@um.edu.my

Abstract: Authentication, in the cloud context, is the process of validating and guaranteeing the identity of cloud service subscribers or users. It is deemed essential since its strength directly impacts the reliability and security of the cloud computing environment. Many efforts, within in the literature, have surveyed cloud security and privacy, but lack a detailed analysis of authentication for the cloud. In view of the research gap and the importance of a valid authentication infrastructure, this survey critically investigates different authentication strategies and frameworks proposed for cloud services. This paper discusses the pros and cons of different authentication strategies and presents the taxonomy of the state-of-the-art cloud service authentication. The paper concludes with the open issues, main challenges and directions highlighted for future work in this relevant area.

Keywords: cloud computing; authentication; authentication-as-a-service; identity management; access control

1 Introduction

Cloud computing is widely adopted for delivering services such as data storage and management over the Internet. There are large varieties of cloud solutions and services to be accessed by a large number of devices such as workstations, smart phones and tablets. The security requirement becomes more complex with flexible content processing and sharing among a large number of users through cloud-based applications and services. When clouds contain applications from multiple organizations on a single managed infrastructure, application data is vulnerable not only to external attacks, but also to attacks from other organizations sharing the same infrastructure. Financial services corporations reported spending over a billion each year to safeguard against online security breaches. Unfortunately, these problems remained unsolved with many incidents of cloud service losing or

corrupting user data. In the latest hit, Adobe suffered significant security breaches compromising the data of 2.9 million customers and valuable source code after the company shifts to a cloud-based delivery model [1].

The cloud environment being distributed, in nature, is facing challenges in managing user's identity, authenticating and authorizing users. Cloud service providers have access to the information stored by subscribers for authentication purpose and this presents a privacy issue to their private information. It is difficult for cloud subscribers or users to make sure the proper Service Level Agreement (SLA) rules are enforced since there is a lack of transparency in the cloud that allows the users to monitor their own information. Besides, cloud users who subscribed to multiple cloud services will have to store passwords in all the clouds for authentication and hence authentication data are replicated and withheld in multiple clouds. These redundant actions of exchanging authenticating data may lead to an exploit of the authentication mechanism.

From the cloud service providers' standpoint, managing and authenticating users in the cloud is becoming inevitably complicated. More specifically, the process of authentication, which is to verify an entity that is trying to access protected resources, is very visible to users. It directly influences their perception of trust. Cloud providers and brokers try to overcome security and privacy-related issues by offering security solutions to its customers. Security-as-a-Service (SEaaS) is a new instance of a cloud service model that delivers security solutions to enterprises by means of cloud-based services from the cloud. These services may be delivered in different forms, and Authentication-as-a-Service (AaaS) is one of the variants. AaaS is the new form of user authentication that cloud users should embrace for mitigating the risk of compromising sensitive information. However, when we look at authentication aspects of cloud computing, most discussions today point towards various forms of identity federation between clouds. As stated previously, there is no information in the literature discussing the authentication strategies and the architectural perspective of the components that together form a strong authentication system. This paper helps to bridge the gap in the literature that would be of interest to both academia and industry alike.

Section 2 provides the overview and the research approach undertaken to investigate cloud service authentication. The taxonomy of cloud service authentication security is presented in Section 3. Section 4 highlights the current state-of-the-art work proposed to authenticate cloud services and their potential problems. Finally, conclusions and directions for future research are identified in Section 5.

2   Overview

In this paper, we provide a survey of cloud service authentication strategies with the approach depicted in Figure 1. The NIST cloud security architecture [3] [4] is adopted and referenced for the baseline terms and definition. The taxonomy of cloud AaaS is proposed after deriving the main properties for cloud service authentication. Literature in the field is analyzed and the authentication strategies highlighted. The previous steps provide the input to derive the open issues and the future directions in the field of cloud service authentication.
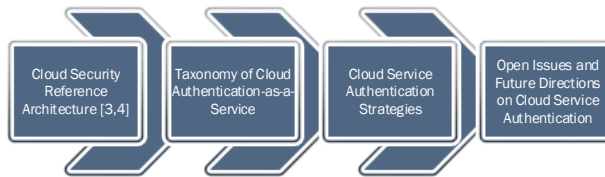
Figure 1
Research Methodology

Following the reference of NIST definition [3] and [4], the stakeholders involved in cloud service authentication are cloud providers, brokers, consumers, auditors and carriers. A cloud consumer represents a person or organization that maintains a business relationship with, and uses the service from, a cloud provider. The cloud provider of SaaS assumes most of the responsibilities in managing and controlling the applications and the infrastructure, while the cloud consumers have limited administrative control of the applications. Cloud broker on the other hand is an entity that manages the use, performance, and delivery of cloud services, and negotiates relationships between cloud providers and cloud consumers. Different aspects of the secure cloud service management can be supported and implemented by either a cloud provider or by a cloud broker, depending upon the structure of each cloud ecosystem.

Meanwhile a cloud auditor is a party that conducts independent assessment of cloud services, information system operations, performance, and the security of a cloud computing implementation. A cloud auditor can evaluate the services provided by a cloud provider in terms of security controls, privacy impact, performance, and adherence to service level agreement parameters. A cloud carrier acts as an intermediary that provides connectivity and transport of cloud services between cloud consumers and cloud providers. It provides access to consumers through network, telecommunications, and other access devices.
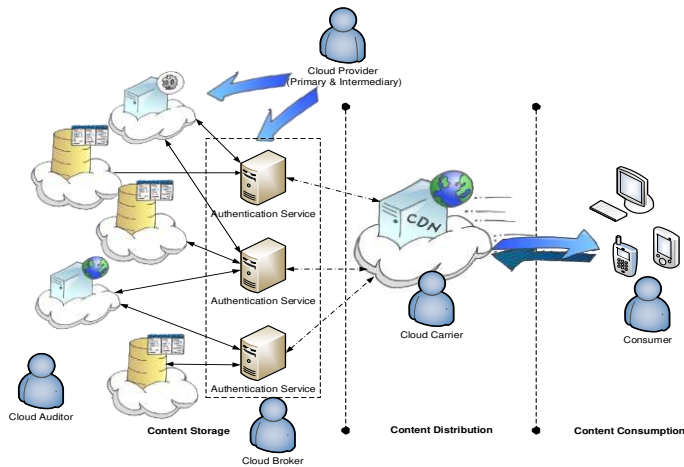
Figure 2

Stakeholders in cloud service authentication

3   Cloud Service Authentication Architecture
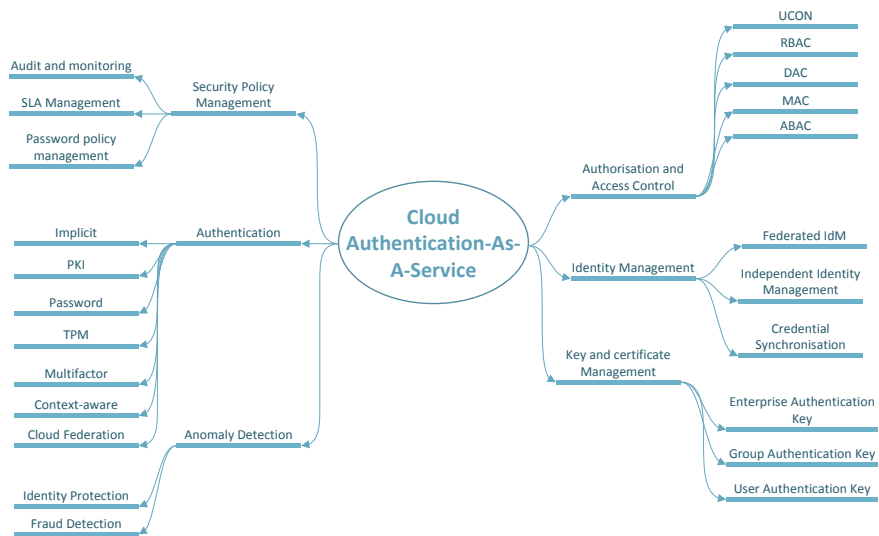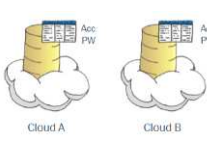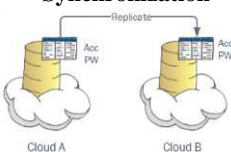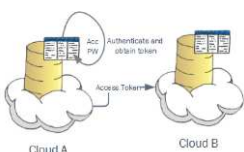
Figure 3

Taxonomy of cloud Authentication-as-a-Service

Both cloud providers and consumers are concerned about security issues associated with the cloud environment. Besides upholding the confidentiality of the system which is a generic security requirement across clouds, different cloud services may have different security and policy characteristics corresponding to specific functionality and usage of the system. The important aspects of cloud authentication service provisioning include Identity Management, Authentication, Access Control and Authorization, Security Policy Management, Key and Certificate Management and Fraud and Anomaly Detection. Taxonomy of the cloud service authentication infrastructure is presented in Figure 3.

3.1   Identity Management

Identity management (IdM) refers to the mechanisms and standards that create and maintain a cloud user's identity, and the de-provision of the user account when the user leaves the cloud service. IdM is a broad administrative area that deals with identifying individuals in a system and controlling access to the resources by placing restrictions on the established identities. Sound identity management and governance are needed to manage identities in hybrid cloud environments, a combination of cloud services and enterprise networks. Identity management in cloud involves three perspectives, namely the identity provisioning paradigm, the log-on paradigm, and the service paradigm [2]. Identity management is required to simplify the user provisioning process especially in enterprise environment. Enabling new users to get access to cloud services and de-provisioning users to ensure that only the rightful users have access to cloud services and data. This creation and deletion of identities are done without regard to user access rights to the service. The log-on paradigm involves exchanging of data by users to log-on to a cloud service. Lastly the service paradigm delivers personalized services to users and their devices after successful log-on attempts. There are three different models adopted by SaaS vendor to provide sign on and identity management service, namely independent IdM, credential synchronization, and federated IdM. Differences between the models are depicted in Table 1.

Table 1
Comparisons of IdM Model

IdM Model	Independent IdM	Credential Synchronization	Federated IdM
Description	• Consumer account information is managed by individual cloud	• Consumer account is replicated and information is shared between clouds	• Consumer account is managed independently

Strengths	• Easy to implement with no enterprise directory integration required	• Consumer manage fewer credentials	• No enterprise directory integration required • Lower risk as credentials are not replicated but propagated on demand
Weaknesses	• Consumers need to manage separate credentials for different accounts	• Requires integration with enterprise directory. • Higher security risk due to replication and distribution of consumer credentials	• Relatively more complex to implement • Require proper agreement and trust relationship between cloud services.

3.2   Authentication Strategy

Authentication is the process for confirming the identity of the user. Cloud consumers have to complete the user authentication process required by the cloud provider. The cloud provider can choose to provide different authentication mechanisms with different security strength and the strength depends on the reliability and integrity of the mechanism. Besides, authentication mechanisms must function properly in order to maintain data confidentiality and integrity. In view of the emergence of hybrid cloud environments, interoperability for user authentication has also become a major concern. The differences in authentication strategies are detailed in Table 2.

3.2.1   Password Authentication

Password authentication is simple and easy to use, but it has to have a certain level of complication and regular renewal to keep the security [3]. It is an authentication technology with well-known weaknesses in the sense that even if the correct username and password combination is provided; it is still difficult to prove that the request is from the rightful owner. Users frequently reuse their passwords when authenticating to various cloud services. Weak password practice brings high security risks to the user account information. Nonetheless, password authentication is still the most frequently used authentication technology with more than 90% of the transactions. Password authentication comes in different forms of challenge-response protocol in current cloud deployment.

3.2.2   Trusted Platform Module-based Authentication

Trusted Platform Module (TPM) is a hardware-based security module that uses secure cryptoprocessor that can store cryptographic keys that protect information. A variant of it, Mobile Trusted Module (MTM) [4] is a proposed standard by

Trusted Computing Group a consortium (TCG) founded by AMD, Hewlett-Packard, IBM, Intel, Microsoft. It is mainly applied to authenticate terminals from telecommunications. However, it is being considered as a cloud computing authentication method with Subscriber Identity Module (SIM) due to the generalization of smartphones. Cloud subscribers' devices can utilize unique hardcoded keys to perform software authentication, encryption, and decryption. TPM chips can also be used for other security technologies such as firewalls, antivirus software, and biometric verification. However, there are some inherent problems with the TPM technology, such as when an attacker manages to bypass disk encryption during a cold boot attack and reveal the master password with social engineering. In cloud environment, the biggest challenge is still the "Bring-your-own-device" (BYOD) concept that does not facilitate the implementation of TPM devices in enterprise network.

3.2.3   Public Key Infrastructure-based Authentication

Employing Trusted Third Party (TTP) services within the cloud leads to the establishment of the necessary trust level and provides ideal solutions to preserve the confidentiality, integrity and authenticity of data and communication. The public key infrastructure (PKI), together with TTP, provide a technically sound and legally acceptable means to implement strong authentication and authorization. PKI is an authentication means using public-key cryptography. It enables users to authenticate the other party based on the certificate without shared secret information. One example of TTP authentication in cloud is Single-Sign-On (SSO). When a user gets authentication from a site, it can go through to other sites with assertion and no authentication process is required. However, the existence of a trusted third party, as an authentication server or certification authority, is becoming a security and fault intolerant bottleneck for systems.

3.2.4   Multifactor Authentication

Multi-factor authentication ensures that a user is who they claim to be by combining a few means of authentication. The more factors used to determine a person's identity, the greater the trust of authenticity [5]. ID, password, biometrics and certificates are used traditionally for single factor authentication. With the emergence of mobile network, second factor authentication takes the form of SMS, e-mail, and telephony OTPs, PUSH Notifications, and mobile OATH Tokens. Even though it is rather effective for closed communities such as enterprise cloud, these second factor methods are too costly, inconvenient, and logistically difficult especially for the distribution, administration, management and support in the cloud.

3.2.5   Implicit Authentication

This approach uses observations of user behavior for authentication and it is well-suited for mobile devices since they are capable of collecting a rich set of user information, such as location, motion, communication and their usage of applications. A number of profiling techniques have been studied to provide a suitable service for user and personal profile information in the mobile cloud environment [6] [7] [8]. But to date, a formal model for this approach has yet to be realized and limited device resources are the technical constraints that need to be overcome. Studies on intelligent mobile authentication service are still inadequate.

Table 2

Strengths and weaknesses of the authentication strategies

Authentication Strategy	Strengths	Weaknesses
Password Authentication	Simple to deploy	Regular renewal to keep the security due to deterministic function of storing password
Trusted Platform Module based authentication	Low deployment cost if TPM is integral part of an enterprise system Could be extended to incorporate other security services	Inherent TPM weaknesses such as cold-boot attack and master password revelation through social engineering BYOD does not facilitate TPM-based authentication in enterprise cloud
Public Key Infrastructure based authentication	Easy establishment of the necessary trust to provide cryptographically strong authentication solutions	Challenging certificate and key management in distributed environment
Multifactor authentication	Greater trust of authenticity	Logistically difficult for the distribution, administration, management and support in the cloud
Implicit authentication	Cloud subscriber device is capable of collecting a rich set of user information for profiling User authentication can be done dynamically, improves usability of authentication	Data collected for profiling risk disclosing user private information Limited device resources are the technical constraints to overcome
Context-aware cloud authentication	Context-aware cloud constantly retrieves structured information from users and profile users through active classification and inference User authentication can be done dynamically, improves the usability of authentication	Lack of privacy of user authentication data in the cloud

3.2.6   **Context-Aware Cloud Authentication**

While a full user and device profiling is not feasible as this stage, a context-aware cloud can assist in the dynamic and adaptive authentication method. Inherent components in this are enhanced authentication services, flexible access control and an adaptable security subsystem responding to current conditions in the environment. Since context-aware cloud constantly retrieves structured information from users and through active classification and inference, a model of the user who has legitimate access to systems and resources can be built. Given the characteristics of a context aware cloud, authentication can be done dynamically depending on the changing conditions of user's risk factor at a particular time. This type of risk-based authentication goes hand in hand with implicit authentication. It is a non-static authentication system which takes into account the profile of users requesting access to the system and relates the risk profile associated with that transaction. Higher risk profiles lead to stronger authentication need, whereas a static username and password may suffice for lower-risk profiles. Adapting authentication levels based on risk increases security beyond secret-knowledge techniques and provides transparent authentication without inconveniencing the user. Users continuously re-authenticate themselves to the cloud service to maintain confidence in their identity.

3.3   **Authorization and Access Control**

Authorization is of vital importance since it involves determining what the user is allowed to do after they have gained access. Authorization can be determined based on the user identity alone, but in most cases, it requires additional attributes about the user, such as their roles or titles. Meanwhile access control is more concerned with allowing a user to access a number of cloud resources. Though this process is typically handled by the applications being accessed, there is now consideration for centralizing the authorization policy decisions regardless of the location of the user or the application. The implementation of proper access control models for the cloud is one of the areas that has been critically evaluated since current access control models are not specifically designed to tackle the requirements of cloud systems.

As data and applications are shifted to the cloud, new challenges emerge to manage consistent and unified access policies for enterprises. Authorization is the means for ensuring that only properly authorized users are able to access resources within a system. An Authorization Service (AS) is responsible for evaluating an authorization query, collecting necessary information about the user and the resource, potentially from an attribute service and identity directory, and evaluating a policy to determine if access should be granted or denied.

Extensive research is being carried out in the area of access control in collaborative systems [9] [10], namely the Mandatory Access Control policies

(MAC), the Discretionary Access Control policies (DAC) and the Role Based Access Control policies (RBAC). Each one of them serves specific security requirements in different working environments. It is noteworthy that an attempt started along with the advancement of RBAC [11] for the design of a series of Attribute Based Access Control models (ABAC) [12] The ABAC model was mainly introduced to overcome a number of RBAC's shortcomings. Besides, UCON [13] [14] access control model is also introduced, along with RBAC, being the most prominent access control models for the Cloud. Nonetheless, further examination is demanded, due to the partial or weak fulfillment of security requirements in the Cloud.

3.4   Anomaly and Fraud Detection

The cloud consumer account and identity can be well protected by using a baseline user profile. Historical information about past user transactions, IP geolocation data, device authentication can be fed into a heuristic engine and by allowing the forming of user profiles, anomalies can be detected. Services such as fraud detection are also offered based on customized rules. Consumer transactions are analyzed in real-time to detect any activities that has similar patterns associated with crimes.

3.5   Security Policy Management

Security policy has become a critical concern of IT and businesses, in general. More specifically, security policy management in AaaS involves the SLA, password policy, audit and monitoring of the service. In order to properly audit the access or management of data governed by a cloud platform, all security operations based upon security identities and policies need precise audit information to be recorded. Meanwhile the purpose of the SLA is to define the basis for interoperable authentication or identity management solutions between consumers and providers [15].

3.6   Key and Certificate Management

Managing access to cloud services includes handling encryption keys and certificates. Key and certificate management have both been offered as part of the security to improve security, compliance and operational efficiency. In a case where the cloud service is being used to store a master encryption key, the data owner who deposited the key can define policies for how that key can be retrieved. This is important for establishing the necessary step in preventing unauthorized access and meeting compliance requirements for safeguarding keys. Centralizing management of keys and certificates on a cloud-based manager

allows the storage of any security object, enforcement of retrieval and revocation policies, and prevention of unauthorized access to sensitive data and systems.

The implementation of key and certificate management service is not without its challenges. The additional complexity in cloud environments as compared to enterprise IT environments is due to the difference in ownership (between cloud consumers and providers) and the control of infrastructures on where the management system and security artifacts are located [16]. However with the correct implementation, this AaaS component provides an effective security layer against malicious attacks and streamlines the auditing process.

4   Related Work

In this section, solutions that have been proposed in academic research pertaining to cloud service authentication is presented and discussed. A summary of the literature presented in Table 3. Authentication architecture and approaches are illustrated in Figure 4.

4.1   Trusted Platform Module-based Authentication

A. Ahmad et al. proposed that cloud services rely on the existing International Mobile Subscriber Identity (IMSI) and Universal Subscriber Identity Module (USIM) cards for authentication [17]. Since mobile network authentication in GSM/ 3G/ WLAN is the first level of security authentication for a mobile subscriber, it could be a valuable asset for the security foundation of cloud computing. The framework includes a mapping between users and services, to determine if a user may access a specific service; and a mapping of IMSIs and user IDs to be recognized as the same user, allowing multiple devices to share the same account. By expanding the use of the trusted platform and USIM, and the introduction of virtualization, mobile cloud services authentication can be achieved.

Z. Song et al. proposed TrustCube [18], a policy-based cloud authentication platform using open standards that supports the integration of various authentication methods. The scheme uses Trusted Computing technologies like TPM, TNC and remote attestation to ensure trust in users, platforms and environment of the platform. The scheme includes an Integrated Authenticated (IA) Service that retrieves policy for access request, extract information and send inquiry to IA server though a trusted network. TrustCube is an end-to-end infrastructure that offers measurements of essential elements of clients, platform and the environment. The cloud service can make an informed decision based on the certifiable report of measurements. Besides, cloud service can also evaluate

the risk of dealing with particular users. The authentication phase is token-based, where successful authentication a token is generated and used for further request from the same user. Biometric characteristics such as fingerprints and palm veins are used to provide biometric reference data in a shared secret key calculation. Nonetheless the proposal is based on the Trusted Platform Module (TPM) of corresponding clients' device for generating secure keys and detecting changes to client's platform (Figure 4, authentication flow 3). However, this is not feasible when enterprise users bring their personal smartphones and tablets into the workplace and use them to access corporate networks. The idea of BYOD (Bring Your Own Device) presents a unique security challenge for IT organizations and is driving the need for stronger authentication and access control policies on employee-owned mobile devices.

4.2   Implicit Authentication

R. Chow et al. proposed a flexible platform [8] for supporting authentication decision based on a behavioral authentication approach [6] [19], to translate user behavior into authentication score and thus allowing users to access cloud services transparently (Figure 4, authentication flow 3). This approach results in a new authentication paradigm which successfully strikes a balance between usability and trust. However, these behavior profiling techniques proposed have not been successfully implemented for mobile computing platforms due to complexity and intensive computation.

4.3   Context-Aware Cloud Authentication

Covington, M. J., *et al.* [8] proposed that context aware cloud to focus on security services incorporating the security-relevant context, making policy enforcement and access control flexible within a system-level service architecture. Manjea Kim et al. [20] and H. Jeong *et al.* [7] on the other hand make use of the context-aware platform of the cloud and proposed a scheme that considers user's context information and profile for authentication (Figure 4, authentication flow 1&2). This platform is able to gather the users' personal information and preferences as well to provide suitable services for them. The scheme includes methods that interpret and infer the high level context and resources management technique that manages distributed IT resources effectively. Further to that, H. Ahn *et al.* extended the context aware cloud authentication to include access control [21] [22]. The authors proposed a context-aware RBAC model which provides efficient access control to user through active classification, inference and judgment of the users assessing systems and resources.

4.4   Public Key Infrastructure-based Authentication

Many efforts have been made to improve identity management and authentication for cloud computing. Most of them focus on designing a trusted third party (Figure 4, authentication flow 2). For instance, Z. Wang *et al.* [23] [24] constructed a TTP with a homomorphic signature for identity management and access control in mobile cloud computing. In the identity management scheme, a mobile user firstly computes a full signature on all his sensitive personal information and stores it in a TTP. During the valid period of the full signature, the user can authenticate his/her identity to the cloud service provider through TTP. Meanwhile, for authorization and access control, a user's full signature on all identity attributes is stored in advance in the access controlling server. For a user who wants to access a cloud service, which has special requirement on one's identity attribute, the user only needs to notify the access controlling server of the particular cloud service name. According to the user's instruction, the access controlling server can compute a partial signature on the special identity attribute, and send it to the cloud server for identification. However, the scalability and extensibility of this scheme are under scrutiny. In the event of identity addition, update and removal, a full signature computation is required. This is very time-consuming and costly due to the heavy computation, and users are usually on very limited computational resources.

H. Li *et al.* [25] and Qin B. *et al.* [26] both proposed a framework for secure authentication and data upload in an identity-based setting. The identity-based feature eliminates the complicated certificates management in signature encryption schemes in the traditional public key infrastructure (PKI) setting. Besides, Mishra, D. *et al.* proposed another identity-based mutual authentication in cloud storage sharing using Elliptic Curve Cryptography (ECC) and claimed that their scheme can resist various attacks in the cloud infrastructure [27].

4.5   Password Authentication

Password-based authentication is relatively easy to construct and deploy (Figure 4, authentication flow 1). However in cloud deployment, without the randomization of password, a user is susceptible to dictionary attacks since the server has a deterministic function of the user's password. To overcome such a limitation, the authentication protocol must be a challenge-response type protocol where the server never learns any deterministic function of the client's password. Besides, to prevent malicious servers and cross-site impersonation, we require that the server never learn the client's password. Asymmetric Password-Authenticated Key Exchange (APAKE) attempts to remedy this problem. Only the client knows the password, while the server stores a one-way function of the password. However, APAKE schemes are still vulnerable to dictionary attacks to the server. Other notable research in this area includes single-password authentication scheme [28]

and a multi-level authentication technique which generates and authenticates the password in multiple levels to access the cloud services [29]. A more secure scheme is the multi-factor authentication that requires a second factor (such as finger print, token, OTP) with username/password proposed by [30] [31]. Nevertheless, the feasibility of two-factor authentication is largely limited by high device cost and the deployment complexity.

4.6   Cloud Federation Authentication

Federated identity is a useful feature for identity management and single sign on. OAuth, OpenID and SAML are the main concepts for federated cloud service authentication (Figure 4, authentication flow 2). Celesti A. *et al.* [32] proposed the "Horizontal Federation" of cloud resources. One cloud service provider, lacking in internal resources, can cooperate with another cloud service provider in order to supplement required resources by means of external ones. The model consists of three phases: discovery of available external cloud resources, matchmaking selection between discovered cloud providers, and authentication for trust context establishment with selected clouds. The main focus of this model is the authentication phase, which is the cloud SSO. Through Cloud SSO a cloud provider authenticates itself with other heterogeneous cloud providers regardless of their implemented security mechanism and accesses all needed external cloud resources. In order to establish trust relationship between home and foreign clouds, a trusted Identity Provider is required to verify digital identities of clouds and provides SAML authentication assertions.
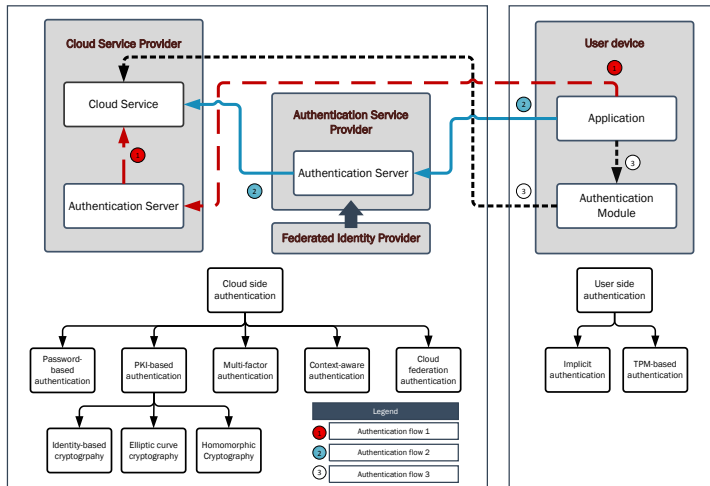
Figure 4
Authentication architecture and strategies

Table 3
Summary of the evaluated cloud service authentication schemes

Scheme	Basic Theory	Type	Description	Idm[1]	Auc[2]	Ac[3]
[20]	Context aware profiling	User profiling	Context aware architecture with user profiling	No	Yes	Yes
[7]	User profiling	User profiling	User profiling for mobile cloud authentication	No	Yes	Yes
[21]	Context aware profiling	User profiling	Context aware architecture with user profiling and RBAC for access control	No	Yes	Yes
[33]	Group Location-based security framework	TPM, Risk-based	Authentication with IMSI and location based service	No	Yes	No
[23]	Homomorphic Signature	PKI,TTP	Access control using Boneh-Boyen signature	No	Yes	Yes
[24]	Homomorphic Signature	PKI,TTP	Homomorphic signature for identity management and access control	Yes	Yes	Yes
[25]	Identity-based cryptography	PKI,TTP	Identity-based authentication protocol for cloud services	No	Yes	No
[30]	Anonymous One-Time Password	Password, PKI	Two factor authentication with OTP & RSA signature	No	Yes	No
[27]	Identity-based cryptography	PKI,TTP	Identity-based (ECC) authentication framework	No	Yes	No
[29]	Password concatenation	Password	Multi-level password concatenation for authentication	No	Yes	No

---

[1] Identity Management
[2] Authentication
[3] Access Control

[8]	User profiling	Implicit, Risk-based	Cloud authentication framework with behavioral authentication approach	No	Yes	No
[6]						
[28]	Single password authentication	Password	Challenge-response authentication protocol for cloud	No	Yes	No
[32]	Cross cloud federation authentication	Cloud federation	Cloud federation authentication	Yes	Yes	No
[34]	Cross cloud federation authentication	Cloud federation	Cloud federation authentication in enterprise	Yes	Yes	Yes
[17]	Trusted Platform Module based authentication	TPM, TTP	USIM/SIM-based authentication framework	No	Yes	No
[18]	Trusted Platform Module based authentication	TPM, biometrics	End-to-end Infrastructure built on TPM for authentication	No	Yes	No

5   Open Issues and Future Directions

Based on the related literature, there are several issues that have not been sufficiently addressed. The gap in the existing solutions would prove to be the direction for future work.

5.1   Privacy of Authentication Data

Although an issue of paramount importance, little research has been carried out in this regard. Existing cryptographic techniques can be utilized for data security but privacy protection and outsourced computation need significant attention. Personal data should always remain in the user control, and the user decides what and whom they share their data with. Especially when implicit and context aware cloud authentication strategy is used, the identity provider needs access to real-time information about the user. They need to feel confident when supplying their context information for profiling and authentication, and at the same time ensure that their privacy would not be violated.

Furthermore, data in the cloud typically resides in a shared environment. Even with SLA in place to support contract negotiation and enforcement, privacy and trust are non-quantitative and thus difficult to bargain. The challenge is that only

authorized entities can access the data. There is a need for appropriate mechanisms to prevent cloud service providers from misusing customer's data.

5.2   BYOD Challenge in the Cloud (Bring Your Own Device)

Along with the growth of mobile consumer devices in the enterprise cloud, securing various types of employee-owned device access to cloud services has become a critical component of the IT value-chain. Unfortunately, classical authentication mechanisms, such as TPM-based authentication cannot respond to the new challenges. BYOD also brings true challenge to develop access control policies in enterprise and hybrid cloud environment.

5.3   Usable and Scalable Authentication

Research should be focusing on the ability to deliver authentication services that are considered usable and scalable across cloud environments. They also should be easy to learn, use, administer, and inexpensive to maintain. The question remains in how the usability of authentication mechanism can be improved, where user logs in once and gain access to all services without being prompted to log in again at different cloud service. Identity federation is the way to go but solutions and control policies must be enforceable across clouds which are difficult to coordinate. While implicit and adaptive type of authentication is trying to increase the usability of the authentication by making authentication as transparent and seamless as possible, they have yet to provide sufficient confidence in realizing a full secure authentication mechanism.

5.4   Future Work

Future research can be directed at putting trust back to the users to ensure that they are in full control of their data. In order to instill the trust in cloud users, cloud provider's technical competency has to be enhanced and at the same time the data owner should have full control over who has the right to use their data and what they are allowed to do with it once they gain access. This is where homomorphic encryption comes into the picture. Homomorphic encryption is a form of encryption which allows specific types of computations to be carried out on cipher text and generate an encrypted result which, when decrypted, matches the result of operations performed on the plaintext. This is a desirable feature in modern communication system architectures that allow us to maintain confidentiality and privacy of outsourced data in cloud. With homomorphic encryption, only the users are equipped with encryption keys while operations are allowed over encrypted bits. This advancement put the trust back to the users that they are in full control of their data.

In addition to the increasing dominance of the mobile device as a primary point of access to cloud services, Cloud Security Alliance (CSA) specifically highlighted the need to provide usable and scalable authentication from mobile devices to multiple, heterogeneous cloud providers as an important step toward the maturity of cloud solutions. Current cloud-based authentication service offers solutions like single sign-on (SSO) to help simplify the management of access to services both in the cloud and behind the firewall. Single Sign-On has inherited the limitation of a password-based authentication and poses significant risks. Besides, its usability has rarely been investigated. In SSO once a user entered its credentials, the user gets signed in to all the subscribed services and re-authentication is required only after the credential is stale or time-out. This undeniably exposes users to more threats and security attacks. We suggest re-authentication when user is requesting for different types of service with different security requirements. In this case users re-enter credential only when the need arises and this can be a risk-based decision. With user risk profiling in place, an authentication service can then decide if the requested service implicitly needs a fresh authentication based on the risk factor. A risk profile of the cloud user is to provide a non-subjective understanding of risk by assigning numerical values to variables representing the threats and danger the user pose. It can be a useful tool for determining the magnitude of authentication required in cloud service authentication. Authentication that requires user mediation only when necessary addressed both the security and usability challenges.

Conclusion

Significant opportunities exist today to develop a strong cloud service authentication mechanism. A complete solution that incorporates identity management, authentication mechanisms, authorization and access control will make a substantial contribution to a correct and effective system of authentication. Organizations must ensure that service providers provide the flexibility to deliver varying levels of strong authentication to meet the required security policies and extend existing security implementations by incorporating identity federation. Addressing the security and usability challenges is a good direction in which to go and a more promising open standard authentication mechanism is needed to achieve a secure cloud ecosystem.

References

[1]   Brandon, J., Adobe data breach highlights the company's security shortcomings, experts say. 2013

[2]   Subashini, S. and V. Kavitha, A survey on security issues in service delivery models of cloud computing. Journal of Network and Computer Applications, 2011. 34(1): pp. 1-11

[3]   Keszthelyi, A., About Passwords. Acta Polytechnica Hungarica, 2013. Vol. 10, No. 6

[4]   Recordon, D. and B. Fitzpatrick, OpenID Authentication 1.1. Finalized OpenID Specification, May, 2006

[5]   Nagaraju, S. and L. Parthiban, SecAuthn: Provably Secure Multi-Factor Authentication for the Cloud Computing Systems. Indian Journal of Science and Technology, 2016. 9(9)

[6]   Markus Jakobsson, E.S., Philippe Golle, Richard Chow, Implicit authentication for mobile devices, Proceedings of the 4th USENIX conference on Hot topics in security. 2009, USENIX Association: Canada. p. 9-9

[7]   Jeong, H. and E. Choi, User Authentication using Profiling in Mobile Cloud Computing. Aasri Conference on Power and Energy Systems, 2012. 2: pp. 262-267

[8]   Chow, R., et al., Authentication in the clouds: a framework and its application to mobile users, Proceedings of the 2010 ACM workshop on Cloud computing security workshop. 2010, ACM: USA. pp. 1-6

[9]   De Capitani di Vimercati, S. and P. Samarati, Mandatory Access Control Policy (MAC), in Encyclopedia of Cryptography and Security, H.A. van Tilborg and S. Jajodia, Editors. 2011, Springer US. pp. 758-758

[10]   Sandhu, R.S. and P. Samarati, Access control: principle and practice. Communications Magazine, IEEE, 1994. 32(9): pp. 40-48

[11]   Tang, Z., et al., A new RBAC based access control model for cloud computing, in Advances in Grid and Pervasive Computing. 2012, Springer. pp. 279-288

[12]   Ruj, S., M. Stojmenovic, and A. Nayak, Decentralized Access Control with Anonymous Authentication of Data Stored in Clouds. Parallel and Distributed Systems, IEEE Transactions on, 2014. 25(2): pp. 384-394

[13]   Park, J. and R. Sandhu, The UCON ABC usage control model. ACM Transactions on Information and System Security (TISSEC), 2004. 7(1): pp. 128-174

[14]   Danwei, C., H. Xiuli, and R. Xunyi, Access Control of Cloud Service Based on UCON, Cloud Computing, 2009, Springer Berlin Heidelberg. pp. 559-564

[15]   Brussels, Cloud Service Level Agreement Standardisation Guidelines. 2014

[16]   Chandramouli, R., M. Iorga, and S. Chokhani, Cryptographic Key Management Issues and Challenges in Cloud Services. 2014: Springer

[17]   Ahmad, Z., et al., Considerations for mobile authentication in the Cloud. Information Security Technical Report, 2011. 16(3-4): pp. 123-130

[18]   Song, Z., et al., TrustCube: An Infrastructure that Builds Trust in Client, in Future of Trust in Computing, 2009, Vieweg+Teubner. pp. 68-79

[19] Hocking, C.G., et al., Co-operative user identity verification using an Authentication Aura. Computers & Security, 2013. 39, Part B: pp. 486-502

[20] Manjea Kim, H.J., Euiin Choi, Context-aware Platform for User Authentication in Cloud Database Computing. 2012

[21] Ahn, H., et al., User Authentication Platform Using Provisioning in Cloud Computing Environment, in Advanced Communication and Networking, 2011, Springer Berlin Heidelberg. pp. 132-138

[22] Ahn, H., et al., User Authentication Using Context-Awareness RBAC Model on Cloud Computing, in Future Communication, Computing, Control and Management, 2012, Springer Berlin Heidelberg. pp. 253-257

[23] Wang, Z., K. Sha, and W. Lv, Slight Homomorphic Signature for Access Controlling in Cloud Computing. Wireless Personal Communications, 2013. 73(1): pp. 51-61

[24] Wang, Z., G. Sun, and D. Chen, A new definition of homomorphic signature for identity management in mobile cloud computing. Journal of Computer and System Sciences, 2014. 80(3): pp. 546-553

[25] Li, H., et al., Identity-Based Authentication for Cloud Computing, in Proceedings of the 1st International Conference on Cloud Computing. 2009, Springer-Verlag: Beijing, China. pp. 157-166

[26] Qin, B., et al., Simultaneous authentication and secrecy in identity-based data upload to cloud. Cluster Computing, 2013. 16(4): pp. 845-859

[27] Mishra, D., V. Kumar, and S. Mukhopadhyay, A Pairing-Free Identity Based Authentication Framework for Cloud Computing, in Network and System Security, 2013, Springer Berlin Heidelberg. pp. 721-727

[28] Acar, T., M. Belenkiy, and A. Küpçü, Single password authentication. Computer Networks, 2013. 57(13): pp. 2597-2614

[29] Dinesha, H.A. and V.K. Agrawal. Multi-level authentication technique for accessing cloud services. International Conference on Computing, Communication and Applications (ICCCA), 2012

[30] Yassin, A., et al., Cloud Authentication Based on Anonymous One-Time Password, in Ubiquitous Information Technologies and Applications, 2013, Springer Netherlands. pp. 423-431

[31] Abdellaoui, A., Y.I. Khamlichi, and H. Chaoui, A Novel Strong Password Generator for Improving Cloud Authentication. Procedia Computer Science, 2016. 85: pp. 293-300

[32] Celesti, A., et al. Three-Phase Cross-Cloud Federation Model: The Cloud SSO Authentication. in Advances in Future Internet (AFIN), 2010 Second International Conference on. 2010

[33]   Chen, Y.-J. and L.-C. Wang, A Security Framework of Group Location-Based Mobile Applications in Cloud Computing. 2011: pp. 184-190

[34]   Noureddine, M. and R. Bashroush, An authentication model towards cloud federation in the enterprise. Journal of Systems and Software, 2013. 86(9): pp. 2269-2275

Modeling of III-V-based Nanohole Filling

Antal Ürmös, Zoltán Farkas, Ákos Nemcsics

Institute of Microelectronics and Technology, Óbuda University, Tavaszmező utca 17, H-1084 Budapest, Hungary; e-mails: urmos.antal@phd.uni-obuda.hu, farkas.zoltan@kvk.uni-obuda.hu, nemcsics.akos@kvk.uni-obuda.hu

Abstract: It is well-known, that several areas of science (for example electronics) have been revolutionized by the application of nano structures. For this reason, it is important to simulate the evolution of these nano structures. The III-V- composite semiconductor based nanohole backfilling is modeled in this paper. The backfilling is described with the viscosity of the liquid gallium, instead of modeling of the real atomic displacements. The construction of the model begins with the macroscopic interpretation of the viscosity. In order to model the atomic displacement at microscopic level, the microscopic viscosity was introduced. It is shown, that under certain conditions we get back the original Arrhenius-Andrande equation from the microscopic viscosity model. During the simulation, both the viscosity and the equilibrium height of the backfilled nanohole was determined as a function of temperature.

Keywords: GaAs; nanohole backfilling; viscosity; modeling

1 Introduction

It is well-known, that several areas of science were revolutionized by the application of nano structures. Recently, increasing proportion of electronic devices containing nanostructures are found.

One excellent example is the III-N-based semiconductor devices, for example the GaN-based blue LEDs. The active regions of these devices consist one or more InGaN quantum wells. These wells are inserted between thicker GaN layers, in a sandwich-like manner. By topologically changing the InN and GaN regions in the InGaN wells, the wavelength of emitted light extends from the ultraviolet to the color of amber. Another possibility is the application of quantum dots, instead of quantum wells. In this case, the large light intensity of the LED is generated by InGaN quantum dots with the aid of phase separation effect [1].

Another example is the nanostructured solar cells. The efficiency of these cells is significantly influenced by the electronic band structure of the semiconductor. A good solution is the application of various nanostructures such as quantum dots and quantum wells [2-4] for increasing of the efficiency of the structure.
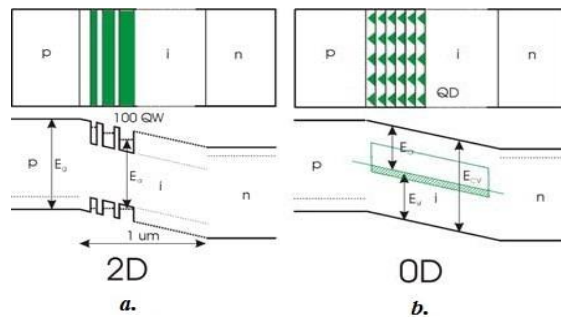
Figure 1
The efficiency of the solar cells can be increased with the application of the multiple quantum wells (a) and quantum dots (b) (source: [3])

With the application of these structures, the solar cells can utilize wider wavelength range of the solar spectrum. Efficiencies of over 40% can be achieved with multilayer solar cells that contain quantum dots. These devices operate on the principle of tunnel effect. In the case of another nano structured solar cells containing quantum wells, up to 40% efficiency can be achieved through the utilization of tunnel effect (Figure 1a). This excellent solar cell efficiency can be surpassed by the usage of quantum dots (Figure 1b). In this case, intermediate energy levels are created in the energy band gap of the semiconductor. This structure is able to utilize the energy of photons, the energy of which differs from the energy determined by gap of the original semiconductor. This way, an efficiency of 60% can be achieved.

An important requisite for the fabrication of the GaAs-based nanostructures, presented in this chapter, is the controlled growth of semiconductor crystal layers and nanostructures. Primarily, these thin, several monolayer thick layers and other nano objects can be grown with the method of molecular-beam-epitaxy (MBE) [5] [6].

2   The Droplet Epitaxy

The various nanostructures can be fabricated with several growth method. There are three growth regimes such as the traditional Frank Van Der Merve, Volmer-Weber and Stransky-Krastanov methods. The difference among these methods, is the difference of the bonding energies between atom-substrate and the atom-neighboring atom relations. In the case of Frank Van Der Merve growth, the atom-substrate bonding energy is bigger than the atom-neighboring atom bonding energy, so mainly two dimensional layers are formed. In case of Volmer-Weber growth the atom-substrate bonding energy is smaller than the atom-neighboring atom bonding energy, so clusters are primarily formed. In case of Stransky-Krastanov growth the adatom-substrate bonding energy and the adatom-

neighboring atom bonding energy are comparable, so both clusters and layers can be formed.
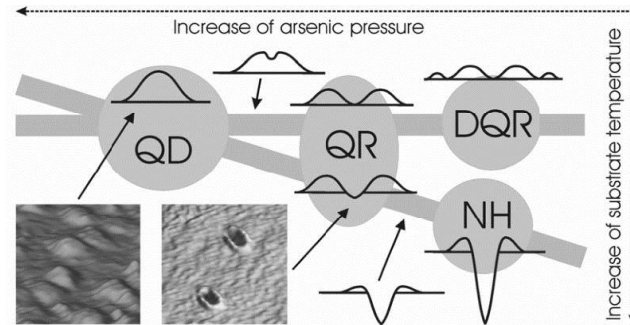
Figure 2

The development of different nanostructures in the function of the temperature of the sample and the pressure of the remaining gas (source: [11])

With the aforementioned methods, only layers or clusters or dots can be formed. A novel technique, called droplet epitaxy allow us to create not only quantum dots but also quantum rings, double quantum rings and nanoholes. The conception of the droplet epitaxy was developed by Koguchi and his co-workers, in the beginning of the 1990 years [7, 8, 9, 10]. The droplet epitaxy is a two stages technology. Firstly a metal droplet of the III. main group of the periodic system (for example Ga) is deposited onto the substrate surface (for example GaAs). After that metallic nano-clusters are created, the final shape which is determined by the sample temperature and by the gas pressure of the residual main group V element (for example arsenic). In this way, various nanostructures can be developed (Figure 2).

As it can be seen on Figure 3, the nanoholes are formed in high temperature and minimal arsenic pressure [12]. The formation of the nanoholes are the following: a gallium droplet is deposited to the substrate surface. Because of the concentration gradient, the arsenic atoms diffuse into the gallium droplet. In the same time, the gallium atoms leave the droplet in the manner of surface diffusion. Finally, the nanohole forms. The process is explained in the Figure 3.
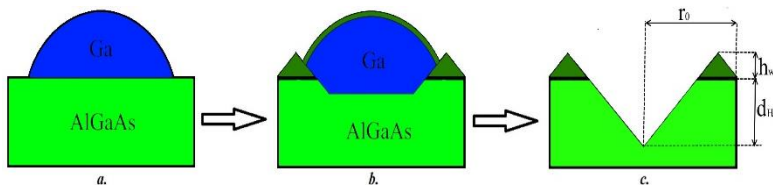
Figure 3

The formation of the nanoholes (source: [12]). The formation of the nanoholes are the following: a gallium droplet are deposited to the substrate (***a***). The arsenic atoms move to the droplet and the same time the gallium atoms leave the droplet (***b***). Finally nanohole forms (***c***)

The variables of Figure 3/c are explained below: r_0 is the outer radius (nanohole and the ring), h_w is the height of the ring, d_H is the depth of the nanohole. If the nanoholes, formed on this way, were filled with metal atoms (for example gallium), then we talk about inverted quantum dots. There are 4 types of inverted quantum dots: type 1 quantum dots are the white light emitters, type 2 quantum dots are the uniform quantum dots, type 3 quantum dots are the ultra low-density quantum dots, type 4 quantum dots are the vertically stacked quantum dot molecules (Figure 4) [12].

In Figure 4, Type 1 quantum dots (white light emitters) are shown which were fabricated by local droplet etching on AlGaAs surface at T=550-620 °C temperature. Bimodal, shallow and deep nanoholes are formed with this method. During the backfilling process of the nanoholes, the shallow ones are fully and the deep ones are partially filled. Local droplet etching creates Type 2 (uniform) forms (Figure 4b) and the density of shallow holes reduced at elevated temperature. These quantum dots are partially buried and their size accurately controlled through the amount of the deposited GaAs. The ultra-low density (ULD, Type 3) quantum dots, filled with GaAs, can be characterized with even lower density and deeper holes (Figure 4c). Figure 4d displays vertically stacked quantum dot molecule, which consists of two closely placed quantum dots. This is the simplest interacting system, made of nanostructures. This structure contains two double filled ultra-low density quantum dots. The two quantum dots are separated with two well-defined AlGaAs barriers.
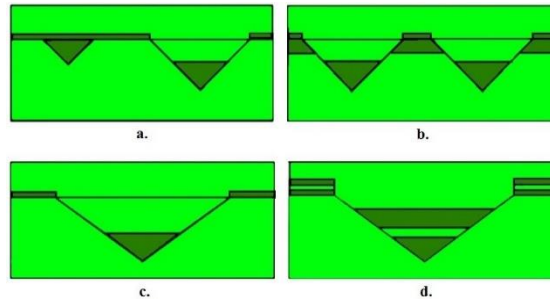
Figure 4

Formation of the inverted quantum dots. The **a)** is the group of Type 1 quantum dots, the **b)** is the group of Type 2 quantum dot, the **c)** is the group of Type3 quantum dots and **d)** is the vertically stacked quantum dot molecule (source: [12]).

3   The Simulation Algorithm

In the simulation, the initial state is the nanohole, showed in the Figure 5 [13]. "**A**" denotes the area adjacent to the ring, "**B**" denotes the ring and "**C**" is the nanohole. The D_{ring} is the width of the ring (radius of outer circle of the ring less

radius of inner circle of the ring), the D_{hole} is the diameter of the hole, the L is the height of the ring above the substrate surface, the H is the height of the ring above the bottom of the nanohole. The angle α is the half angle of the orifice of the nano hole, the value of which value is 55°.
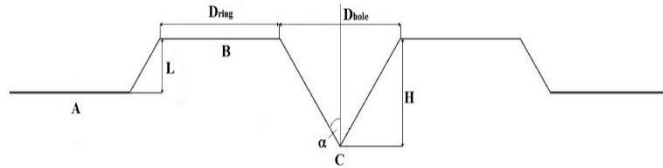
Figure 5

The base structure of the nanohole. The point "**A**" is the area next to the ring, the point "**B**" is the ring and the point "**C**" is the nanohole. A D_{ring} is the width of the ring, a D_{hole} is the diameter of the hole, the L is the height of the ring above the surface of the substrate, the H is the height of the ring above the bottom of the nanohole. The α angle is the half angle of the orifice of the nano hole.

During the simulation, in the first step, Ga was deposited to the substrate surface (Figure 6). In this chapter, the ideal case is investigated, when there is no surface diffusion, meaning that the deposited atoms do not migrate on the surface.
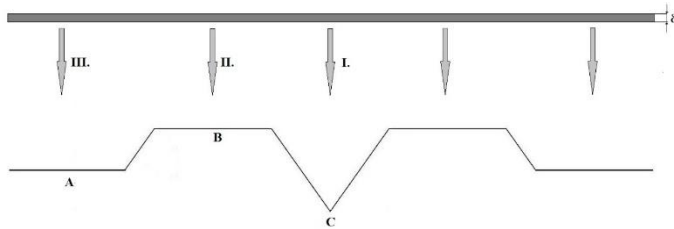
Figure 6

The parts of the volume, that get to different parts of the nanohole, during the deposition (base layer)

From the volume of deposited layer, the volume I goes directly to the nanohole ($V_{I.}$). The volume that is deposited directly to the nanohole can be obtained from the followng equation:

$$V_{I.} = r_{hole}^2 * \pi * \delta \tag{1}$$

where the r_{hole} is the radius of the nanohole and δ is the thickness of the deposited layer. The height of the newly deposited volume in the cone can be computed by the cone volume formula:

$$m = \sqrt[3]{\frac{3*V_{I.}}{\pi*tg^2\alpha}} \tag{2}$$

where the m is the backfilled height from the point of the cone and the $V_{I.}$ is the volume gets directly into the nanohole. The volume II, is the additional ring

volume, the volume *III* is the volume that is deposited aside the ring, δ denotes the thickness of the layer (Figure 7).
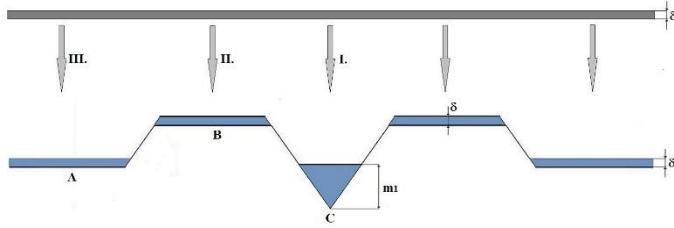
Figure 7

The parts of the volume, that get to different parts of the nanohole, during the deposition (first layer). The heights on the ring and next to the ring is δ.

In the course of the second step another layer is deposited, the thickness of which is also δ (Figure 8). The volumes in the ring and aside ring to the ring are equal, and the monolayer thickness will be also δ. There are two ways to compute the filled height in the nano hole. According to the first method, we compute it with the volume of the cone, which is the following:

$$m_2 = \sqrt[3]{\frac{3*(V_{I.1}+V_{I.2})}{\pi*tg^2\alpha}} \tag{3}$$

where m_2 is the height backfilled in the second step, a $V_{I.1}$ is the backfilled volume in the first step and $V_{I.2}$ is the backfilled volume in the second step. This formula can be generalized, in the following way:

$$m_i = \sqrt[3]{\frac{3*\sum_{k=1}^{i} V_{I.k}}{\pi*tg^2\alpha}} \tag{4}$$

where m_i is the backfilled height in the *i*-th step and $\sum_{k=0}^{i} V_{I.k}$ is the the sum of the volume deposited in *i*-th step and the previous steps.
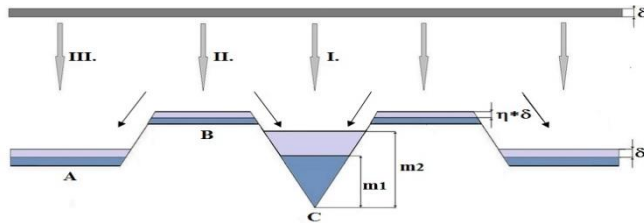
Figure 8

The parts of the volume, that get to different parts of the nanohole, during the deposition (second layer). The heights on the ring and next to the ring is δ.

Another possible solution is, that we can compute the height of the truncated cone, solving a third order equation. This equation is as follows:

$$m_k^3 * \left(\frac{\pi}{3} * tg^2 55\right) + m_k^2 * (\pi * r_{k-1} * tg55) + m_k * (\pi * r_{k-1}^2) - V_k = 0 \qquad (5)$$

where m_k is the height of the truncated cone, r_{k-1} is the radius of the circle of the top of the deposited material within the cone, computed in the previous step, the V_k is the volume of the truncated cone, which is equal to the volume $V_{I.k}$, which gets directly to the nanohole in the k-th step. In the next step, this height is added to the total sum of the backfilled heights, computed in the previous steps:

$$m_i = \sum_{k=1}^{i} m_k. \qquad (6)$$

The ring volume $V_{II.}$ can be computed, that the volume of the inner truncated cone (V_{itc}) is subtracted from the total volume (which is the sum of the volume of the inner truncated cone (V_{itc}) and the ring volume $V_{II.}$). The V_{itc} volume of the inner truncated cone can be calculated by the following formula:

$$V_{itc} = \frac{\pi}{3} m_{itc}(R_{itc}^2 + r_{itc}^2 + R_{itc} * r_{itc}) \qquad (7)$$

where R_{itc} is the larger radius, the r_{itc} is the smaller radius and m_{itc} is the height of the inner truncated cone. The r_{itc} is the smaller radius can be computed by the following formulae:

$$r_{itc} = H * tg\alpha \qquad (8)$$

where H is the depth of nanohole, the α is the half angle of the orifice of the nanohole (its value is 55°). The R_{itc} is the larger radius can be computed by

$$R_{itc} = H * tg\alpha + m_{itc} * tg\alpha \qquad (9)$$

formula, where H is the depth of nanohole, the m_{itc} is the height of the inner truncated cone and the α is the half angle of the orifice of the nanohole (its value is 55°). As a consequence, the volume of the inner truncated cone:

$$V_{itc} = m_{itc}^3 * \left(\frac{\pi}{3} * tg^2\alpha\right) + m_{itc}^2 * (H * \pi * tg^2\alpha) + m_{itc} * (H^2 * \pi * tg^2\alpha). \quad (10)$$

Figure 9 illustrates the parameters. R_{itc}, r_{itc}, m_{itc} are the larger radius, the smaller radius and the height of the inner truncated cone respectively. The α is the half angle of the orifice of the nanohole (its value is 55°), H is the depth of nanohole and the y an auxiliary variable, the value of which is $y = m_{itc} * tg\alpha$.
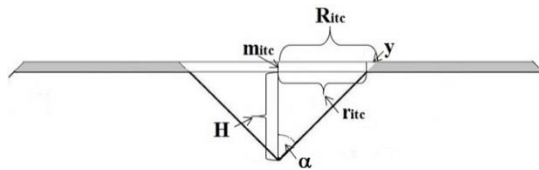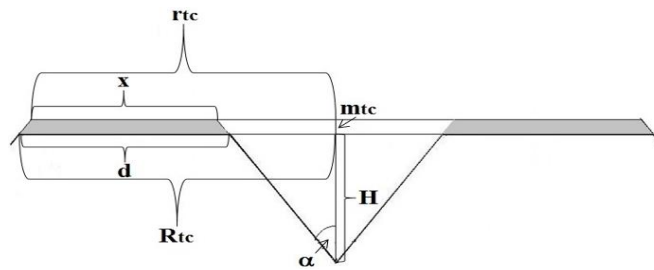

Figure 9

The interpretation of the parameters. In the figure R_{itc} is the larger radius, the r_{itc} is the smaller radius and m_{itc} is the height of the inner truncated cone, the α is the half angle of the orifice of the nanohole (its value is 55°), H is the depth of the nanohole and the y an auxiliary variable, the value of which is $y = m_{itc} * tg\alpha$.

The total volume (V_{tc}) can be computed by the following formula:

$$V_{tc} = \frac{\pi}{3} m_{tc}(R_{tc}^2 + r_{tc}^2 + R_{tc} * r_{tc}) \tag{11}$$

where R_{tc} is the greater radius, r_{tc} smaller radius, az m_{tc} the height of the total truncated cone. The r_{tc} smaller radius can be computed by the following formula

$$r_{tc} = H * tg\alpha + d - m_{tc} * tg\alpha \tag{12}$$

where H is the depth of the nanohole, m_{tc} the height of the truncated cone and α is the half angle of the orifice of nanohole (its value is 55°). The R_{tc} bigger radius can be computed by the following formula:

$$R_{tc} = H * tg\alpha + m_{tc} * tg\alpha \tag{13}$$

where H is the depth of the nanohole, m_{tc} the height of the complete truncated cone (in this case it is the thickness of the deposited layer) and α is the half angle of the orifice of the nanohole (the value of which is 55°). As a consequence, the volume of the truncated cone:

$$V_{tc} = m_{tc}^3 * \left(\frac{\pi}{3} * tg^2\alpha\right) - m_{tc}^2 * (d * \pi * tg\alpha - H * \pi * tg^2\alpha) + m_{tc} * (d^2 * \pi + 2 * d * H * \pi * tg\alpha + H^2 * \pi * tg^2\alpha) \tag{14}$$

where H is the depth of the nanohole, m_{tc} the height of the truncated cone (in this case is the thickness of the deposited layer) and α is the half angle of the orifice of the nanohole (its value is 55°).

Figure 10 illustrates the parameters. R_{tc}, r_{tc}, m_{tc} are the greater radius, the smaller radius and the height of the inner truncated cone, respectively. The α is the half angle of the orifice of the nanohole (its value is 55°), H is the depth of the nanohole, the d is the lower plateau of the ring and the x is the upper plateau of the ring.

Figure 10

The interpretation of the parameters. In the figure R_{tc} is the larger radius, the r_{tc} is the smaller radius and m_{tc} is the height of the inner truncated cone, the α is the half angle of the nanohole (its value is 55°), H is the depth of the nanohole, the d is the lower plateau of the ring and the x is the upper plateau of the ring.

Taking into consideration the parameters defined above, the $V_{II.}$ ring volume can be easily described by the following formula:

$$V_{II.} = V_{tc} - V_{itc} \tag{15}$$

where V_{tc} is the volume of the complete truncated cone, a V_{itc} is the volume of the inner truncated cone. This substraction takes the form (with substituting the respective/appropriate parameters):

$$V_{II.} = -m_{ring}^2 * (d * \pi * tg\alpha + 2 * H * \pi * tg^2\alpha) + m_{ring} * (d^2 * \pi + 2 * d * H * \pi * tg\alpha) \tag{16}$$

where H is the depth of the nanohole, m_{ring} the height of the ring (in this case is the thickness of the deposited layer) and α is the half angle of the orifice of the nanohole (its value is 55°), the d is the width of the lower plateau of the ring.

4   The Mechanism of the Atomic Displacement

In the reality, at finite temperature $\left(n * k * T > E_{bonding}\right)$ the atoms migrate on the surface. The parameters of migration depend on the temperature. This surface migration can be modeled by many methods, for example Kinetic Monte-Carlo method [14, 15]. In this paper, instead of this algorithm the dynamic viscosity of liquid gallium will be applied. This is an approximation, and this attribute originates from the macroscopic description. The viscosity is the function of the temperature. From macroscopic point of view, the dynamical viscosity is a proportionality factor, depends on the properties of the liquid material. Dynamic viscosity is the ratio of shear stress to shear velocity:

$$\tau = \eta * \frac{d\gamma}{dt} \tag{17}$$

where τ is the shear stress, η is the dynamic viscosity, $\frac{d\gamma}{dt}$ is the shear velocity. In case of liquid metals, the viscosity can be determined in a multiple ways [16-20]. In this paper, the Arrhenius-Andrande equation is used [20], its formula is the following:

$$\mu(T) = \mu_0 * e^{\frac{E}{R*T}}. \tag{18}$$

In this formula, the $\mu(T)$ is the dynamic viscosity, the μ_0 is a pre-exponential factor (in case of gallium its value is 0.436), E is the activation energy (its value in case of Ga is 4000 $\frac{J}{mol}$), the R is universal gas constant (its value is 8,3144 $\frac{J}{K*mol}$) and the T is the temperature (on Kelvin scale).

In the microscopic approach the probability of the occurrence of the event is given at atomic level in order to determine the dynamic viscosity [15, 21]:

$$P(E_a, T) = e^{-\frac{E_a}{R*T}} \tag{19}$$

where E_a is the activation energy of the given event, T is the temperature (on Kelvin scale), R is the universal gas constant (its value is $8,3144 \frac{J}{mol*K}$). The k hopping frequency is calculated by this formula:

$$k = k_0 * e^{-\frac{E_a}{R*T}} \tag{20}$$

where k_0 is the atomic vibration frequency. Parameter k_0 can be determined in the subsequent way:

$$k_0 = \frac{2*k_B*T}{h} \tag{21}$$

where k_B is the Boltzmann constant, h is the Planck constant. After this step the $\rho = \rho(E_a, T)$ geometrical factor is introduced, in the following way [26]:

$$\rho(E_a, T) = l * e^{\frac{2E_a}{R*T}} \tag{22}$$

where l is a scalar value, but it can be both a single variable or a multivariable function. Let $E_0 = E_a[n]$, where E_0 is the bulk activation energy, $E_a[n]$ an density functional, which gives the binding energy of the atomic multeity, consists of n binding energy, with taking into consideration of atom-atom, atom-electron, electron-electron interactions. Consequently the dynamic viscosity is:

$$\mu(T) = \rho(E_0, T) * k = k_0 * l * e^{\frac{E_0}{R*T}} \tag{23}$$

where the $k_0 * l$ product is the μ_0 pre-exponential factor. Thus the equation 18 is obtained. The effect of the viscosity is taken into consideration according the subsequent way in this paper: the volume of the ring is multiplied by an $\eta(T)$ volume proportional factor, which can be computed with the next formula:

$$\eta(T) = \frac{\mu(T)}{\mu(T_m)} = e^{\frac{E}{R}\left(\frac{T_m-T}{T*T_m}\right)} \tag{24}$$

where T is the temperature (on Kelvin scale), the $\eta(T) = \eta$ is a volume proportional factor, $\mu(T)$ is the temperature dependent dynamical viscosity, $\mu(T_m)$ is the melting temperature dependent dynamical viscosity. According to the formulae above, the $V_{II.}$ ring volume can be computed in the subsequent way:

$$V'_{II.} = \eta * V_{II.} \tag{25}$$

where $V_{II.}$ is the volume of the ring, $V'_{II.}$ volume that remains on the ring and η is the viscosity. The thickness of the layer remaining on the ring is proportional with the viscosity, thus $m_{ring} = \eta * \delta$. As a consequence, if this formula was substituted to the 11 equation, then:

$$V'_{II.} = -(\eta * \delta)^2 * (d * \pi * tg\alpha + 2 * H * \pi * tg^2\alpha) + (\eta * \delta) * (d^2 * \pi + 2 * d * H * \pi * tg\alpha) \tag{26}$$

where η is the volume proportionality factor, the δ is the thickness of the deposited layer, H is the depth of the nanohole, the α is the half angle of the orifice of the nanohole (its value is 55°), the d is the width of the lower plateau of the ring. As a consequence the volume that moves into the nano hole is a sum of the volume moving directly to the nanohole and that moving indirectly into the nanohole, i.e.:

$$V'_{I.} = V_{I.} + V''_{II.} \tag{27}$$

where $V_{I.}$ is the volume, moving directly to the nanohole and $V''_{II.}$ is the volume moving indirectly to the nanohole. The latter originates from the ring. This volume can be computed by the following formula:

$$V''_{II.} = \frac{(1-\eta)}{2} * V_{II.} \tag{28}$$

where volume of $V''_{II.}$ that gets indirectly from the ring to the hole, and $V_{II.}$ is the volume remaining to the ring. The reason of ½ factor is, that the half of the volume that leaves the ring flows to the nanohole, and the other half of this volume flows to the area adjacent to the ring.

Similarly as in the previous chapter, a δ thick layer is deposited also (Figure 11). The arrows show, that $\frac{(1-\eta)}{2}$ percent of the ring volume flows down into the hole and aside the ring. The height of the layer on the ring will be $\eta * \delta$. The height of the volume in the nanohole can be calculated by the volume formula of the cone:

$$m = \sqrt[3]{\frac{3*V'_{I.}}{\pi*tg^2\alpha}} \tag{29}$$

where m is the backfilled height, $V'_{I.}$ is the volume that arrives to the nanohole, which can be computed by the formula 4 and 5.
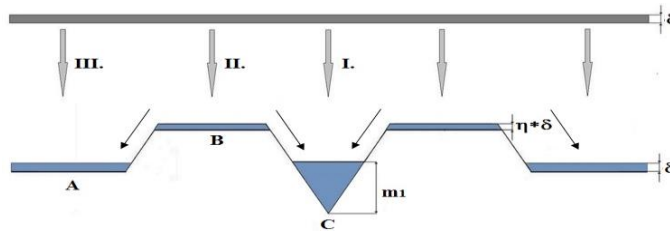
Figure 11

The parts of the volume, move to different parts of the nanohole, during the deposition (first layer). The arrows shows, that $\frac{(1-\eta)}{2}$ percent of the ring volume flows down to the hole and aside the ring. The height of the layer on the ring will be $\eta * \delta$ and aside the ring is δ.

In the second step also a δ thick layer is deposited (Figure 12). The arrows (similarly as is the previous figure) shows, that $\frac{(1-\eta)}{2}$ percent of the ring volume flows down into the hole and aside the ring. The height of the layer on the ring

will be $\eta * \delta$ and aside the ring will be δ. The backfilled height in the nanohole, - similarly to the previous chapter - can be computed in two different ways. According to the first method, using the volume formula of the cone, the formula of the backfilled height is

$$m_2 = \sqrt[3]{\frac{3*(V'_{I.1}+V'_{I.2})}{\pi*tg^2\alpha}} \tag{30}$$

where m_2 is the backfilled height in the second step, the $V'_{I.1}$ is the backfilled volume in the first step and the $V'_{I.2}$ is the backfilled volume in the second step. This formula can be generalized according to the subsequent way:

$$m_i = \sqrt[3]{\frac{3*\sum_{k=1}^{i} V'_{I.k}}{\pi*tg^2\alpha}} \tag{31}$$

where m_i is the filled height in the i-th step and $\sum_{k=0}^{i} V'_{I.k}$ is the the sum of the volumes deposited in the i-th and in all of the previous steps. Another possible solution is that with equation 5 a third-order equation is solved and the height of the truncated cone is obtained this way. Afterwards the height is added to the sum of the heights which is in already in the nanohole (equation 6).
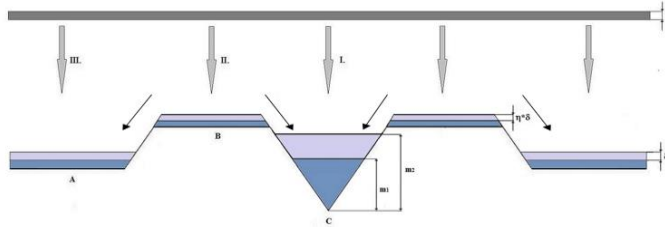
Figure 12

The parts of the volume, goes to different parts of the nanohole, during the deposition (first layer). The arrows shows, that $\frac{(1-\eta)}{2}$ percent of the ring volume flows down to the hole and aside the hole. The height of the layer on the ring will be $\eta * \delta$ and next to the ring is δ.

5   Discussion

During the simulation it was investigated as to how the nanohole was filled with GaAs after crystallization. (It means, that the atomic movement described with the help of viscosity of metallic Ga, but the thicknesses calculated from the crystallized GaAs.) The backfilling process (Ga movement) was simulated as the function of the temperature. The height of the ring next to the nanohole increases as the layers are built upon each other. In other words, it was investigated that when the heights of the material inside of the hole and that on the ring became approximately equal. The time was determined as the number of layers already deposited. These two heights considered to be equal, when the differences of the

top levels of the appropriate structures are closer to each other than unit. The average value of these two heights is called equilibrium height. During the simulation it was our goal to determine, the temperature, where for a given number of layers the height difference between the hole and the ring was minimal. Figure 13 shows the backfilling process of the nanohole. The number of layers already deposited were 1 (denoted with a), 3 (denoted with b), 5 (c) and 7 (d), respectively. If the temperature dependence of the viscosity is neglected, then equilibrium height can be observed at the 22-nd layer. In our example each of the layers contain 2 monolayer thick GaAs deposition. It is important, that in case of GaAs the lattice constant is 0.526 nm, which corresponds 2 monolayers (because this material consists of two components). Here, one monolayer corresponds to 10 pixels, (which is $\frac{0,284}{10} = 0.0284$ nm) in the software, used for the simulation.
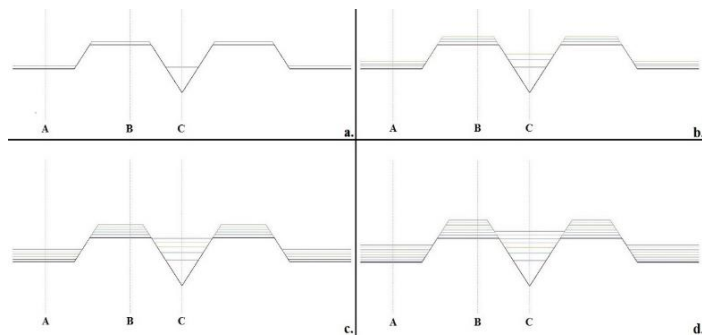
Figure 13

The filling process of the nanohole, in case of 1 (**a**), 3 (**b**), 5 (**c**) and 7 (**d**) layers

The flow-chart of the simulation algorithm is shown in the Figure 14.

This ideal case exists below 49 °C substrate temperature. Over this temperature, the equilibrium height and number of layers corresponding to equilibrium height (hereinafter equilibrium number of layers) decrease. As Fig. 15 shows, at 51.6 °C substrate temperature the nanohole totally filled with the 20-th layer. At this point the equilibrium height is 15.2945 nm. The bottom line of the charts below represents the filled height in the nanohole (measurement point "C"), while the top line (measurement point "B") represent the ring height. The point, where these two lines cross each other is the place of the equilibrium height. The absolute error value is maximal at the layer 18, its value is 0.0078 (the relative error is 1.561*10^-3%). On each parts of chart, the equilibrium height and equilibrium layer pairs are marked with a red lines. It can also be seen that a unit change of the equilibrium height corresponds to increasing change of temperature as the temperature itself rises. Point "A" is selected as reference point (denoting 0 height).
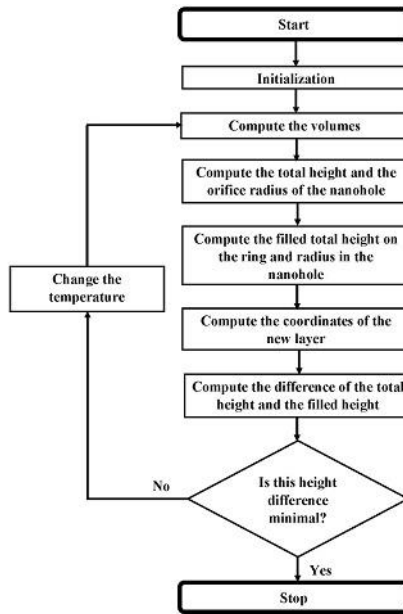
Figure 14
The flowchart of the simulation algorithm
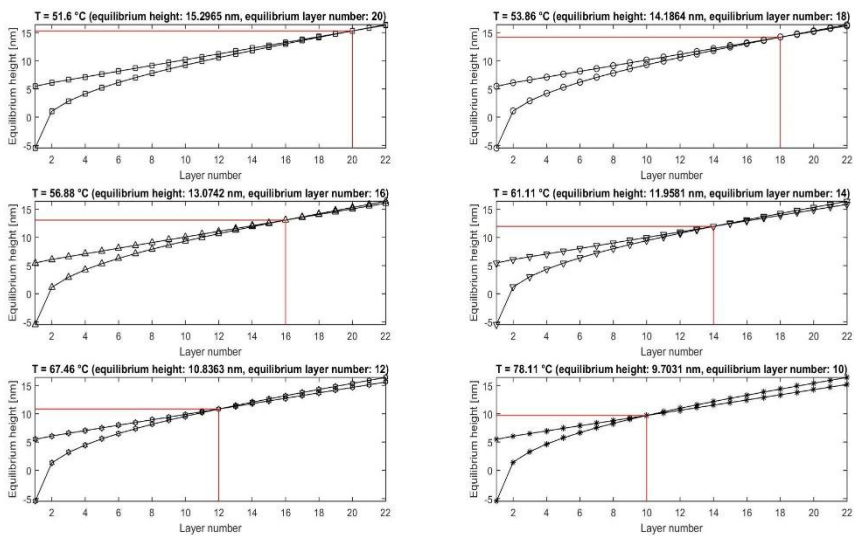
Figure 15
Layer number and equilibrium height diagrams in different temperatures

On Figure 16 the top diagram shows the equilibrium height as the function of the temperature and the bottom ones shows the equilibrium height as the function of the viscosity.
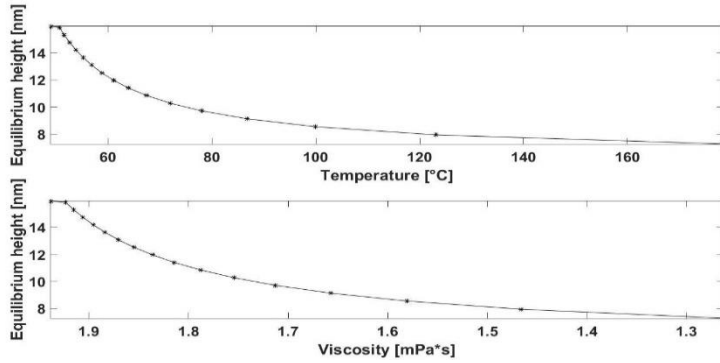
Figure 16

Equilibrium height-temperature (top) and equilibrium height-viscosity (bottom) diagrams

Here, the macroscopic viscosity of Ga is calculated by Arrhenius-Andrande formula (18 equation). Figure 17 shows the macroscopic temperature-viscosity diagram of Ga.
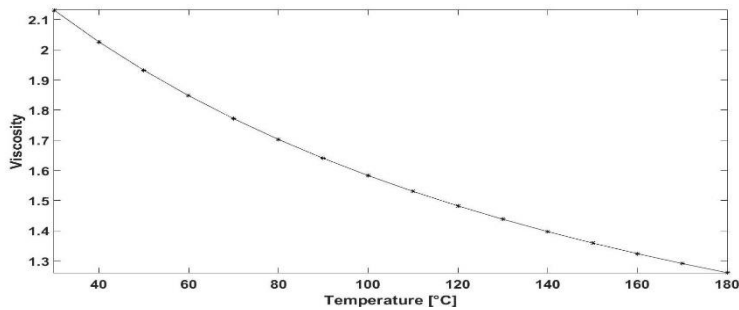
Figure 17

The macroscopic temperature-viscosity diagram of gallium

For the purpose to determine the microscopic temperature-viscosity diagram, first the maximum value of the microscopic viscosity is calculated. For the determination of this, the starting point was the equation 18. At first, the activation energy of the particles is determined, using the following formula [22]:

$$E_p = E_0 * \left(1 - 6\alpha * \frac{r}{D}\right) \tag{32}$$

where the E_0 the bulk activation energy (its value is in case of gallium $4000 \frac{J}{mol}$), E_p is the cohesive energy of the metal droplets, the α shape factor (in case of sphere shaped particles its value is 1), the r is the radius of the atom, a D is the

size of particle. Let the melting point of the gallium droplets be 254 K (-19.15 ^{o}C) [23]. The size of the particle is determined by the following formula [24]:

$$D = \frac{9*r*T_{mb}}{T_{mb}-T_{mp}} \tag{33}$$

From this equation, in case of Ga, the atomic radius is 1.36*10^{-10} m and the size of the particle (at 254 K temperature) is 7.5688*10^{-9} m. If the activation energy is $E_p = 3568 \frac{J}{mol}$ and the particle melting point is $T_{mp} = 254K$, then the maximum value of the microscopic viscosity is 2.3626.

In microscopic level, the temperature-viscosity value pairs of Ga showed in Figure 18 and 19. If one-term power-law is fitted on this value pair series, then the maximum relative error of the fitting is 1.798%, while in case of one-term exponential function the maximum relative error of the fitting is 3.499%. Despite this two-term power and exponential functions are fitting, because the maximum values of the relative error are under 1%. If two-term power law is fitted to the point series (Figure 18), then the temperature-viscosity formula will be the following [25]:

$$\mu_{p,Mi}(T) = a * x^b + c = a * T^b + c \tag{34}$$

where $\mu_{p,Mi}(T)$ is the viscosity as a function of teperature, the a, b, c values are material dependent empirical constants, the $x = T$. Here, the temperature is given in Kelvin. The value of a is 7.766*10^5, b is -2.292 and c is -0.7616. The upper part of the figure is the power function, fitted to the point series, the bottom graph of the figure shows the error values of the function fitting in the individual points. The maximum value of the fitting error is 0,0008515 (the relative error is 0.0523%), which is at 123.15 ^{o}C (396.3 ^{o}K) temperature.
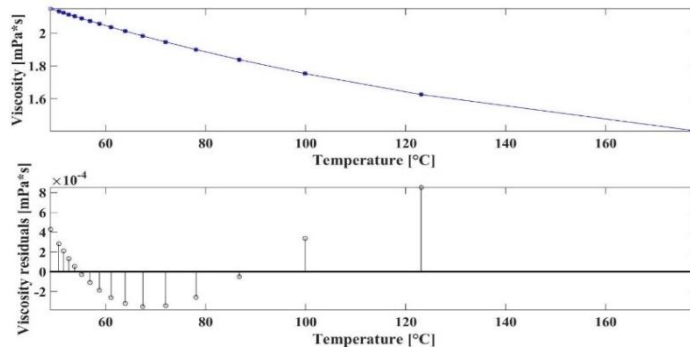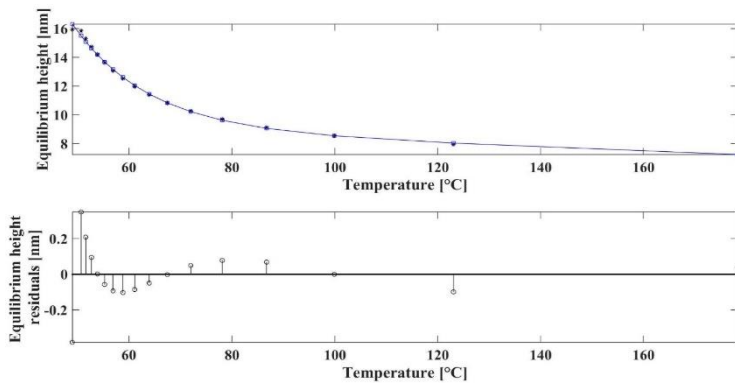

Figure 18

The temperature-viscosity diagram of gallium in microscopic level. Power function is fitted to the point series. The top part of the diagram is the power function, fitted to the point series, the bottom part of the diagram shows the residuals of the function fitting in the individual points.

If the a coefficient is $k_0 * l = \tau_0$, the b coefficient is 1, the c coefficient is 0 and the $x = e^{\frac{E_0}{R*T}}$, the the equation 18 is obtained. From this we can conclude, that the equation 35 is the microscopic form of equation 18. If exponential sum function was fitted to the point series (Figure 19), then the temperature-viscosity formula would be as follows:

$$\mu_{e,Mi}(T) = a * e^{bx} + c * e^{dx} = a * e^{bT} + c * e^{dT} \tag{35}$$

where $\mu_{e,Mi}(T)$ the viscosity as a function of teperature, the a, b, c values are material dependent empirical constants, the $x = T$. The value of a parameter is 41.77, b parameter is -0.01341, c parameter is 1.382 and d parameter is -0.001554. The upper part of the diagram is the exponential function fitted to the point series, the bottom graph of the diagram shows the error values of the function fitted to the individual points. The maximum value of the fitting error is 0.0001338 (the relative error is 0.00822%), which is at 123.1 ^{o}C (396.3 ^{o}K) temperature.
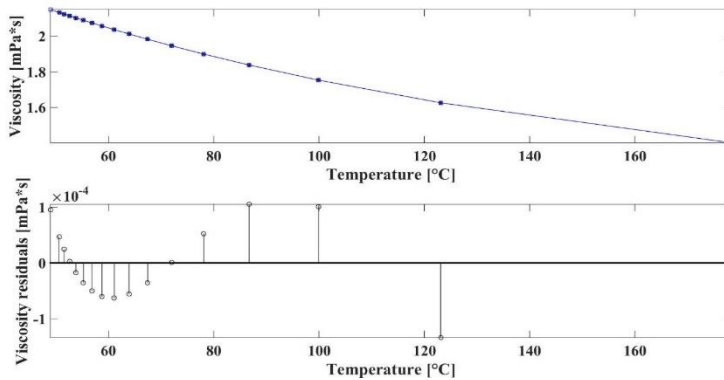
Figure 19

The temperature-viscosity diagram of gallium in microscopic level. Exponential function is fitted to the point series. The top part of the diagram is the power function, fitted to the point series, the bottom part of the diagram shows the residuals of the function fitting in the individual points.

If the a coefficient is $k_0 * l = \tau_0$, the b coefficient is 1, the c coefficient is 0 and the $x = \frac{E_0}{R*T}$, the the equation 18 is obtained. From this we can conclude, that the equation 5.5 is the microscopic form of equation 18.

The analytical determination of equilibrium height, power function and exponential function are fitted to the temperature-equilibrium height point series. If one-term power function is fitted to this point pair series, then the maximum relative error of the fitting is 36.31%, while in case of one-term exponential function the maximum relative error of the fitting is 39.81%. Two-term power function does not fit at all, if Kelvin scale is used. In spite of this, exponential sum

function can be fitted well (Figure 20), so the temperature-equilibrium height formula will be the following:

$$h_e(T) = a * e^{bx} + c * e^{dx} = a * e^{b*T} + c * e^{d*T} \tag{36}$$

where $h_e(T)$ the equilibrium height in the function of temperature, the a, b, c, d values are material dependent empirical constants, the $x = T$. The value of a parameter is $3.278*10^{10}$, b parameter is -0.06904, c parameter is 16.2 and d parameter is -0.001783. The top part of the diagram is the two-terms exponential function, fitted to the point series, the bottom part of the diagram shows the value of error of the fitting at individual points. The maximum value of the fitting error is 0.3822 (the relative error is 2.39 %), which is at 48.95 °C (322.1 °K).

Figure 20

The temperature-equilibrium height diagram of gallium in microscopic level. Exponential function is fitted to the point series. The top part of the diagram is the power function, fitted to the point series, the bottom part of the diagram shows the residuals of the function fitting in the individual points.

Conclusion

The nanohole backfilling was modeled on a III-V substrate, in this paper. The backfilling was described with the viscosity of liquid gallium, instead of modeling of the real atomic displacements. The crystallization by arsenic was considered as a quick bulk process. During the modeling, we started with the macroscopic interpretation of the viscosity. Microscopic viscosity was introduced in order to enable he presentation of a statistical, microscopic description of the atomic displacement. This was shown for the cases when the original Arrhenius-Andrande equation that the microscopic viscosity is returned. During the simulation, the viscosity and the equilibrium height of the filled nanohole was determined as a function of temperature.

References

[1]   G. E. Weng, A. K. Ling, X. Q. Lv, J. Y. Zhang, B. P. Zhang, „III-Nitride-Based Quantum Dots and their Optoelectronic Applications," Nano-Micro Lett. 3 (3) pp. 200-207, 2011

[2]　Á. Nemcsics, "Novel Frontiers in Solar Cell Technology", In Hungarian, Elektronikai Technológia, Mikrotechnika, pp. 39-45, 2009

[3]　Á. Nemcsics, „Novel Alternative for the GaAs-based Self-organized Quantum-Structure," Óbuda University e-Bulletin, pp. 193-199, 2011

[4]　A. Luque, A. Martí, C. Stanley, "Understanding Intermediate-Band Solar Cells" Nature Photonics 6(3):146-152

[5]　Á. Nemcsics, I. Réti, V. G. Tényi, P. Kucsera, L. Tóth, P. Harmat, A. Mieville, M. Csutorás, B. Kupás-Deák, T. Sándor, J. Bozsik, "Technological Conditions of the Production Nanostructures by Molecular Beam Epitaxy", In Hungarian, GÉP LXI, p. 29, 2010

[6]　M. A. Herman, H. Sitter, Molecular Beam Epitaxy, Fundamentals and Current Status 2nd ed., Berlin: Springer, 1996

[7]　N. Koguchi, S. Takahashi, T. Chikyow, „New MBE Growth Method for InSb Quantum Well Boxes," Journal of Crystal Growth, pp. 688-692, 1991

[8]　N. Koguchi, K. Ishige, „Growth of GaAs Epitaxial Microcrystals on an S-Terminated GaAs Substrate by Successive Irradiation of Ga and As Molecular Beams," Japanese Journal of Applied Physics, Vol. 32, Issue 5A, p. 2052, 1993

[9]　T. Chikyow, N. Koguchi, „MBE Growth Method for Pyramid-shaped GaAs Micro Crystals on ZnSe(001) Surface Using Ga Droplets," Japanese Journal of Applied Physics, Vol. 29/2, Issue 11, p. 2093, 1990

[10]　S. Sanguinetti, N. Koguchi, "Droplet Epitaxy of Nanostructures," in Molecular Beam Epitaxy: From Research to Mass Production, 1st Edition, Waltham, MA, USA, Elsevier Science, 2013, pp. 95-111

[11]　Á. Nemcsics, „Growth of GaAs and Related Materials-based Nano-Structures by Droplet Epitaxial Method," in Recent Res. Devel. Mat. Sci. Eng., Vol. 4, Kerala, India, Transworld Research Network, 2011, pp. 53-87

[12]　C. Heyn, D. Sonnenberg, W. Hansen, „Local Droplet Etching: Self-assembled Nanoholes for Quantum Dots and Nanopillars," in Nanodroplets, New York, Springer Science, 2013, pp. 363-383

[13]　Á. Nemcsics, B. Pődör, L. Tóth, J. Balázs, L. Dobos, J. Makai, M. Csutorás, A. Ürmös, „Investigation of MBE Grown Inverted GaAs Quantum Dots," Microelectronics Reliability, Vol. 59, pp. 60-63, 2016

[14]　M. Biehl, "Lattice Gas Models and Kinetic Monte Carlo Simulations of Epitaxial Growth," in Birkhäuser Basel, ISNM International Series of Numerical Mathematics, 2004

[15]　W. Miller, "Simulation of Epitaxial Growth by Means of Density Functional Theory, Kinetic Monte Carlo, and Phase Field Methods," in Handbook of Crystal Growth: Fundamentals (Volume I, Part A:

Thermodynamics and Kinetics), 2nd ed., New York, Elsevier B. V, 2015, pp. 521-559

[16] L. Battezzati, A. L. Greer, „The Viscosity of Liquid Metals and Alloys," Acta Metallurgica, Volume 37, Issue 7, pp. 1791-1802, 1989

[17] C. J. Seeton, „Viscosity–Temperature Correlation for Liquids," Tribology Letters, Vol. 22, Issue 1, pp. 67-78, 2006

[18] K. E. Spells, „The Determination of the Viscosity of Liquid Gallium over an Extended Nrange of Temperature," Proc. Phys. Soc, 1936

[19] M. J. Assael, I. J. Armyra, J. Brillo, S. V. Stankus, J. Wu,W. A. Wakeham, „Reference Data for the Density and Viscosity of Liquid Cadmium, Cobalt, Gallium, Indium, Mercury, Silicon, Thallium, and Zinc," J. Phys. Chem. Ref. Data, Vol. 41, Issue 3, pp. 033101-1-17, 2012

[20] W. F. Gale, T. C. Totemeier, „General Physical Properties," in Smithells Metals Reference Book (Eighth Edition), Elsevier Inc., 2004, pp. 1-45

[21] L. Nurminen, A. Kuronen, K. Kaski, "Kinetic Monte Carlo Simulation of Nucleation on Patterned Substrates," Phys. Rev. B, Vol. 63, No. 3, p. 035407, December 2000

[22] W. H. Qi, M. P. Wang, "Size and Shape Dependent Melting Temperature of Metallic Nanoparticles," Materials Chemistry and Physics, Vol. 88, No. 2-3, pp. 280-284, 2004

[23] G. A. Mansoori, S. A. Rice, "Confined Fluids: Structure, Properties and Phase Behavior," in Advances in Chemical Physics, Vol. 156. Hoboken, NJ, USA: John Wiley & Sons Inc., 2014, pp. 1-97

[24] D. Xie, M. P. Wang, W. H. Qia, L. F. Cao, "Thermal Stability of Indium Nanocrystals: A Theoretical Study," Materials Chemistry and Physics, Vol. 96, No. 2-3, pp. 418-421, 2006

[25] I. N. Bronshtein, K. A. Semendyayev, G. Musiol, H. Mühlig, „Arithmetic," in Handbook of Mathematics, Berlin Heidelberg, Springer, 2007, p. 8

[26] L. C. Yang, "Frequency Factor in Arrhenius Decomposition Kinetics for Insensitive Energetic Materials," in American Institute of Aeronautics and Astronautics Inc., 45th AIAA/ASME/SAE/ASEE Joint Propulsion Conference & Exhibit, 2009, pp. 1-28

[27] Z. Erdélyi, M. Pasichnyy, V. Bezpalchuk, J. J. Tomán, B. Gajdics, A. M. Gusak, Stochastic Kinetic Mean Field Model, Computer Physics Communications, Volume 204, July 2016, pp. 31-37, http://dx.doi.org/10.1016/j.cpc.2016.03.003

[28] J. Zheng, W. Zhuang, N. Yan, G. Kou, H. Peng, C. McNally, D. Erichsen, A. Cheloha, S. Herek, C. Shi, Y. Shi, Classification of HIV-I-mediated

Neuronal Dendritic and Synaptic Damage using Multiple Criteria Linear Programming, Neuroinformatics, Vol. 2, No. 3, pp. 303-326, 2004

[29]   R. E. Precup, S. Preitl, M. Balas, V. Balas, Fuzzy Controllers for Tire Slip Control in Anti-Lock Braking Systems, Proceedings of IEEE International Conference on Fuzzy Systems, Budapest, Hungary, Vol. 3, pp. 1317-1322, 2004

[30]   S. B. Ghosn, F. Drouby, H. M. Harmanani, A Parallel Genetic Algorithm for the Open-Shop Scheduling Problem using Deterministic and Random Moves, International Journal of Artificial Intelligence, Vol. 14, No. 1, pp. 130-144, 2016

[31]   J. K. Tar, J. F. Bitó, I. J. Rudas, Contradiction Resolution in the Adaptive Control of Underactuated Mechanical Systems Evading the Framework of Optimal Controllers, Acta Polytechnica Hungarica, Vol. 13, No. 1, pp. 97-121, 2016

Analysis of Conveyor Belt Impact Resistance Data Using a Software Application

Anna Grinčová[1], Miriam Andrejiová[2], Daniela Marasová[3]

[1]Faculty of Electrical Engineering and Informatics, Technical University of Kosice, Letna 9, 042 00 Kosice, Slovak Republic

[2]Faculty of Mechanical Engineering, Technical University of Kosice, Letna 9, 042 00 Kosice, Slovak Republic

[3]Faculty of Mining, Ecology, Process Control and Geotechnology, Technical University of Kosice, Park Komenskeho 14, 042 00 Kosice, Slovak Republic

anna.grincova@tuke.sk, miriam.andrejiova@tuke.sk, daniela.marasova@tuke.sk

Abstract: Transportation is an important logistics activity within production systems, affecting continuity of the production process. High-capacity, cost-efficient, and energetically undemanding, continuous transportation systems include the conveyor belt system. A crucial and the most sensitive component thereof is a conveyor belt which is exposed, within the production processes, to the impacts that give rise to the damage and degradation process. Therefore, the issues related to increasing the service life and the operational reliability of conveyor belts are highly topical. The aim of the article is to discuss the issues regarding the measurement analysis within the testing of conveyor belts in terms of software optimization. The role of the created AnCoBE (Analysis of Conveyor Belt) software application is to apply selected mathematical and statistical methods and Java technology in order to support processing and evaluation of data obtained by experimental research of evaluation of utility properties of conveyor belts in terms of their dynamic stress. The experimental tests herein were focused on the determination of conveyor belt impact resistance and the description of damage caused to conveyor belts by the related stresses.

Keywords: conveyor belts; regression model; software application

1 Introduction

The belt conveyor systems are used in a broad extent for the transportation of various types of materials and are exposed to various surrounding conditions, weather, and environmental impacts. According to Kulinovski [1], for belt conveyors, the transport task can be defined as a process the purpose of which is

to transport the set quantity of handled material within a defined period of time between the set loading and offloading locations.

In terms of belt conveyor operation, the most important structural component the **conveyor belt**. According to Zur [2], the belt conveyor is a limited-range, continuously moving transport facility that carries material on the belt surface, between two belts or inside a belt. According to Boroska [3], the conveyor belt transfers the resistances formed during the belt movement and serves for the transportation of materials, loads, or persons. Practical experience indicates that the most critical point is the place where the material is loaded onto the conveyor, the so-called chute. At the chutes, point impact loads are formed, being one of the main reasons for conveyor belt damage. If the impact energy is higher than the ability of the supports and the conveyor belt to absorb this energy, then conveyor belt damage occurs, especially at its upper cover layer, in form of transversal and longitudinal scratches, disruptions, and punctures, which can even cause damage to a conveyor belt carcass.

The conveyor belts impact resistance is classified as the ability of a conveyor belt to absorb the impact energy of the material's fall onto the belt, i.e. absorb the impact energy by the deformation work of the conveyor belt without damaging the belt.

In the theoretical part herein, the Authors focused their efforts on the creation of mathematical models, with an aim to accurately describe conveyor belt properties. The mathematical apparatus for the determination of reliability and optimal service life of the conveyor belts, while applying the renewal theory, was described in papers [4, 5]. The issues regarding conveyor belt damage were also discussed in papers [6, 7, 8, 9]. Models related to punctures of rubber-fabric and steel-cord conveyor belts were also discussed in papers [10, 11, 12].

The aim of this article is to provide the information on the solution of the problem regarding the analysis of measurements in the testing of conveyor belts during the impact tests, with regard to software optimization. At present, there is a simple tool for the measurement analysis; however, this tool was designed with regard to the measurements obtained from the device which is not currently used for the measurement recording. The aim of the research was to design a more complex tool for a deeper analysis of the problem with regard to new data provided by a new measuring device. The developed software application contains the tools for the analysis of the data measured, using the PP065 electronic device, for the manipulation therein, representation of relationships and the creation of regression models.

2 Material and Methods

2.1 Experiment and Data Acquisition Description

The testing equipment (Fig. 1) located at the Institute of Logistics of the Technical University in Košice is used for the impact testing of the conveyor belts. The purpose of the conveyor belt testing is to determine their impact resistance. The testing equipment contains a hydraulic system for belt specimen clamping, another hydraulic system for stretching a specimen during the test, and a laser detector recording the real-time height of the drop hammer. The structure of the testing equipment is based on the current requirements resulting from the conducted research as well as on the requirements of the conveyor belt manufacturer in Slovakia.

Figure 1
Testing equipment

During the test, the specimen of a selected conveyor belt is fixed in hydraulic clamps and is stretched using the force equal to 1/10 of the belt strength specified by the manufacturer. Subsequently, a drop hammer of a selected weight and shape is dropped onto it in a free fall. The drop hammer weight may range between 50 kg and 110 kg. In the testing, two types of end-pieces are typically used: pyramid-shaped and sphere-shaped (Fig. 2). The type of the used drop hammer end-piece characterizes the type of the transported falling material, in particular, a pyramid represents the impact of sharp-edge material and a sphere simulates the impact of non-compact, crumbling material. The drop hammer may be dropped onto the

conveyor belt from the maximum height of 2.6 m. The various belt types are used in the testing: rubber-fabric or steel-cord conveyor belts. During the testing, real-time height of the drop hammer, real-time impact and stretching forces of the conveyor belt are recorded [2].

Figure 2
Types of impact end-pieces

The test evaluation, in case of identification of the wear caused by the puncture, consists of the visual inspection of a conveyor belt and the identification of particular values of input parameters (weight, height) at which the conveyor belt damage occurred. Examples of the damage of the conveyor belt specimens are in Fig. 3.

| a) | b) | c) |
| d) | e) | f) |
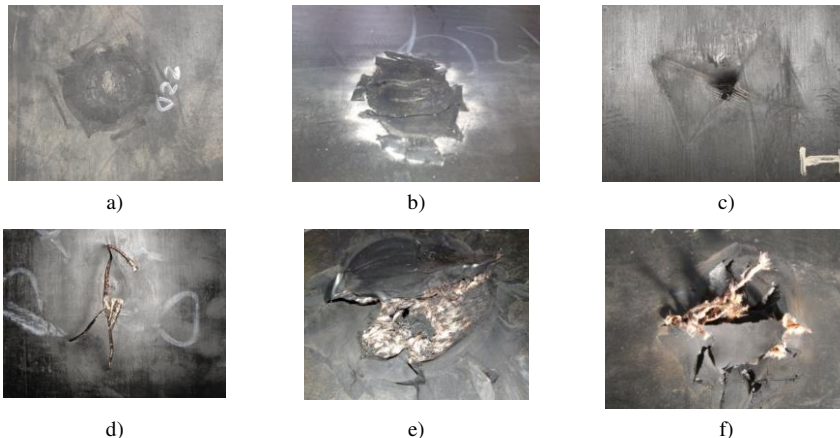
Figure 3
Examples of the damaged conveyor belt specimens

Figure 3a) shows the specimen without any visible signs of damage. The upper cover layer was undamaged, without any cracks caused by the falling material. Similarly, the lower cover layer did not carry any damage signs either. At the same time, no anomalies were observed in a tested specimen. The damage was manifested in the specimens (3b, 3c) by cracks or puncture at the site of simulated

material's impact. The conveyor belt fabric carcass was not visible through the crack and no torn fibers were present. Complete damage occurred in specimens 3d and 3e, where the fibers from the fabric carcass of the specimen were clearly visible; they protruded to the surface through the perforation in the cover layers. The most severe degree of damage – puncture (specimen 3f), is characteristic with the perforation interconnecting the lower and the upper cover layers. In the formed crack, torn carcass fibers may be observed; they may protrude to the surface. The upper and the lower cover layers are visibly damaged. Locally, parts of the upper cover layer may also be torn.

2.2   Regression Analysis

Many scientific and engineering problems include the examination of relationships between two or more variables. Regression analysis is a tool appropriate for solving these types of problems. The main objective of multiple linear regression is to search the linear relationships between a dependent variable and several independent variables and thus enable concurrent examination of the effects of several variables. The relationship between the explained (dependent) variable y and explanatory variables x_j, ($j = 1, 2, \ldots, k$) may be expressed by the following general linear regression in Eq. 1:

$$y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \alpha_4 x_4 + \alpha_5 x_5 + \ldots \alpha_k x_k + \varepsilon , \tag{1}$$

where α_0 is referred to as the intercept and coefficients α_j for $j = 1, 2, \ldots, k$ are referred to as the regression model parameters and ε is a random error of the regression model. In a standard case we assume that a random error has normal division with the mean value of 0 and the constant variance. A matrix notation of the linear regression model is shown in Eq. 2:

$$y = X\beta + \varepsilon , \tag{2}$$

where y is the n-element column vector of the measured values of the explained variable, X is the matrix of the $n \times (k+1)$ type of the measured values of the explanatory variables x_k, β is the $(k+1)$-element column vector of unknown model parameters, and ε is the n-element vector of random errors which represent the random component of the regression model.

To estimate the regression model parameters, we will use the method of least squares. To verify the statistical significance of the model, we will use the F-test of the statistical significance of the model. Statistical significance of the individual regression model parameters will be verified using the test of statistical significance of a regression parameter. Evaluation of the strength of the determined relationship between quantitative variables is carried out using the correlation analysis.

2.3   Applied Technologies

The problem in question was solved used using the Java language together with **AWT** and **Swing** technologies provided by this platform and the **JMathPlot, Apache POI, Apache Commons Math** libraries. The main advantage of the Java programming language is the independence of the platform. Java is currently, one of the most important and most frequently used programming languages [13] in the world. It is derived from C++, which is a direct descendant of C. Plenty of Java characteristics are derived particularly from these two languages.

3   Software Application

3.1   Hardware Requirements and Program Functions

The program serves for the analysis and subsequent work with measurements. The measured values are recorded using the PP065 electronic device (Fig. 4) which facilitates the control of the testing equipment and the recording of the measurement results.

Figure 4
PP065 electronic device

Recommended configuration for launching the program: Pentium class processor (500 MHz and more), 128 MB of RAM main memory and 10 MB of disc space. This recommended configuration only provides for launching the program; the main memory capacity depends on how many measurements we want to work with at the same time. Complete measurement, collection and evaluation of data are carried out as shown in Figure 5.

Recording of the measured values is carried out using the USB interface. The measured values are saved in form of text files. Data obtained from them are further processed using the created application. On the basis of these data, the software creates the mathematical models and verifies their statistical significance and makes graphical comparisons of the measured and the modeled data.
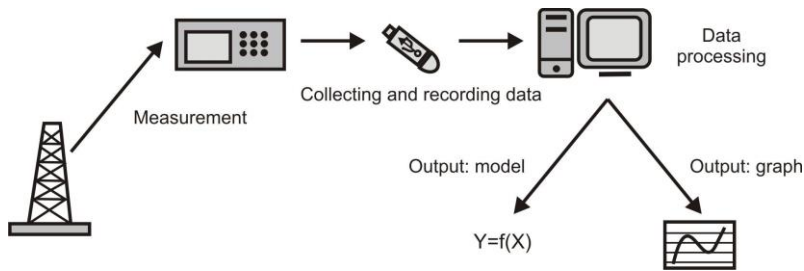
Figure 5
Collection and evaluation of data

Individual items of the saved file represent: time in milliseconds, height in millimeters, stretching force in kilonewtons, and impact force in kilonewtons. One text file contains 10,000 lines. Basic program functions are listed in Figure 6.
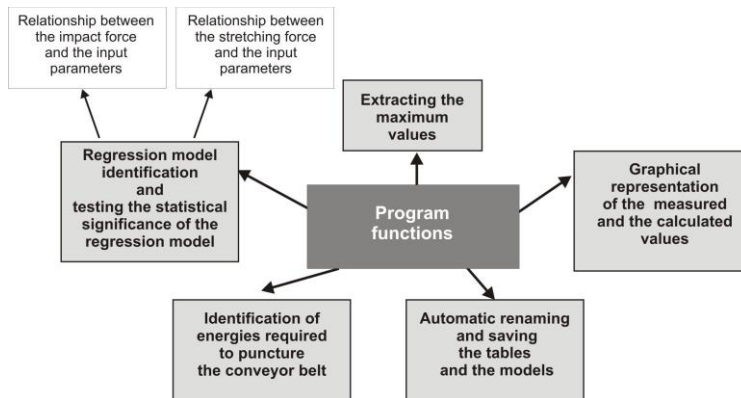
Figure 6
Program functions

The graphical user interface (GUI) in the entire program is created using the Swing library. The main program structure consists of the main menu, information panel, and the main panel. The information panel displays short messages on the program status and error messages. Error messages which require closer user attention are presented in a pop-up window.

The main panel also serves as the storage for program panels. This principle facilitates persistence of the main menu and the information panel when switching between the program functions. All program tools are thus independent panels that are inserted in the main program panel, as required. The program inputs are thus the measurements obtained from the PP065 electronic device in form of text files. These are subsequently analyzed using the amplitude searching algorithm. All the data, the measured and the analyzed, are saved in the memory, until the program is terminated or until other measurements are loaded.

3.2 Analysis of Input Files

As the drop hammer falls down onto the belt, the sensor records, using a laser, its real-time height; other sensors record the impact and stretching forces. For the illustration purposes, we will present the analysis and evaluation of the data obtained when testing the impact resistance of rubber-fabric conveyor belts.
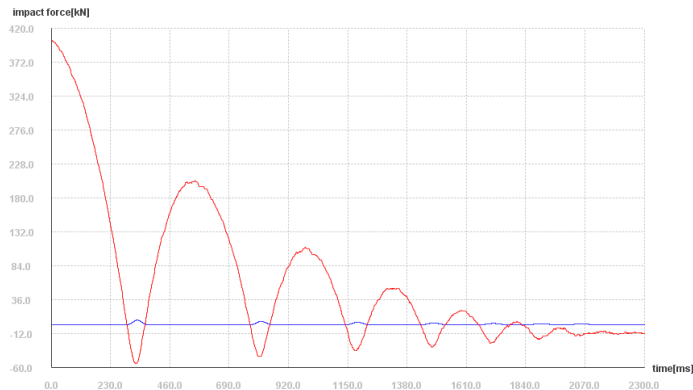
Figure 7

Graphical representation of impact force (blue) and height curve (red)

The plot in Figure 7 shows the development of the impact force and the height of the drop hammer in time for measurements on the textile-rubber conveyor belt with a 50 kg drop hammer dropped from the height of 0.4 meters. As shown in [12], there is a relationship between the impact force and the stretching force; therefore, to solve the problem of analysis of files, it is sufficient to focus only on one of the two forces.

The plot in Figure 8 shows the curve taken from the previous plot with added points designated with Arabic numerals. This plot represents one, particularly the first amplitude of this file:

- Point **0** coordinates represent the time and height from which the drop hammer started to fall
- Point **1** coordinates represent the time and height when the drop hammer fell down onto the belt
- Point **2** coordinates represent the time and height where the drop hammer was burrowed the deepest into the belt
- Point **3** coordinates represent the time and height when the drop hammer bounced off the belt
- Point **4** coordinates represent the time and height where the drop hammer bounced and from where it started to fall onto the belt again
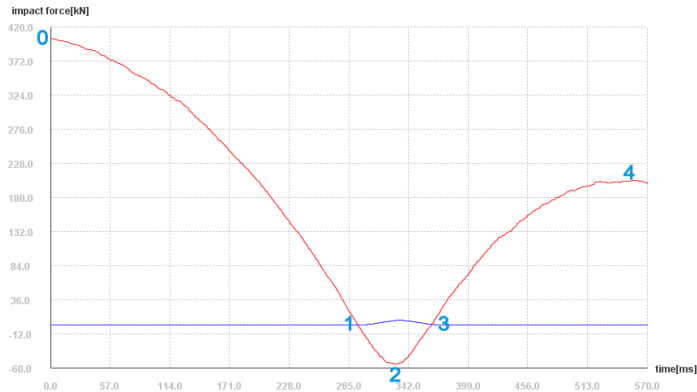
Figure 8
Graphical representation of the curve of one amplitude

The amplitude is thus the interval since the drop hammer started to fall from certain maximum height fell onto the belt and then bounced off to the following maximum height. Each file contains several such amplitudes; accurate determination thereof requires creation of the algorithm, i.e. the procedure how to identify these amplitudes correctly. Firstly, we will specify the data required for the correct identification of amplitudes. An important fact is that the measurement always begins with the drop hammer's fall; it means that at time 0 the first amplitude has the maximum height, i.e. the height from which the drop hammer is dropped.

Data required for amplitude identification:

- The maximum force and the time of maximum force

- The minimum height and the time of minimum height

- The maximum height prior to the impact and the time of maximum height prior to the impact

- The maximum height after the bounce and the time of maximum height after the bounce

- The drop height of the drop hammer and the time of impacting the belt

- The bounce height of the drop hammer and the time of the drop hammer's bounce of the belt

The measurement lasts 10 seconds and data are recorded every millisecond, it means that we have 10,000 data on the drop hammer height, the impact and the stretching forces. It is important that the number of records is final. Another important factor is that when the drop hammer falls down from the height x, after

the bounce the height can no longer achieve this value; this fact has the same effect in both measured forces. It means that every maximum value of the force and the height of the amplitude must be higher that the appertaining maximum value of the following amplitude. This also applies for the minimum height, i.e. how deeply the hammer penetrates the belt, except in the case of belt puncture. It is also necessary to note that if the height is 0, it means that the drop hammer is on the belt's level. Using these facts, the algorithm was created for searching amplitudes in each measured file. With the use of the found amplitudes, required plots and models will be created in the following step.

With the algorithm that searches all amplitudes we have to bear in mind that small inaccuracies occur at the measurements (e.g., at vibration, etc.) which result in the recorded amplitude. Therefore, not all the found amplitudes may be considered as valid. These invalid amplitudes are always located behind the valid amplitudes, unless certain invalid amplitude exceeds the values of normal amplitude.

3.3   Program Settings

Program settings are used to set certain behavior rules for the program which are saved in the disc. These settings include an option to set

- The force (impact or stretching) we primarily want to work with in the program

- The algorithm to be used in the model creation

- Whether the program is to search all amplitudes or only the first ones,

- The data elimination prior to the analysis (elimination of invalid amplitudes)

- The renaming certain variables that will be used for the creation of plots and tables

Standard settings are listed in Table 1. (All variables relate to the current amplitude.)

In the *Force* item we will select the force we want to work with. The program works with the selected force in cases when the force selection option is not enabled: *Stretch* (stretching force), *Impact* (impact force). In the *Algorithm* item we will select the algorithm for analyzing the input files: *Only first amplitudes* (the algorithm will select only the first amplitudes from each file), *All amplitudes* (the algorithm will select all amplitudes).
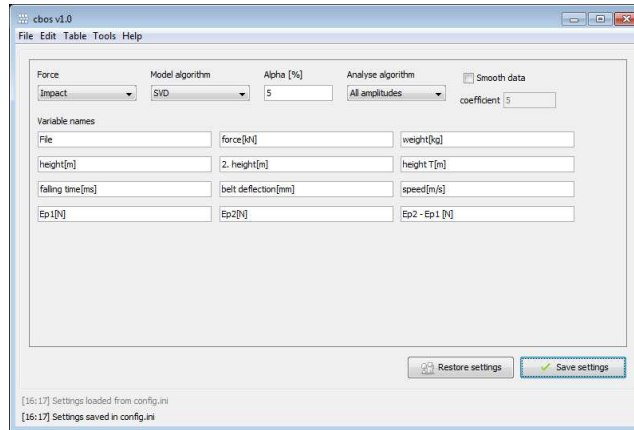
Figure 9
Settings panel

Table 1
Standard settings

Name	Explanation	Name	Explanation
File	File name	**falling time[ms]**	Time of the drop hammer's fall onto the belt
force[kN]	Selected stretching force	**belt deflection[mm]**	The depth into which the drop hammer was burrowed
weight[kg]	Drop hammer weight	**speed[m/s]**	Speed of the drop hammer's impact on the belt
height[m]	Drop height	**Ep1[N]**	Potential energy in the drop height
2. height[m]	Bounce height	**Ep2[N]**	Potential energy in the bounce height
height T[m]	Theoretical height obtained from the file name	**Ep2 - Ep1 [N]**	Difference between potential energies

In the *Model algorithm* item we will select the algorithm for solving the *Ax = b* problem in the model calculation using the method of least squares. In the *Alpha* item we will set the level of significance for the determination of significance of regression parameters. When we put a cross next to *Smooth data*, the *coefficient* value will be used and the input data will be eliminated prior to the analysis using the moving average method. The *Restore settings* button changes all setting to default values.

The *Save settings* button will save the made changes; without saving the changes, all changes will be lost. The changes are also saved in the *config.ini* file, so they will not be lost after the program is restarted. In the *Variable names* it is possible to set the name under which individual amplitude values will be displayed, for example when saving the model in the .xls format, creating the table using the *Edit data,* etc.

3.4   Program Tools

All the data, the measured and the analyzed, are stored in the memory until the program is terminated or until other measurements are made. They are stored in the structure accessed using the created interface. In addition to the data necessary for the identification of amplitude, also the data on belt name, drop hammer weight, assumed height from which the drop hammer was dropped, and data on energies are obtained and kept; these data are listed in the Analysis of Energies section. The program tools are accessible from the main menu and some of them from the *Loaded files* menu.
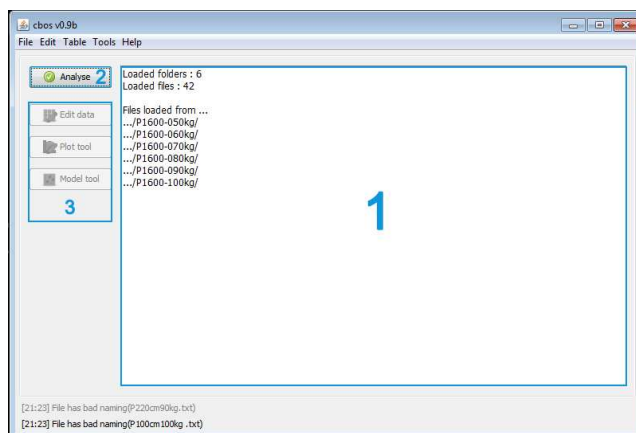
Figure 10
Loaded measurements

3.4.1   Loading of Measurements

The measurements may be loaded in any of the program statuses (except the error status) from the main menu via *File > Open Files*. It is possible to select only the folders, while their name must consist of the chain ***NNNkg**, where * is an arbitrary chain and NNN is the weight of the used drop hammer in kilograms in measurements contained in this folder. The folders contain .txt files with measurements which should have the ***NNNcm.txt** format, where * is an arbitrary chain and NNN is the height in centimeters from which the drop hammer was

dropped at such measurement. Unless the files end with NNNcm.txt, it will not be possible to work with the assumed height; this, however, currently does not represent any limitation, as we can identify the height from the file. The program only works with the theoretical height, if the user wants to sort out the measurements into categories by the height from which the drop hammer began to fall onto the belt.

After the measurements are loaded, the program automatically switches to *Loaded files* panel (Fig. 10) which contains, inter alia, the *Analyze* button (2) used to launch the analysis of files. The information on the analysis of loaded files and the results are displayed in the text section (1). Fast access to the program tools is possible by using the 3 button.

3.4.2   File Preparation

A tool (Fig. 11) is created to provide assistance with renaming the files.
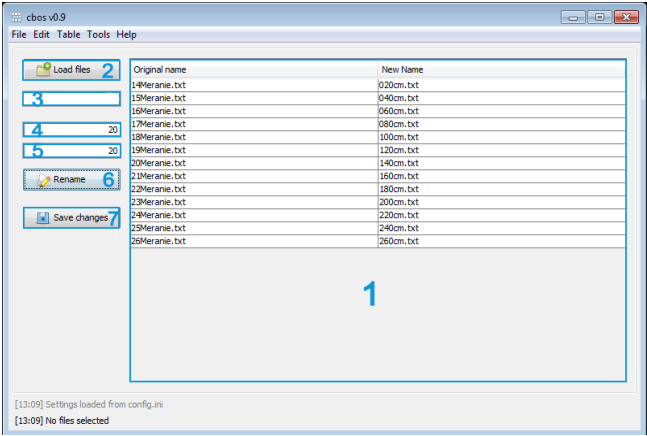
Figure 11
File renaming tool

The main table (1) shows the original and new names of files. Button (2) is used to load the recorded files. In section (3) we will select the prefix of renamed files (**NNNcm**), in (4) we will set the height at which the measurement began for the first file (in cm), in (5) we will select the value in which the height in cm should be increased in each following file. Button (6) is used to rename the files according to the set parameters and button (7) is used to save the required changes.

3.4.3   Data Editing

To edit the analyzed data, the tool (Fig. 12) was created; it facilitates displaying these data in form of tables. The tool enables viewing the data after individual measurements, whereas the data may be displayed either as the data on all amplitudes of this measurement or by individual amplitudes. The data may also be modified or all amplitudes may be deleted. It is also possible to display the plot (Fig. 12) which shows the points identified by the algorithm for searching the amplitudes. These points are identical to points in the plot in Fig. 8 and their meaning is explained in the Analysis of Input Files section.
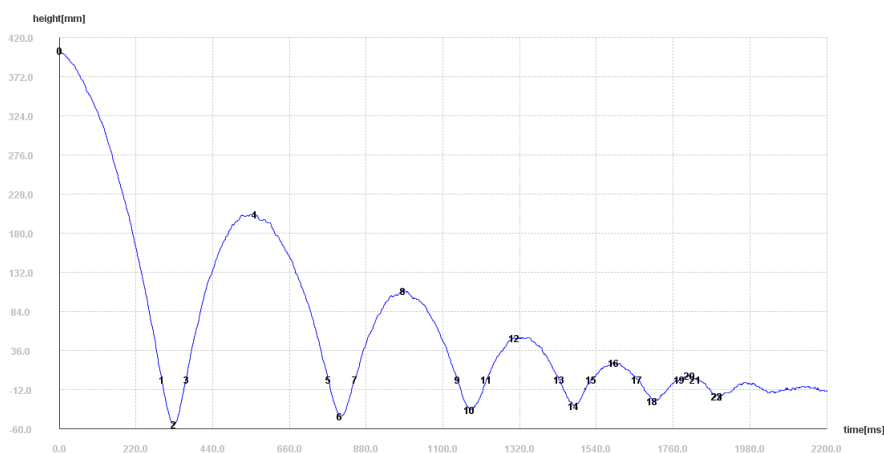
Figure 12
Graphical representation of height curve

Data editing serves for the adjustment of already analyzed measurements; to access the panel, we will enter the route *Edit > Edit data*, or *Edit data* from the *Loaded files* panel. Individual functions of this panel are explained in Figure 13.

When the table (1) is selected, we can choose *Amplitudes* (the table will display the data from all files for the amplitude selected in 2) or *Files* (displayed data on all amplitudes for the file selected in 2). For the output table we can select various parameters (4, height, weight, impact force, speed...) and the selected parameters will be displayed in the window (5). Parameter addition or removal is carried out using buttons (6). Using the *Plot file* function (7), we can draw the plot of the file showing the height in time and the data on where the algorithm has found the amplitude. The *Delete amplitude* functions enables deletion of the selected amplitude. The resulting table will be displayed in the window (3).
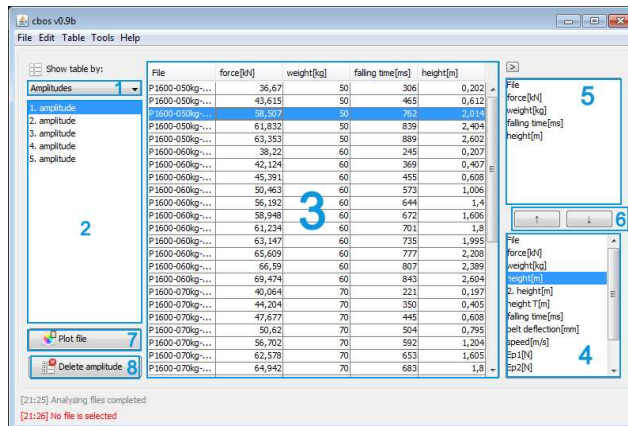
Figure 13
Data editing tool

3.4.4   Data Depiction

A tool for plot creation enables creating 2D and 3D plots from the measured and analyzed data. Depicted data may be displayed as a point chart, line chart, or a combination of both. The tool allows sufficient freedom when combining the depicted data. The data may also be automatically differentiated according to the selected parameters (e.g., by weight or height).

Data depiction is carried out using the *Plot tool* panel; to access the panel, we choose *Tools > Plottool*, or *Plot tool* from the *Loaded files* panel. Individual functions of this panel are explained in Figure 14.
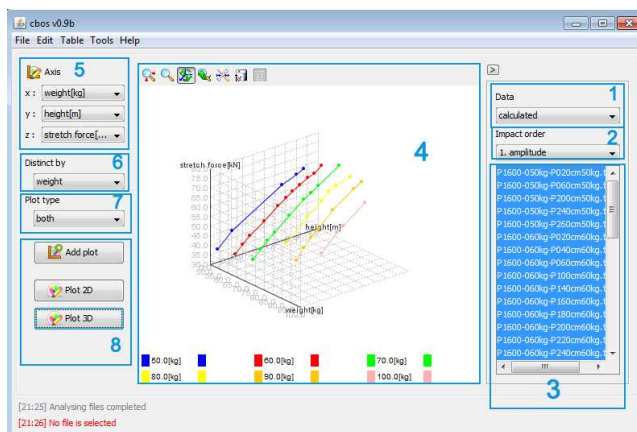
Figure14
Tool for plot creation

The *Data* function enables selection of data (1), selection of the file (3) and the amplitude (2, *Impact order*) for which we want to depict the data. When selecting the data to be used for a plot creation, we may choose the *Measured Data* (original data) or the *Calculated data* (analyzed data). Coordinate system selection is enabled by the *Axis* function (5). An option to sort the data by weight or height, or not to sort them (only for the analyzed data), is enabled by the *District by* function (6). When drawing the plot, it is possible to choose a plot type (7, *Plot type*): point, line, combined (point and line charts will be drawn at the same time using the same color. Buttons (8) serve for depicting or adding the data in the plot (*Add plot*). The plot as such is displayed in the window (4).

Zoom mode works exactly as a zoom tool, so if we want to adjust plot borders, they must be set in plot settings where it is also possible to rename the axes and adjust the range. The plot may be saved in the .png format. In table with the depicted data, it is possible to change color of individual plots, but not to change the data.

3.4.5   Model Creation

The regression models are created using the tool (Fig. 15) enabling the real-time calculation of various models according to the set parameters. The inputs for the creation of a regression model are regression parameters and the data from which the model will be calculated. The resulting data on the created model includes regression coefficients, determination of appropriateness of individual coefficients and a determination index. The model is evaluated by applying the method of least squares.

The *Model tool* panel serves for the creation of regression models from the analyzed measurements; to access the panel, we choose *Tools > Model tool*, or *Model tool* from the *Loaded files* panel. Individual functions of this panel are explained in Figure 15.

The *Add/Remove* function (1) serves for adding/removing a regression parameter from the regression function. Each parameter may be determined using the drop-down menu (2). The *For amplitudes* function (3) enables selection of the amplitude for which we want to create the model. Selection of the data we want to use for the model calculation is carried out using the *Use data* function (4). Model solving is enabled by the *Solve* function (5), whereas the model results are displayed in a separate window (6).

The *Plot model* function (7) enables model depiction. The resulting model may be saved using the *Save results* function (8) in the .xls format. This function enables saving also the depicted model plot.
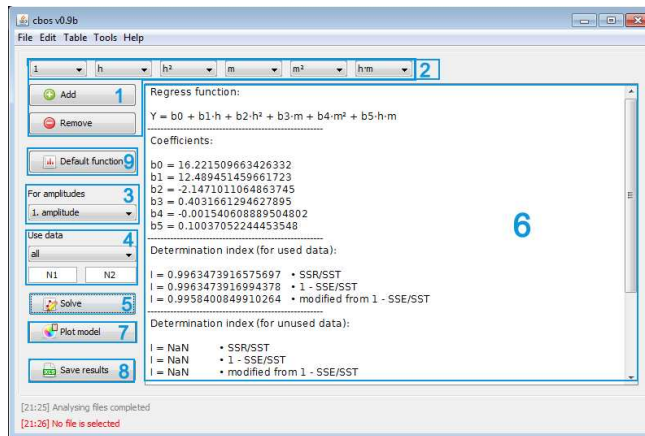
Figure 15
Model tool

3.4.6   Analysis of Energies

The software application also enables the analysis of energies. The real-time records for the drop hammer height may be used to identify the energy that the belt is able to absorb, just before the damage occurs on the basis of potential energies before and after the drop hammer bounce off the belt. Therefore, it is possible to make appropriate adjustments of input parameters to avoid the belt damage by puncture.

Conclusion

The electronics used in experimental measurements, aimed at evaluation and assessment of utility properties of conveyor belts, in terms of their dynamic stress, have until recently, facilitated only recording of measured values. Complicated and often time consuming initial processing and evaluation of the measured data is replaced with the AnCoBe software application described herein. This software application is particularly specialized for the solution of conveyor belt impact resistance in dynamic stress. Installation of this application does not require any special hardware or software. Similarly, work with the application does not require any special computer skills or elaborate training. The AnCoBe application includes a user guide, even though this software application control is intuitive. The application is freely distributable, including all sources and documentation, enabling the potential development and update by users as an open source method.

Acknowledgements

References

[1]   P. Kulinowski: Simulation Studies as the Part of an Integrated Design Process Dealing with Belt Conveyor Operation, Maintenance and Reliability. Vol. 15, 2013, pp. 83-88

[2]   T. Żur: Belt Conveyors in Mining. Poland: Śląsk Publishers, Katowice 1979

[3]   J. Boroska, Daniela Marasova: Nowe kierunki i doswiadczenia w zakresie budowy i eksploatacji tasm transporterowych i urzadzen z nimi wspólpracujacych, Gliwice: Politechnika Slaska, 1995, pp. 55-66

[4]   A. Pavliskova: Reliability and Continuous Regeneration Model, Acta Montanistica Slovaca. Vol. 11 (2) 2006, pp. 119-121

[5]   A. Pavliskova: Deterministic Strategies and Life-Time, Acta Montanistica Slovaca. Vol. 12 (3) 2007, pp. 265-269

[6]   U. Bugarić et al.: Reliability of Rubber Conveyor Belts as a Part of the Overburden Removal System - Case Study: Tamnava-East Field Open Caste Mine, Tehnički vjesnik. Vol. 21 (5) 2014, pp. 925-932

[7]   U. Bugarić et al.: Lost Production Costs of the Overburden Excavation System Caused by Rubber Belt Failure, Eksploatacja i Niezawodnosc .Vol. 14 (4) 2012, pp. 333-341

[8]   M. Bajda, D. Gancarek: Badania odporności taśmy aramidowej na symulowane uszkodzenia eksploatacyjne, Transport Przemysłowy i Maszyny Robocze. Vol. 1, 2015, pp. 26-30

[9]   L. Ambrisko et al.: Belt Damage Aspect to Impact Loading, Applied Mechanics and Materials. Vol. 683, 2014, pp. 102-107

[10]   G. Fedorko et al.: Failure Analysis of Belt Conveyor Damage Caused by the Falling Material. Part 1: Experimental Measurements and Regression Models, Engineering Failure Analysis. Vol. 36, 2014, pp. 30-38

[11]   M. Andrejiova, A. Pavliskova: Analysis of Regression Model of Functional Dependency in Impact Force from Height and Weight of Ram for Conveyor Belt, Annals of Faculty Engineering Hunedoara: International Journal of Engineering. Vol. 8, 2010, pp. 267-270

[12]   A. Grincova et al.: Regression Model Creation Based on Experimental Tests of Conveyor Belts against Belt Rips Resistance, Acta Montanistica Slovaca. Vol. 14 (1) 2009, pp. 113-120

[13]   S. Szalai et al.: An Overview of Low-Cost EGSE Architectures Improvement, Acta Polytecnica Hungarica, Vol. 13, 2016, pp. 139-158

Analysis of the Efficiency of the Recloser Function of LV Smart Switchboards

Judith Pálfi, Miklós Tompa, Péter Holcsik

Óbuda University, Kandó Kálmán Faculty of Electrical Engineering, Power System Department, Research Group of Applied Disciplines and Technologies in Energetics, Bécsi út 96, H-1034 Budapest, Hungary
palfi.judith@kvk.uni-obuda.hu, tompa.miklos@kksz-projekt.hu, peter.holcsik@elmu.hu

Abstract: The improvement of the quality of services is the main target of the electricity suppliers. This can be achieved by minimizing the number and duration in time of the power outages. The reliability indicators show the quality of the system. The quality of the services can be increased by proper maintenance, proper renovation activities and the implementation of advanced tools and functions into the system. Smart switchboard (SSB) devices are provided with modern and innovative functions which support the activity of the low voltage (LV) distribution network operators. Remote control and monitoring, fault location identification support, supply of on-line load-flow input data, and automatic recloser are only some examples of the newly available technologies. In this paper the improvement of the LV reliability indicators by using the recloser function of the SBB's is presented.

Keywords: LV network; quality of service; smart switchboard; recloser; fault location

1 Introduction

The AD&TE research group recommends the introduction of advanced switchboards to the power supply companies. This equipment is the smart switchboard (SSB). By the introduction of the SSB's, the duration of the power outages can be significantly reduced. This can be achieved by using the reclose function of the SSB (more detailed information about the reclose function of the SSB is provided in Section 1.1). Currently available LV swichboards are not equipped with the reclose function [1] which has been developed by the authors of this paper.

In this paper the research team will formulate recommendations to the electricity companies about the location of these devices. This recommendation is presented in the form of a function, the so-called efficiency function, which shows the expected improvement of the quality of services [2] [3].

1.1 LV Distribution Network Switchboards Today and Tomorow

The fuse protecting a section of the LV network melts down when a short circuit current occurs. The fuse protects the network from the damages caused by the short circuit currents. (The eventual faults of the measuring stations are not considered in this paper.)

The fuses are located in the outdoor or indoor switchboards (see Figure 1).

Figure 1
Outdoor and indoor switchboards in current use

Within the LV distribution network the switchboards can be found in different positions [4]. The LV network begins from the distribution equipment on the 0,4 kV side of the medium voltage (MV) / LV transformer stations (see upper part of Fig. 1) and ends by the consumers. A large number of switchboards can operate within a network depending on the protection design (see bottom part of Figure 1). In Figure 2 a typical LV network topology is shown and possible locations for the switchboards are represented.

Nowadays, if a fuse melts down, the supervisory dispatcher can detect it only from the error related calls [5]. On-line remote alarm- or remote monitoring systems are not available. Due to the technology of the fuse, the remote reclose is not possible.
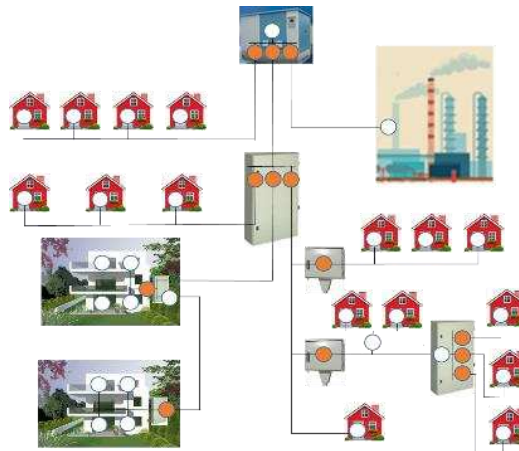
Figure 2
Typical LV network topology. The swichboards are highlighted with orange dots.

The remote alarm and monitoring functions are a basic condition for the on-line determination of the fault location (for details, please see earlier [6] paper of the research group on this subject). In the near future remote operation will play a significant role in handling the temporary short circuits.

1.2   Temporary Short Circuits

The temporary short circuits are caused by external influences (e.g. tree branch reaching to the overhead lines, arc-over caused by heavy rains, phase swinging due to powerful wind blasts, etc.) or short period overloads in the network. The common characteristic of these faults is that their repair requires only the change of the fuse(s) (1, 2 or 3 fuses, depending on how many phases were involved in the incident). No other intervention or installation works are required. The electricians will not have to be sent out to the location of the temporary short circuits as soon as the swicthboards will be provided with remote control systems [7].

1.3   Smart Switchboards

In order to implement the remote operation, the fuses will have to be replaced with circuit breakers. Contrary to the fuses, the circuit breakers do not have to be replaced following a short circuit power breakdown. Depending on their producer and type, the circuit breakers are capable of numerous operations during their working life. Measuring equipment and analysis systems [8] for the measured values (hardware and software) are to be integrated into the SSB's in order to detect the short circuits and to be able to issue the OFF instructions to the circuit breakers.

The SSB concept stands for a remote controlled LV switchboard which uses a circuit breaker for the dismantling of the short circuit current. The detection of the short circuit current is done by using a detection equipment together with the corresponding measurement analysis system. It is suitable for remote switch-on (circuit breaker activation) which, if necessary, can be turned to clogging mode. It contains the possibility of visible interruption point and earthing as well.

The visible interruption point and the earthing functions are required for ensuring the life, health and safety protection during the maintenance, reconstruction, etc. works.

The remote monitoring functions could actively or passively monitor the current, the voltage and the performance of the LV system.

The implementation of an automatic recloser, a so-called reclose function into the SSB is also possible.

1.4   The Reclose Function

When the reclose function is activated, the circuit breaker recloses after a previously specified time period following the defense action. If the defense continues to detect the short circuit, the circuit breaker opens again. Following this and after another pre-set time period, the automatics will close the contacts of the circuit breaker again. The automatics seeks to switch back two times (two cycles). The final release is activated if the short circuit persists. Nowadays, the reclose technology is already in use for MV networks [9]. Its working principle and the pre-set time periods for MV networks are shown in Figure 3.
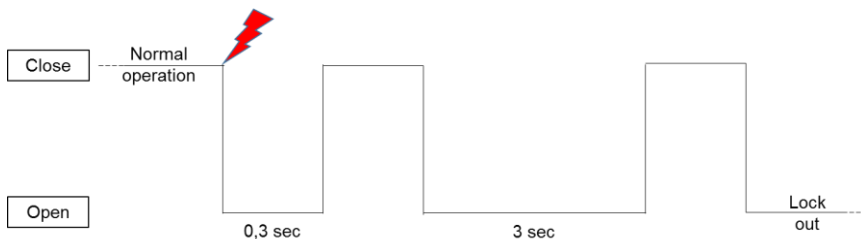
Figure 3

Working principle and pre-set time periods of the reclose function in case of MV networks

The implementation of the reclose function into the LV network will optimize the work of the electricians (they do not have to spend any more time going out to detect the faults in the network) and minimize the length of the LV power outages caused by short circuits. The earlier time consuming process (see Figure 4) is reduced to a few minutes (see Figure 3).
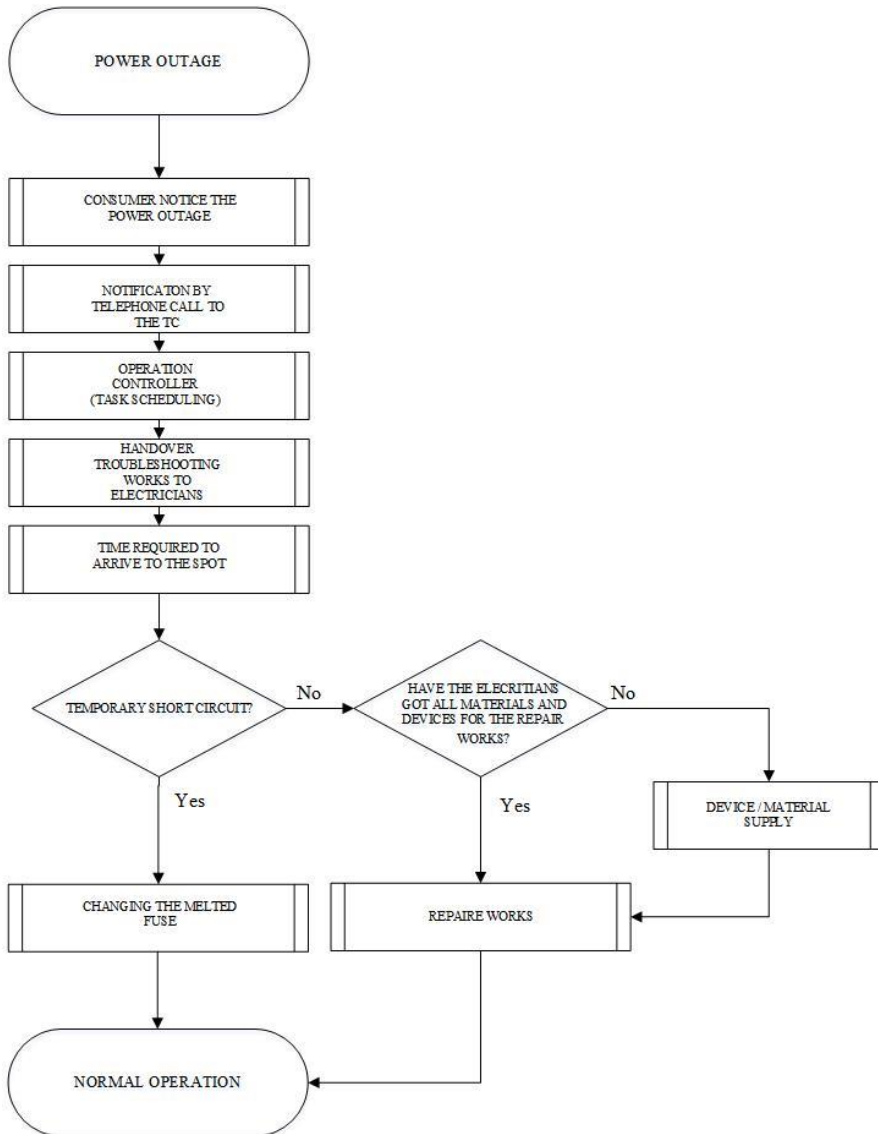
Figure 4
The LV power outage protection flow chart

The AD&TE research team has analized the improvement of the quality of the LV networks following the implementation of the SSB's and the efficiency of these advanced tools. The efficiency can be measured by the network's indicators of quality (for network quality indicators please see Section 2.1).

2   Method Used for Analysing the Efficiency of SSB's

When switchboards serving a large number of customers are upgraded to SSB's the related efficiency of these tools increases, that is a small number of devices protect a large number of customers [10]. This means that the efficiency of the reclose function implementation will be higher for those swichboards of which melting down can impact a larger number of customers. The efficiency will decrease as the number of customerss served by the upgraded switchboards is smaller.

The research group recommends to the power supplier companies to begin the implementation of the SSB's with the switchboards connected to larger number of customers*.*

By carrying out the study of the efficiency the research group determined the number of customers for which it is worthwhile upgrading the existing equipment. A cumulative efficiency function has been developed. This function shows the SAIDI improvement (Section 2.1) which could be achieved when changing some of the fuses to SSB's.

The research group examined the data collected from the LV power outages during years 2014 and 2015. The improvement of the SAIDI indicator has been determined assuming that during these 2 years SSB's with reclose function would have been used instead of the existing equipment (SAIDI indicator in section 2.1).

The input data for the study was provided by ELMŰ Network ltd and ÉMÁSZ Network ltd.

2.1   The Impact of LV Power Outages on SAIDI and SAIFI Indicators

The electricity suppliers use several indicators for measuring the quality of electricity networks. The Hungarian Energy and Public Utility Regulatory Authority is following two indicators and expects their improvement by the Hungarian electricity suppliers. These two indicators are the System Average Interruption Duration Index (SAIDI) and the System Average Interruption Frequency Index (SAIFI). [2] [11]

Nowadays, a LV power outage duration is taken into account from the time of the notification of the error until the time of reconnection of the customers. This time period is the outage time *(Ui).*

Another factor influencing the SAIDI is the number of customers affected by the power outage *(Ni).*

The electricity suppliers keep record of how many customers are affected by the breakdown of a certain equipment. Consequently, the impact on customers by the LV outage *i* can be determined from the IT databases.

The total number of customers (N_T) during a particular year for a particular company, e.g. ELMŰ or ÉMÁSZ is a constant value (pieces).

"The SAIDI network quality indicator is given by [2]:

$$SAIDI = \frac{\sum\limits_{x=1}^{n}(U_i \cdot N_i)}{N_T} \text{[sec]} \tag{1}$$

where N_i is the number of customers and U_i is the annual outage time for location *i,* and N_T is the total number of customers served.

In other words [2],

$$SAIDI = \frac{sum\ of\ all\ customer\ interruption\ durations}{total\ number\ of\ customers\ served} \tag{2}$$

SAIFI stands for the average number of power outages in a year related to the total number of customers.

The SAIFI network quality indicator is given by [2]:

$$SAIFI = \frac{\sum\limits_{x=1}^{n}(\lambda_i \cdot N_i)}{N_T} \text{[sec]} \tag{3}$$

where λ_i is the failure rate, N_i is the number of customers for location *i* and N_T is the total number of customers served. In other words,

$$SAIFI = \frac{total\ number\ of\ customer\ interruptions}{total\ number\ of\ customers\ served} \tag{4}$$

".[2]

The SSB device is installed on LV networks. Table 1 shows the ratio of MV and LV outages in the SAIFI and SAIDI indicators.

Table I

MV/LV ratio in SAIFIs and SAIDIs datas of ELMŰ Hálózati ltd., head of Pest,

01. – 08. 2015

	SAIFI	SAIDI
LV	0.09	0.72
MV	0.53	0.85
Sum of LV and MV	0.62	1.57
Ratio of the LV / MV	14.51% / 85.49%	45.85% / 54.14%

From the data in Table 1, it can be seen that the power outages in the LV networks have a significantly smaller influence on the SAIFI indicators (14.51%) than on the SAIDI indicators (45.85%). Therefore, the AD&TE research group considered the calculation method of SAIDI for setting up the cumulative efficiency function. The results are given in minutes, similarly to the SAIDI indicator.

The power outage data (outage time, number of affected customers) are gathered by the electricity suppliers based on the worksheets filled out by the electricians. These worksheets are checked and validated by the LV operation controllers [12] [13] [14].

2.2 The Analysis of the Worksheets

The electricity suppliers are recording into their own IT system the customers' calls related to the power outages [15] [16]. These notices are then converted into digital worksheets by the LV operators. The worksheets are distributed to the electricians taking into account their current geographical location, their skills and other relevant information. Figure 5 illustrates the input interface of a worksheet.

Figure 5

Electrician worksheet from ELMŰ-ÉMÁSZ's work management and LV operation control system

After the repair work has been completed, the electrician fills out the worksheet. This is done with the help of drop-down optional lists. He will choose and complete the lists with information related to the cause of the error, whether one or more customers were affected, the type and category of the failed equipment (e.g. transformer station, switchboard, devices of the measuring station, etc.) and/or device (cable, overhead line, fuse, etc.).

In addition to the drop-down lists, free texts can be introduced as well. These could be notes and/or explanations related to the size of the replaced lining measured in amper or the type of the repair which has been carried out (e.g. replacement of a circuit breaker, measuring equipment exchange, reconnection of a cable, etc.).

The information concerning the type of short circuits which caused the power outage (temporary or long term) is not recorded in the worksheets. Whether only a fuse was changed or a more complex repair work was carried out (e.g. cutting the branches of trees, reconnection of collapsed overhead lines, etc.) is not specified. Due to this reason, the database in its current form does not allow for a clear and rapid detection of the power outages caused by short circuits. Concerning that, the worksheets must be analyzed one by one, taking into account all the information which has been given in the free text notes and in the drop-down lists as well.

It can happen that the electrician replaced a fuse and the reason of the melt down of it was not a temporary short circuit but for example, a cable of the network was cut by a branch of a tree or by an excavator working nearby.

During the two years period of the study a total number of 21351 worksheets were generated (after data cleaning and not including the faults of the public lighting and of the measuring stations). The examination of the worksheets, one by one, would have taken more than 550 working hours. Therefore, the research group decided to apply a sampling method for the analysis of this large amount of information. [17]

2.3   Selection of the Sample

The analyzed population are the worksheets generated by ELMŰ ÉMÁSZ during the years 2014 and 2015. The sampling frame consists of as many as 21351 worksheets. Through the sampling process the randomly selected information from the data gathered in 3-3 weeks of the 2 years period was processed.

The selected weeks are shown in Table 2.

Table 2
Weeks randomly selected in years 2014 and 2015

	2014	2015
The calendar number of the selected weeks	09	05
	13	15
	37	28

The free text entries of the selected weeks' worksheets were filtered from the point of view of whether the reason of the error was clearly due to a temporary short circuit or not. Manual analysis was not carried out in the cases in which the power outage was not due to temporary short circuits. Further possibilities for filtering the free text entries are described in Section 7. After filtering the 3-3 weeks' worksheets 2060 pieces remained for manual analysis.

3 Analysis of the Sample and Results

The sample selected and filtered as described in Section 2.3 consists of 2060 worksheets. The data were analyzed manually, one by one. In the analysis of the worksheets (see Section 2.2) the information related to the duration of the intervention of the electricians and the description of the type of the faults was considered. The faults were categorized in faults which could have been saved by providing the SSB's and faults where the provision of the SSB's would have had no positive impact. A similar selection has been carried out from the point of view of the customers.

The results of the analysis are presented in Table 3.

Table 3

Results of the analyzed sample

Calendar number of selected weeks	Number of events which could have been saved by providing SSB's (pieces)	Number of customers who could have been saved by providing SSB's (pieces)	Outage time which could have been saved by providing SSB's (minutes)
2014. 09th	123	5771	19.448
2014. 13th	88	3289	12.548
2014. 37th	87	3548	14.241
2015. 05th	158	6092	23.566
2015. 15th	94	3609	10.646
2015. 28th	1026	32362	354.155
Total:	**1576**	**54671**	**434604**

Based on the results shown in Table 3 temporary outage was recorded in 1576 cases in which the electricians had to only replace the melted fuse(s). Consequently, 1576 power outages (which affected 5467 customers during a total of 434.604 minutes) could have been minimized by implementing the SSB's. Extrapolating the results the annual theoretical maximum efficiency can be obtained.

3.1 Determination of the Annual Maximum Efficiency

The theoretical maximum of SAIDI improvement can be calculated from the data of Table 3.

In 1st result column of Table 4 the LV power outages are shown which affected more than one customer within the given week. This data shows the load of power outages in that week.

Table 4
Determination of the maximum efficiency

Examined weeks year 2014 and 2015	More than one customer affected by the LV power outages	Given week cons*time (min)	Average cumulative errors per week	Correlation coefficient	SAIDI change with SSB revised for a week period	SAIDI change with SSB for one year revised with calendar effects	Maximum change of SAIDI	Maximum due to the FACTUAL percentage
2014 09th	1107	902801	6171	5.57	2.23	38.83		
2014 13th	2730	425350	6171	2.26	0.43	7.42	50.73	25%
2014 37th	5920	556267	6171	1.04	0.26	4.47		
2015 05th	790	995045	4778	6.05	2.67	46.44		
2015 15th	1410	381228	4778	3.39	0.57	9.97	67.72	29%
2015 28th	36820	11298735	4778	0.13	0.65	11.31		

It can be seen that while during the 5th week of 2015 there were only 790 power outages in the 28th week there were 36.820. The later figure was generated by the extreme weather conditions.

The 2nd result column of Table 4 shows the value of the numerator of SAIDI indicator formula.

In order to extrapolate the measured data it is necessary to examine the way the number of power outages from the considered week relate to the annual week average. This relation is shown in the correlation coefficient column of Table 4.

ELMŰ ÉMÁSZ calculated the quality indicators in 2014 with 2.252.508 pcs customers, in 2015 with 2.246.941 pcs customers. By using these two values and the correction factor the values given in 4th column of Table 4 were determined. The extrapolation was done by carrying out the correction and following the calculation method of the SAIDI indicator.

From the weekly results it can be determined that if SSB's would have been installed in all the possible locations of the LV network, the SAIDI indicator would have been lower with 50.73 minutes in 2014. The 205.2 minutes represents 25% of ÉMÁSZ ELMŰ factual data. In 2015 this value is 67.72 minutes which stands for 29% of the annual factual data of 231.32 minutes.

3.2   Distributions and the Comparison of Distributions

The research group's preliminary assumption was that the distribution of the measured data and the number of customers behind the fuse boxes from the network correlate with each other. In order to verify this hypothesis, the number of customers affected by the power outages was examined.
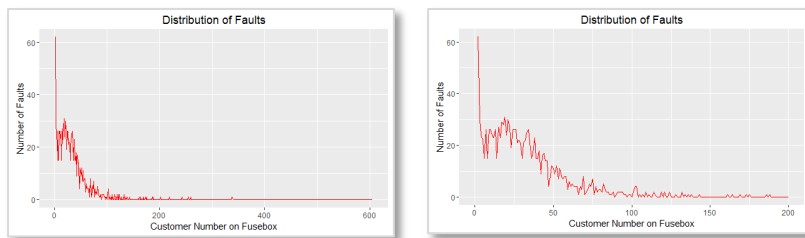
Figure 6

Distributions of the costumers behind the faults, data measured in 2014 and 2015

Left side: normal plot, right side: the plot is zoomed along the horizontal axis

between 0 and 200 costumers

The distribution illustrated on Figure 6 has been compared with the distribution of the number of customers from behind all of the fuse boxes of the ELMŰ ÉMÁSZ LV network.

The distributions of the customers numbers behind the fuse boxes is shown ib Figure 7.
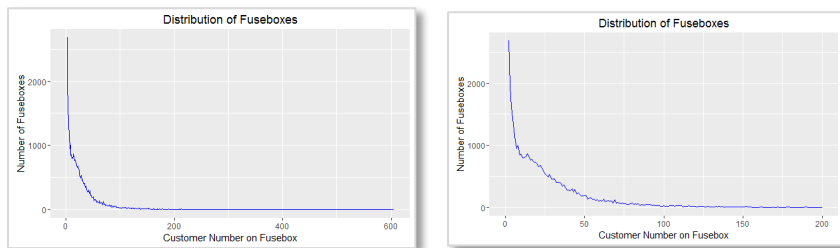
Figure 7

Distributions of the customers behind the fuse boxes on the ELMŰ-ÉMÁSZ network, 2016

Left side: normal plot, right side: the plot is zoomed along the horizontal axis between 0 and 200

customers

Data of Figure 6 and Figure 7 illustrated in one coordinate system:
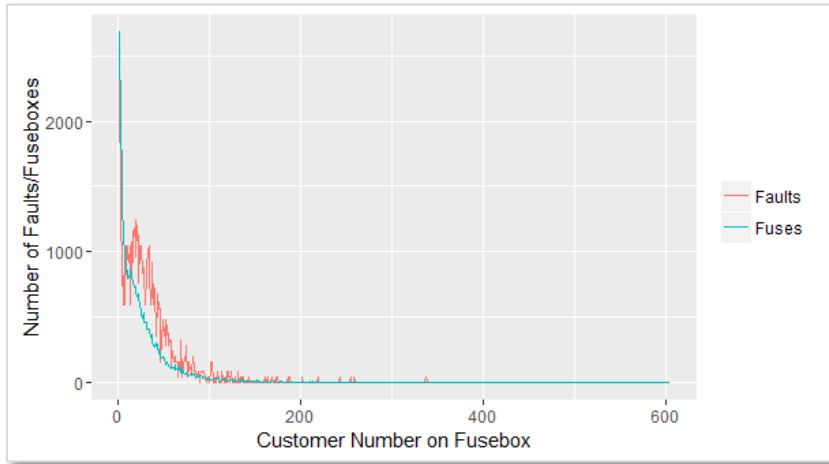
Figure 8

Distribution of the costumers numbers behind the measured fuse boxes and number of faults in 2014
and 2015

Figure 8 shows that the assumption of the research group was confirmed: both
curves show an exponential decreasing along the horizontal axis. However, beside
the correlation, differences can be observed between the two curves. The reason
for these differences is detailed in Section 6.

4   The Efficiency Function

The recommendation described in Section 2 was to begin the implementation of
the SSB's into the equipments connected to larger number of customers.

By carrying out the study of the efficiency, the research group determined the
number of customers for which it is worth upgrading the existing equipment. A
cumulative efficiency function has been developed. This function shows the
SAIDI improvement (Section 2.1) which could be achieved when changing some
of the fuses to SSB's.

For example, in case N=105 customers:

$$Cumulative\ utility = \sum_{N=105}^{N_{\max}}(N_i \cdot U_i) \tag{5}$$

where N is the number of customers in the LV network behind the fuse, $N_{\max}$ is the
maximum number of customers behind a fuse, N_i is the number of customers
affected by the power outage i, U_i is the duration of the power outage i.

The cumulative efficiency function for the measured weeks is given in Figure 9.
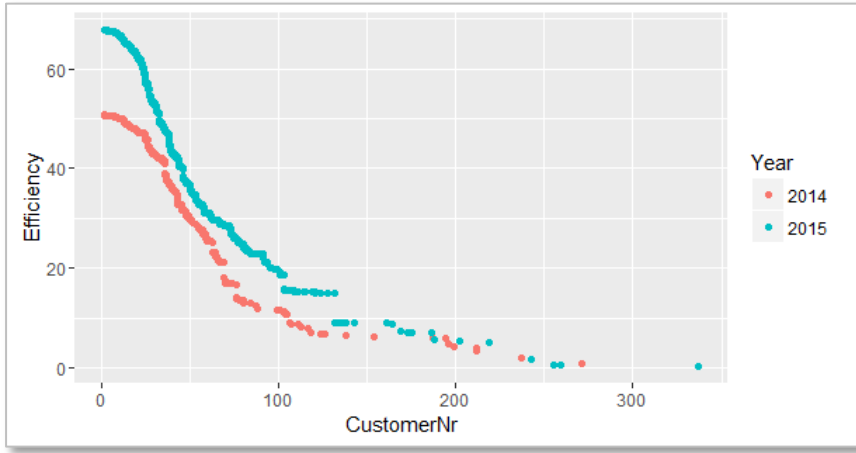
Figure 9
Efficiency in 2014/2015 observations

The data in Figure 9 shows an exponential distribution. The cumulative efficiency function decreases exponentially with the increase of the number of customers *N*, that is with the exponential increase (see Figure 7) of the number of implemented devices.

4.1 The Determination of the Efficiency Function Using the Regression Analysis

The curve from Figure 9 regarding the measured data in year 2015 has been determined using regression analysis. The regression model was developed by using the 3rd order polynomial regression according to which:

$$CU(N) = 74.526 - 0.8611 \cdot N + 0.003317 \cdot N^2 - 0.000004095 \cdot N^3 \qquad (6)$$

The function of the estimation fits closely the data measured according to the statistic features: it describes the measured data inasmuch as 96.96% (see Section 6).

The representation of the function (6) and the measured data of 2015 is given in Figure 10.
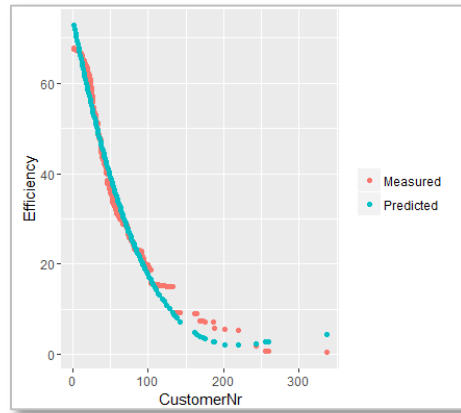
Figure 10
Efficiency and costumer number, 2015 3rd order regression model

The Figure 10 shows that the data for over 100 customers is under represented (for more details see section 6). However, it illustrates properly the matching between the model and the measured data in the lower ranges.

Hence:

The second recommendation of the research group to the electricity suppliers is to use the efficiency function for determining the number of customers by which it is worthwhile to upgrade the switchboards to SSB's with reclose function. Boundary condition: N <100 (for details see section 6).

In Figure 11 the data measured in year 2014 are shown in the same diagramm with the data of the regression model from year 2015. If the regression model would forecast the expected improvement with 100% certainty, then the value of the green "error" curve should converge to zero everywhere. The green "error" curve represents the difference between the two functions.
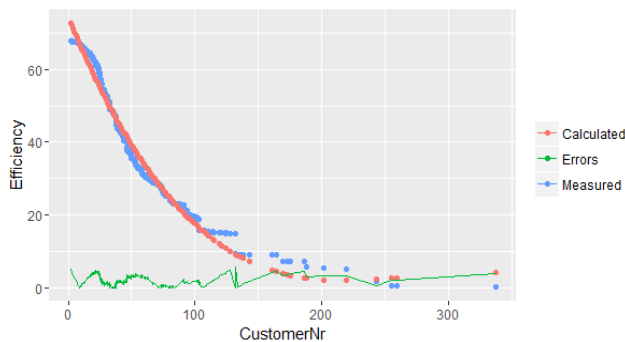
Figure 11
Efficiency and customer number year 2014 using 3rd order regression model of 2015

5 The "Platform Anomaly" in the Measured Results

It can be observed that the shape of the curves for customer numbers between 100 and 180 is levelled. In this interval the efficiency indicator remains almost constant although the number of customers increases. The cause of this so-called "platform" anomaly is discussed in Section 6.
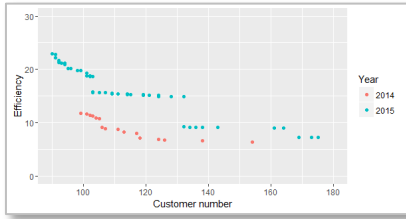
Figure 12

The efficiency curves

in years 2014 and 2015
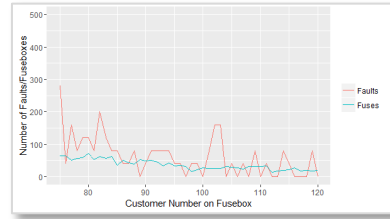
Figure 13

Distributions of the costumers

number behind the faults and fuse boxes

5.1 The Sectioned Efficiency Function

Due to the anomaly presented in Section 4 the research team cut the efficiency function in 3 different sections. The boundaries of the sections are:
1. 0-102
2. 103-131
3. 132-337
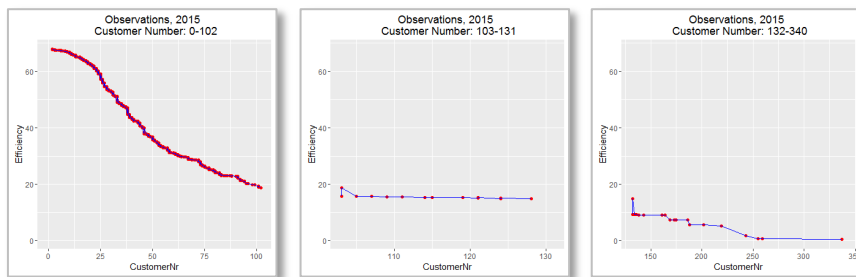
Figure 14

The measured efficiency function in year 2015 divided in three sections

In the first case the polynomial- while in the second case the linear regression model was used. In the third case due to the small number of data (19 pcs observations) the second degree polynomial regression has been applied.

The results obtained by this procedure are the following:

$$CU(N)_{0<N<103} = 66.36 + 0.5309 \cdot N - 0.05254 \cdot N^2 + 0.0007689 \cdot N^3 -$$
$$0.000003444 \cdot N^4 \tag{7}$$

$$CU(N)_{102<N<132} = 22.11972 + 0.5803 \cdot N \tag{8}$$

$$CU(N)_{131<N<337} = 27.82 - 0.1573 \cdot N + 0.0002191 \cdot N^2 \tag{9}$$

The three models supperposed in one coordinate system are shown in Figure 15. The representation of the three functions supperposed with the measured data is given in Figure 16.
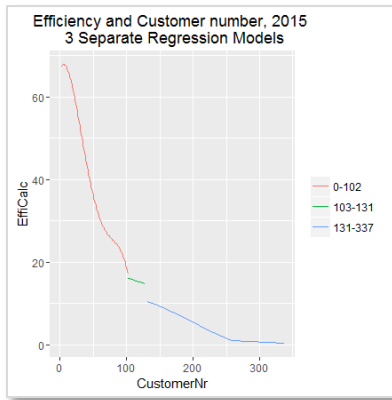
<table>
<tr><td>Figure 15</td><td>Figure 16</td></tr>
<tr><td>Three efficiency functions from year 2015 supperposed in one diagram</td><td>The divided into three efficiency function in 2015 in one figure with the mesured data</td></tr>
</table>
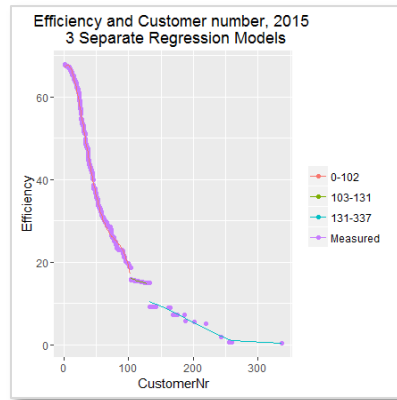
In Figure 16 it can be seen that the separated regression models describes more accurately the measured data than the model shown in Figure 11. However, the research group does not recommend to the electricity suppliers the usage of this model due to the reasons described in Section 6.

6   The Accuracy of the Model

The statistic model is an estimation based on the statistical data used to create the model itself. Thus, a regression model will give the function which describes a curve closest to the sample points. The function itself is not necessarily passing exactly through the given points.

There are available certain indices which qualify the model:

1.   p value and t-test: these show if there is any correlation between the data and the degree we can relay on it. If $|t| > 2$, then we can state with a

reliability of 95% that there exists a correlation. For| t |> 2.6 we obtain 99% reliability.

2. R^2: in the present context its value gives the percentage ratio of the description of the sample by the model. In case of the non-separated model its value is 96.96% which confirms a high level superposition (with the reserves described in section 4.1)

3. Standard Deviation: in the present context it shows how close the actual values are in relation to the expected values. In case of a non-separated efficiency function it is 2.6 at the beginning of the series of data. However, by the end of the series, this number is above 70.

4. Residuals: the difference between the values calculated with the measured and the model data which stands for the error. See Figure 11.

The above confirms that the models describe properly the measured data.

The compliance of the non-separated regression model is limited by the fact that for over 100 customers (*N*) only few measured data is available. This fact could significantly change the model. One of the reasons for "platform anomaly" is also due to this fact. Because of this under-representation, further development of the non-separated model is recommended. The disadvantages of this model can be elimininated by more frequent sampling.

Conclusion

The manual processing of the entire data is limited. The current aim of the AD&TE research group is to organize and to transform the entire data in order to analyze it with machine learning methods. Using the descriptive data (including the free texts) predict with simpler random forest and more complicated neural network whether a certain event is an error manageable by SSB or not. By this method, the entire data could be examined. This would result in the set up of a more accurate model.

As next step, the research group plans to examine the impact of the weather conditions on the LV power outages' data.

Acknowledgment

References

[1]   Dr. F. Novothny, Power System I, Kandó Kálmán Faculty of Electrical Engineering, Óbuda University, Budapest 2010

[2]   "1366-2012 - IEEE Guide for Electric Power Distribution Reliability Indices" Revision of IEEE Std 1366-2003 (Revision of IEEE Std 1366-1998), May 31 2012, DOI:10.1109/IEEESTD.2012.6209381

[3]   T. Portik, L Pokorádi: The Summarized Weighted Mean of Maxima Defuzzification and Its Application at the End of the Risk Assessment Process, in Acta Polytechnica Hungarica, Vol. 11, Issue 3, 2014, pp. 167-180, DOI: 10.12700/APH.11.03.2014.03.11, ISBN: 1785-8860

[4]   B. Barze, A. Sharma, A. L. Barabási: Graph Theory Properties of Cellular Networks, Handbook of Systems Biology – Concepts and Insights, Chapter 9, Academic Press, Elsevier, 2013, pp. 177-193, ISBN-13: 978-01023859440

[5]   H. Li, R. Yeh, Y. Lin, L. Lin, J. Zhao, C. Lin, I. J. Rudas, Medical Sample Classifier Design using Fuzzy Cerebellar Model Neural Networks, Acta Polytechnica Hungarica, Vol. 13, No. 6, pp. 7-24, 2016

[6]   J. Pálfi: Localization of Faults in Low Voltage Networks by the Graph Method, in Proceedings of the 11th IEEE International Symposium on Applied Computational Intelligence and Informatics SACI 2016. Conference: Timisoara, Roumania, 2016.05.12-2016.05.14., 2016, pp. 397-400, ISBN: 978-1-5090-2379-0

[7]   J. Haakana, J. Lassila, T. Kaipia, J. Partanen, Comparison of Reliability Indices From the Perspective of Network Automation Devices, Power Delivery IEEE Transactions, Vol. 25, pp. 1547-1555, 2010, ISSN 0885-8977

[8]   A. Varga, E. Rácz and P. Kádár: New Experimental Method for Measuring Power Characteristics of Photovoltaic Cells at Given Light Irradiation; IEEE 8th International Symposium on Applied Computational Intelligence and Informatics, SACI 2013, Timisoara, Romania, Print ISBN: 978-1-4673-6397-6, DOI: 10.1109/SACI.2013.6609008, 2013, pp. 405-409, May 2-25, ISBN: 978-146736400-3, Source Type: Conference Proceeding, Original language: English, DOI: 10.1109/SACI.2013.6609008

[9]   Zs. Mitrik: Operation of Third Generation Sulphur-Hexafluoride Circuit Breakers in Medium Zone of Short Circuit Currents, Kandó Kálmán Faculty of Electrical Engineering, XXXI. Kandó conference 17. 11 2016, Óbuda University, Kandó Kálmán Faculty of Electrical Engineering, pp. 53-54, ISBN 978-963-7158-07-0

[10]   Pausits P, Szögi G, Nallbani M, Rudas I J, Haidegger T: Identification of Risks in the Application of Surgical Robotics, 19th IEEE International Conference on Intelligent Engineering Systems, Bratislava: INES 2015, pp. 147-151, ISBN:978-1-4673-7938-0

[11]   K. Zou, W. W. Keerthipala, S. Perera, "Saidi Minimization of a Remote Distribution Feeder" in Australasian Universities Power Engineering

Conference AUPEC, Perth, Australia, 2007, pp. 342-346, ISBN 9780646494883

[12] J. Kopják, J. Kovács: Timed Cooperative Multitask for Tiny Real-Time Embedded Systems, Applied Machine Intelligence and Informatics (SAMI) 2012 IEEE 10th International Symposium on, 2012, pp. 377-382, DOI: 10.1109/SACI.2011.5873022

[13] R. E. Precup, S. Preitl: Stability and Sensitivity Analysis of Fuzzy Control Systems. Mechatronics Applications, Acta Polytechnica Hungarica, Vol. 3, No. 1, pp. 61-76, 2006

[14] R. E. Precup, C. A. Dragos, S. Preitl, M. B. Radac, E. M. Petriu: Novel Tensor Product Models for Automatic Transmission System Control, IEEE Systems Journal, Vol. 6, No. 3, pp. 488-498, 2012

[15] I. J. Rudas, J. Gáti, A. Szakál, K. Némethy: From the Smart Hands to Tele-Operations, Acta Polytechnica Hungarica Vol. 13, No. 1, 2016, pp. 43-60, DOI: 10,12700 / APH.13.1.2016.1.5

[16] Impact of TCSC on the Transient Stability / Zsolt Čonka, Michal Kolcun - 2013. In: Acta Electrotechnica et Informatica. Roč. 13, č. 2 (2013), s. 50-54. - ISSN 1335-8243

[17] A. L. Kazakov, A. Lempert: On Mathematical Models for Optimization Problem of Logistics Infrastructure, International Journal of Artificial Intelligence, Vol. 13, No. 1, pp. 200-210, 2015

Synchronized Dancing of an Industrial Manipulator and Humans with Arbitrary Music

Figen Özen[1], Dilek Bilgin Tükel[2], Georgi Dimirovski[2]

[1] Haliç University, Department of Electrical and Electronics Engineering, Sütlüce Mah, Nr 82, İmrahor Cad, 34445, Istanbul, Turkey, figenozen@halic.edu.tr

[2] Doğuş University, Department of Control and Automation Engineering, Hasanpaşa Mah, Zeamet Sok, Nr 21, 34722, Istanbul, Turkey
dtukel@dogus.edu.tr, gdimirovski@dogus.edu.tr

Abstract: An extended Labanotation for an industrial robot was developed and applied. A user-friendly program was designed to help design dance choreography. The program, LabanRobot, has a simple interface that can be used without any prior knowledge of the robot. The choreographer should keep geometry in mind and plan the dance sequence accordingly. The conversion of dance sequences to robot motion is done automatically by the program. Details of the algorithm are given. Examples of simulation and on-stage-performance are shown.

Keywords: industrial robot; Labanotation; dance

1   Introduction

Since the beginning of human existence, the environment and life itself has posed many difficult problems for humans. The strength of humans was limited and insufficient to achieve the many tasks that life on earth required. Humans may have been physically weak, but their imagination and intelligence were strong. To overcome their physical limitations, they imagined artificial humans to help. The dream of artificial servants dates back to before Christ. Aristotle said, "If every tool, when ordered, or even of its own accord, could do the work that befits it, then there would be no need either of apprentices for the master workers or of slaves for the lords." He imagined such a machine even when humanity was far from realizing this dream. Eventually, automata, which can do some humble work for people, were finally created. But they were far from satisfying human imagination. Research and attempts to design better ones accelerated and goals became even more ambitious when more tools were available.

Today, in the age of Industry 4.0, millions of robots are in use. Research continues to build more effective industrial robots to achieve more sophisticated tasks as part of the smart factory concept, where robots act as cyberphysical systems. They communicate with other robots and humans through an Internet of Things (IoT). Human–robot collaboration and this is a challenge, since industrial robots may accidentally harm humans. There has been a lack of sensors that produce data similar to the human senses and a shortage of intelligent control algorithms to avoid accidents during collaboration.

In this work, an industrial robot arm was programmed to perform a popular dance. A simple, user-friendly graphical interface was created to aid inputting the choreography. The user does not have to be a programmer to design a dance. Color codes and simple buttons are used to simplify the design process. Section 2 discusses the notations of movement and dance. Section 3 gives a brief review of the literature. Section 4 describes the extended Labanotation proposed in this paper and its application with a robot. Section 5 draws conclusions and indicates further research topics.

2   Expressing Robot Motion

Robot motion has been a research topic for decades. When it comes to humanoids, movement is expected to be human-like, but modeling human motion is a very complex task. After designing a humanoid, then the inevitable problem arises: How should one tell the robot to behave like a human?

If a human has to open a door, she or he goes to the door and opens it. This task usually requires no deep thinking for a human. On the other hand, if a humanoid is made to open the door, then the task must be split into a sequence of sub-tasks. For example, locating the door handle, raising the arm, grasping the handle with the hand, pushing the handle, pulling the door, releasing the handle, and stopping.

All the sub-tasks must be translated into a language that allows the robot to perform them correctly. This takes the sub-task into another domain, where the application of control results in action. For example, the sub-task of raising the arm must translated to torque at specific angles by certain actuators in the joints.

2.1   Translation of Sub-Tasks into Machine Language

To translate sub-tasks into machine language, various notation systems can be used. There are around ninety systems but only a few are used in robotics. The most popular ones are Labanotation, Benesh notation, and Eshkol Wachman notation, in order of popularity. They have been borrowed from the field of dance, but are not limited to it. They have also been used in gymnastics, mime, circus

performance, anthropology, ergonomics, neurology, clinical research, animation, and human–computer interaction [1].

Due to the complexity of movement, there has been great interest in representing it in simpler terms. There have been many attempts to codify movement, especially in the field of dance. In dance education, the traditional method is supervised learning, where a teacher shows a student how to dance a given pattern. This requires extended one-on-one study, is not very efficient, and relies too much on the student's memory. The dance notation systems strived for efficiency and reliability.

Dance notations are done with many different forms such as letters, numbers, stick figures, and realistic figures. There have been attempts to represent robot movement with dance notation, especially humanoids. There is no notation that can adequately express every possible movement. This leaves the roboticist with fewer options. The majority of the work employing movement notation in the robotics uses Labanotation.

In the following section, the Laban, Benesh and Eshkol Wachman notation systems will be reviewed, since they are by far the most common ones. All three notation systems were developed in the twentieth century. They require a considerable amount of time to learn. Even though they share the same aim, their approaches are different.

2.1.1   Laban Notation

Laban notation was devised by the dancer and the choreographer Rudolph Laban. It is also called Kinetography Laban and Labanotation, which was coined by Ann Hutchinson Guest and has gained popularity. Kinetography Laban presents a compact view of the body.

In Laban notation, successive movements are shown vertically and simultaneous movements are shown horizontally. The duration of each movement is indicated by vertical length. The part of the body involved is also expressed vertically [2].

The Laban system describes a movement in terms of the following elements:
   a)   Body
   b)   Space
   c)   Effort
   d)   Shape

The first element, body, identifies which body parts are involved. The 'space' element indicates the space and extent of the movement. Here, the Kinesphere, the term coined by Rudolph Laban, is to be highlighted. It refers to the volume created and occupied by a human. Depending on the movement, it can become larger or smaller [3]. The 'effort' element refers to the energy involved. The 'shape' element is the form of the body and its evolution during the movement.

Laban notation is very detailed. There is also an abbreviated representation, Laban motif notation. This was devised by Ann Hutchinson Guest and others to use in dance education, and focuses on what is different or similar. Especially when the repetitive patterns are involved, it saves time and energy [4].

Labanotation uses a staff, similar to a music score. Unlike a music score, Labanotation is read vertically from bottom to top. It is divided into columns to represent the body parts (Fig. 1). The centerline represents the vertical center of the body. The left side of the staff represents the left side of the body, and the right side of the staff, the right side of the body.
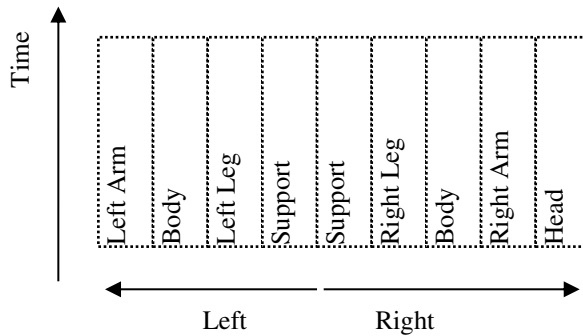
Figure 1

Organization of the staff of a Laban score

Labanotation has symbols for nine different directions, which can be high, medium or low, so there are 27 possible combinations. Fig. 2 shows the direction symbols and position marks. In the figure, the shape of the symbol indicates the direction, shading, and the level or height of the movement.
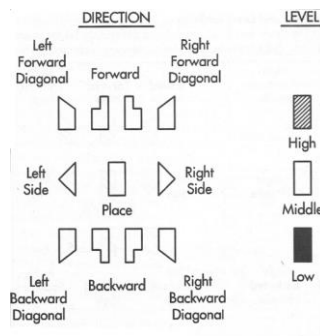
Figure 2

Laban symbols

An example of Laban notation is given in Fig. 3.

Figure 3
Example of Labanotation

Among Labanotation programs, LabanWriter is the most widely used. The program was developed at the Ohio State University. It prepares Labanotation scores and digitizes them [5].

Laban movement analysis is based on Rudolph Laban's work but it deals with dynamics and effects that Labanotation does not include.

The Effort Shape notation was also devised by Laban, but it is less popular. It includes the following parameters:
    a)   Weight (light, strong)
    b)   Space (indirect, direct)
    c)   Flow (free, bound)
    d)   Time (sustained, quick) [6]

Here, time refers to rhythm, tempo, phasing, and duration of the movement. Weight refers to the movement's softness, smoothness, sharpness, and energy. Flow is about the combination of variation of movements, using recurring elements, contrast, and repetition. Basic types of movements are dabbing, gliding, floating, flicking, thrusting, pressing, wringing, and slashing [7].

2.1.2   Benesh Notation

Benesh notation was devised by Rudolph Benesh and Joan Rothwell in the mid-twentieth century. It uses key frames. It is drawn on a staff, like Laban notation, but sequences are written horizontally from left to right. The staff has five vertical lines, each for different body parts. Stronger movements are indicated by a bolder drawing. Fig. 4 shows an example of Benesh notation.
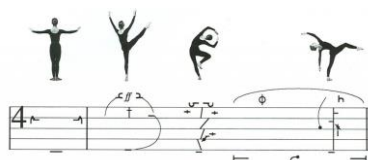
Figure 4
Example of Benesh notation

2.1.3   Eshkol Wachman Notation

Eshkol Wachman notation was devised by the dancer and choreographer Noa Eshkol and the architect Abraham Wachman in the 1960s. In this notation, the body is viewed as a composition of articulated limbs, which can lie between two joints or be connected to a joint at only one end. It is recorded in a matrix, not a staff. The left side of the matrix indicates which body parts move. Sequences are written from left to right and the end of the movement is shown by a bold bar [8]. Fig. 5 shows an example.
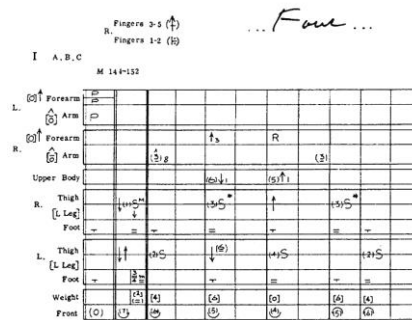
Figure 5

Example of Eshkol Wachman notation

3   Literature Review

Research on dance in the robotics field deals with many problems, which can be grouped as motion capture, processing the captured images/videos, processing music to dance with, representing the required motion using dance notation, converting the notation into inputs for the joints of a robot, and evaluating the dance performance of a robot. Naturally, not all of these issues are dealt with in every publication.

Peng, Zhou, Hu, Chao and Li [9] provide a thorough review of the robotics dance field since the 1990s. They describe the following research areas: robotic dancers that can cooperate with humans, imitate human dance, synchronize with music, and creating robotic choreography. They emphasized that cooperation between a robot and a human is harder to achieve than coordination between robots and that the imitation of human movement by a robot is still in its infancy. They conclude that synchronizing with music requires many aspects of music to be taken into consideration, but so far, only beat, tempo, and rhythm have been studied. Finally, they mention that work remains to be done on creating choreography with human aesthetic values and improvisation.

Afsar, Cortez and Santos [10] surveyed approaches to human behavior detection from video. The steps involved are initialization, tracking, pose, and recognition. They review each step in detail and tabulate the literature to aid future research.

Most papers deal with a certain aspect of robot dance or a specific dance. The rest of this section summarizes some of the outstanding research in the field.

Hachimura and Nakamura [11] segmented motion data and quantize motion direction and duration for Labanotation. A Laban editor with limited scope was developed for dance education.

Kojima, Hachimura, and Nakamura [12] modeled the human body with 21 joints and study the speed of joint motion. Labanotation data was converted into animation data. Only basic Labanotation functions were simulated, thus compromising the smoothness of the motion. An interface for LabanEditor was introduced. Interaction with the environment was not taken into account.

Nakamura, Niwayama, Tabata, and Kumo [13] developed a dance training system consisting of a motion capture system, a projector for projecting the teacher's footprints on the floor, and a robot acting as a mobile mirror. Two traditional Japanese dances were used for testing. The mirror robot had to position itself at a suitable distance as the student rehearses. Since the student cannot look at the mirror robot and the teacher's footprints at the same time, the reasoning has to be modified.

Suzuki and Hashimoto [14] combined a mobile platform with a motion interface system, a controller, and output devices to study human robot interaction with a semiautonomous robot. In response to a performer's dance, the program iDance produces sounds and moves in compliance with the shifting weight of the performer. Neural networks were used to represent emotional states.

Kosuge and Hirata designed [15] a dance robot to study human–robot coordination coupled with physical interaction. The robot has an omni-directional mobile base and it can sense the force and torque applied by its human partner. The robot was tested with the Waltz. More dance forms need to be tested.

Choi, Kim, Oh, and You [16] balanced the dancing arm of a humanoid robot with a posture and walking control scheme that utilizes the kinematic resolution method of Center of Mass (CoM).

Jeong, Seo, and Yang [17] designed a wearable interface equipped with magnetic sensors to teach human motion to a humanoid robot. Recorded motions are processed to obtain key frames. Clustering is applied to the key frames. A computer analyzes and enacts motion primitives and kinematic calculations. The system was tested using simple motions, namely drawing a heart and boxing. More complex motions have to be taken into consideration to imitate human motions effectively.

Kekehashi, Izawa, Shirai, Nakanishi, Okada, and Inaba [18] studied the motion of hula hooping. Torso movement was analyzed and the algorithm was tested using dedicated robots. Some stability problems were encountered. Since hula hooping is simpler than dance, in general, the algorithm needs to be adapted to the challenges of dance and later to arbitrary human motion.

Dong, Wang, Wang, Yan, and Chen [19] studied the lion dance with a 19 degrees-of-freedom, under actuated, multi-legged robot. Kinematics, dynamics, and stability of the lion dance were studied, which was performed by two humans. The results showed that the robot lacked the liveliness and loveliness required to perfectly imitate the lion dance.

Wang and Kosuge [20] considered the case of the Waltz. This dance has a leader and a follower. In the study, the human was the leader and robot was the follower. The human dancer's body was modeled using an inverted pendulum. The position of the leader was estimated with data from laser range finders and an extended Kalman filter. Next step estimation was done with Hidden Markov Models. Since dancing the Waltz requires haptic interaction, human arm dynamics were used in the control loop. The paper leaves the questions of role switching and rotational motions unanswered. Robustness and optimality are issues to be dealt with.

Auguliaro, Schoellig, and D'Andrea [21] used the Laban Effort Shape concepts, namely space, time, weight, and flow, in the choreography of a group of quadrocopters. Motion and music were synchronized, and collision-free flight trajectories were calculated, but the motion library is limited and audience feedback is not incorporated.

Zhou, De la Torre, and Hodgins [22] extended Kernel K Means and Spectral Clustering to partition human motion into a fixed number of classes. Testing was done on with Carnegie Mellon University Motion Capture database videos. The results show that the computational load is very heavy, and some simplifications need to be employed. In addition, the algorithm has to be sped up for real-time application.

Radac and Precup [23] classified the ways of learning from primitives as a) time-scale transformation approaches, b) temporal concatenation of primitive-based approaches, and c) time-based decomposition approaches. The primitives are useful to simplify the task, which can be very complex. They suggested an iterative learning control algorithm, which does not require much prior information about the system to be controlled. They solved the trajectory tracking problem.

Wang and Zhou [24] used a method based on time-based b-spline-decomposition for a continuous and smooth trajectory tracking. The desired trajectory was decomposed online and the control input was synthesized. The application to a microscope probe resulted in better output tracking over feedback control.

Mussa-Ivaldi and Solla [25] investigated the primitives for motion control in neurobiological systems. They studied adaptive behavior in the nervous system for learning, making use of knowledge stored in memory.

Grymin, Neas and Farhood [26] worked on motion planning in an environment with obstacles. Their approach was based on temporal concatenation of motion primitives and the Graph theory. They provided their simulation results on a hovercraft.

Okamoto, Shiratori, Kudoh, Nakaoka, and Ikeuchi [27] dealt with a cyclic Aizu-bandasian dance. The dance was performed by a human and was captured using a motion capture system, sampled, and then key poses were extracted and performed by a humanoid based on key poses. The human dance was decomposed by splines. The humanoid dance consisted of lower-, middle-, and upper-body movements. Each movement was split into sub-tasks to be performed by each body part. The lower-body supported the whole body, the middle-body supplied the expression of dance and maintained balance, the upper-body expressed the dance. The algorithm has not yet been applied to other dance forms to demonstrate universality.

Ramos, Mansard, Stasse, Benazeth, Hak, and Saab [28] used motion capture to form a reference movement for a humanoid dancing with a human. To generate the movements, an operational space inverse dynamics method was used, to solve the problem of following the dance figures demonstrated by a human. The problems of coordination, control, and stabilization of balance were studied; balance was given top priority over other objectives such as visibility and posture. To maintain balance, linear momentum variation was controlled by center-of-mass acceleration and angular momentum variation was controlled by proper contact forces. Due to the lack of sensor feedback, the performance of the platform fell short of expectations.

Yoshida, Shirokura, Sigiura, Sakamato, Ono, Inami, and Igarashi [29] reported an interface for an entertainment system called RoboJockey. The interface controls a mobile and a humanoid robot. Among other actions, the robots automatically perform according to music. The humanoid robot moves its knees in response to the beat. The mobile robot does not have enough actuators to move with the rhythm, thus it simply switches actions in response to the beat. The system and the interface need improvement because high level functionality, including programming loops and sensor feedback, is not supported.

Takano and Nakamura [30] represented human motion through segmentation and encoded the segments into Hidden Markov Models. Afterwards, statistical analysis tools, such as correlation matrices, were used to predict the next motion. The algorithm should be optimized and parameter tuning should be adaptive. The correlations between the parameters need to be tested.

4 Experimental Setup

For the research presented here, a computer program was written to make an industrial manipulator dance together with a human. To design a robot dance similar to human dance, the required movements are expressed with extended Labanotation. The industrial manipulator is a six-degree-of-freedom Mitsubishi RV-7FL, which is geometrically similar to a human arm.

A Laban editor was designed and programmed through a special interface. The controller calculates the input torque parameters of the manipulator after the required motion data has been supplied from the interface.

To handle objects, a hand was designed for the industrial robot. The hand consists of a vacuum pad and a vacuum generator. During dance, the hand carries dance related accessories such as flowers, matador's cape, camera, etc.

The control of the robot arm is done by CR-750D, which has servo drivers and a motion controller inside. Hierarchical trajectories are generated by Matlab and RT Toolbox 2. The host computer communicates with the robot controller via Ethernet. The interconnection is shown in Fig. 6.
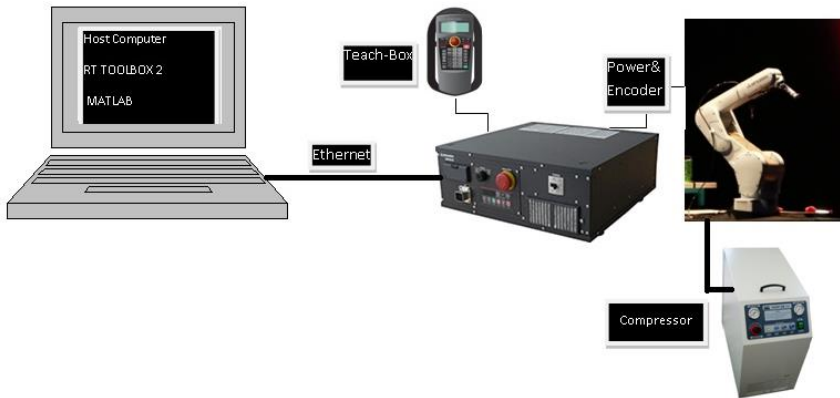
Figure 6

Mitsubishi RV-7FL robot arm with CR-750 controller and the connection of system components

To translate Labanotation into robot movement, the kinematic model of the manipulator is needed. The derivation is done and the specifications of the manipulator are summarized in Table 1.

Link lengths were obtained from the CAD data supplied by the manufacturing company [31]. The external dimensions in Fig. 7 were also supplied by the manufacturing company.

Type	Joint Number	Mitsubishi RV-7FL	Type	Joint Number	Mitsubishi RV-7FL			
Degrees of freedom		6	Maximum load capacity		7 kg			
Maximum reach radius		908 mm	Mass		67 kg			
Operating range (degree)				1	±240	Max speed (deg/sec)	1	288
				2	-110+130		2	321
				3	-0+162		3	360
				4	+200-200		4	337
				5	-120+120		5	450
				6	-360+360		6	10977
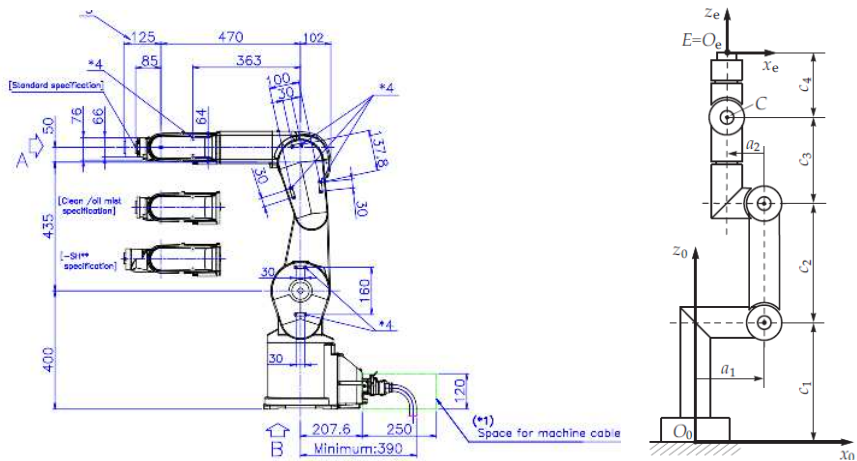
Figure 7
Mitsubishi RV-7FL external dimensions (left) and ortho-parallel basic and spherical wrist robot with seven essential geometric parameters [31]

Mitsubishi RV-7FL has a 3R ortho-parallel basis structure with a 3R wrist. Link and joint offsets were calculated and verified using a simulation toolbox. A summary of the kinematic analysis and the parameter values are shown in Table 2.

Table 2

Structural Kinematic Parameters for Mitsubishi RV-7FL

Joint No	Joint angles	Link Length (mm)	Offsets (mm)
1	Θ_1	$L_1=c_1=400$	$a_1=0$
2	Θ_2	$L_2=c_2=435$	$a_2=-50$
3	Θ_3	$L_3=c_3=470$	$b=0$
4	Θ_4	$L_4=c_4=85$	
5	Θ_5	0	
6	Θ_6	0	

The forward and inverse kinematic problems were solved with Brandstötter, Angerer, and Hofbaur's algorithm [32] for Mitsubishi RV-7FL.

Transformation matrix of the end effector with respect to the base is given as:

$$T_0^6 = T_0^1 T_1^2 T_2^3 T_3^4 T_4^5 T_5^6 = \begin{bmatrix} & & & x \\ & R(rpy) & & y \\ & & & z \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{1}$$

The end effector position is given by:

$$p = [x \quad y \quad z \quad \theta_x \quad \theta_y \quad \theta_z]^T.$$

where

$$T_0^6(1,1) = s_1(c_4 s_6 + s_4 c_5 c_6) - c_1(c_{23}(s_4 s_6 - c_4 c_5 c_6) + s_{23} s_5 c_6)$$

$$T_0^6(1,2) = s_1(c_4 c_6 + s_4 c_5 s_6) - c_1(c_{23}(s_4 c_6 + c_4 c_5 s_6) + s_{23} s_5 s_6)$$

$$T_0^6(1,3) = c_1(s_{23} c_5 + c_{23} c_4 c_5) + s_1 s_4 s_6$$

$$T_0^6(1,4) = L_4(c_1 s_{23} c_5 + c_1 c_{23} c_4 s_5 - s_1 s_4 s_5) + L_3 c_1 s_{23} - L_2 c_1 s_2 - a_2 c_1 c_{23}$$

$$T_0^6(2,1) = c_1(c_4 s_6 + s_4 c_5 c_6) - s_1(c_{23}(s_4 s_6 - c_4 c_5 c_6) + s_{23} s_5 c_6)$$

$$T_0^6(2,2) = c_1(c_4 c_6 + s_4 c_5 s_6) - s_1(c_{23}(s_4 c_6 + c_4 c_5 s_6) + s_{23} s_5 s_6)$$

$$T_0^6(2,3) = s_1(s_{23} c_5 + c_{23} c_4 c_5) + c_1 s_4 s_5$$

$$T_0^6(2,4) = L_4(s_1 s_{23} c_5 + s_1 c_{23} c_4 s_5 + c_1 s_4 s_5) + L_3 c_1 s_{23} + L_2 s_1 s_2 + a_2 s_1 c_{23}$$

$$T_0^6(3,1) = s_{23}(s_4 s_6 - c_4 c_5 c_6) - c_{23} s_5 c_6$$

$$T_0^6(3,2) = s_{23}(s_4 c_6 + c_4 c_5 s_6) + c_{23} s_5 s_6$$

$$T_0^6(3,3) = (c_{23} c_5 + s_{23} c_4 s_5)$$

$$T_0^6(3,4) = L_1 + L_2 c_2 + L_4(c_{23} s_5 - s_{23} c_4 s_5) + L_3 c_{23} + a_2 s_1 c_{23}$$

Wrist position can be calculated as:

$$[x_w \quad y_w \quad z_w]^T = [x \quad y \quad z]^T - L_4 R_6^0 [0 \quad 0 \quad 1]^T$$

$$\theta_1 = atan2(y_w, x_w)$$

$$\theta_2 = acos\left(\frac{M^2 + L_2^2 - K^2}{2ML_2}\right) + atan2(N, z_w - L_1)$$

$$\theta_3 = acos\left(\frac{M^2 + L_2^2 - K^2}{2KL_2}\right) + +atan2(a_2, L_3)$$

$$\theta_4 = atan2(\boldsymbol{T}_0^6(2,3)c_1 - \boldsymbol{T}_0^6(1,3)s_1, \boldsymbol{T}_0^6(1,3)c_{23} + \boldsymbol{T}_0^6(2,3)c_{23}s_1 - \boldsymbol{T}_0^6(3,3)s_{23})$$

$$\theta_5 = atan2(\sqrt{1 - H^2}, H)$$

$$\theta_6 = atan2\big(\boldsymbol{T}_0^6(1,2)s_{23}c_1 - \boldsymbol{T}_0^6(2,2)s_1s_{23} + \boldsymbol{T}_0^6(3,2)c_{23,} - \boldsymbol{T}_0^6(1,1)s_{23}c_1 \\ - \boldsymbol{T}_0^6(2,1)s_{23}s_1 - \boldsymbol{T}_0^6(3,1)c_{23}\big)$$

where

$$N = \sqrt{x_w^2 + y_w^2}$$

$$M = \sqrt{N^2 + (z_w^2 - L_1)^2}$$

$$K = \sqrt{a_2^2 + L_3^2}$$

$$H = (\boldsymbol{T}_0^6(1,3)s_{23}c_1 + \boldsymbol{T}_0^6(2,3)s_{23}s_1 - \boldsymbol{T}_0^6(3,3)c_{23})$$

A new Matlab-based program called LabanRobot was developed and used to write Labanotation for an industrial robot [33]. The user interface of LabanRobot is shown in Fig. 8. The choreographer can easily select the dance sequence, thinking about the geometry of the motion only, and imagining the robot as a human arm. The choreographer does not need to have any prior knowledge about the robot. They can design a dance using the interface and clicking on the rhythm and level boxes. The program plans the trajectory of the specified geometry.

The notation used in LabanRobot is different from traditional Labanotation. There are five levels, as an extra 'extension level' was added. The extension level, depending on the motion, is expressed with vertical or horizontal bars. Traditional Labanotation uses longer symbols for movements of longer duration, whereas LabanRobot does so with different colors. Rhythm is also represented by color. Using colors makes the script shorter.

The choreographer has to select the attributes of each movement: the rhythm (slow, medium, fast), the height or the level (low, medium-low, medium, medium-high, high), the extension (in, medium-in, medium, medium-out, out) and the direction. Next the movement is added to the dance sequence. The choreographer repeats this procedure for each movement in the sequence. Finally, the whole sequence is converted to a program by pressing the button 'Create Robot Program.'
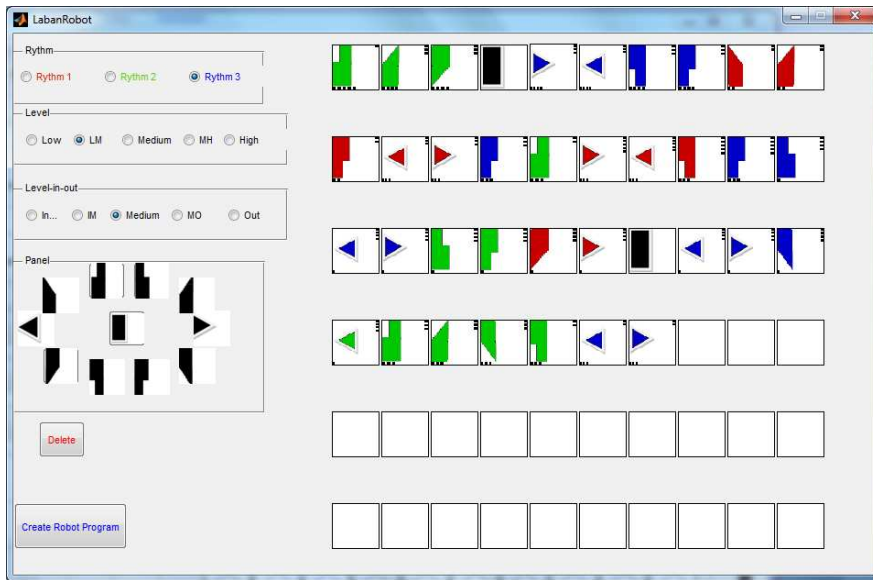
Figure 8
The user interface of LabanRobot

In Table 3, various heights and extensions are shown. The height of the end effector is designated by horizontal bars. The more bars, the higher the effector from the base. The extension is denoted by vertical bars. The more bars, the farther extended from the base.
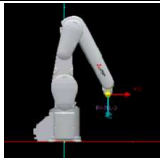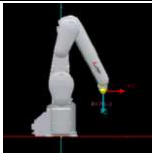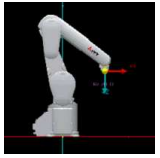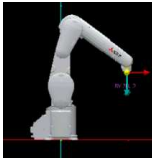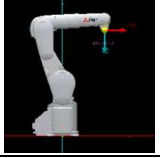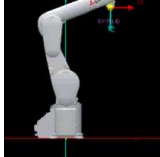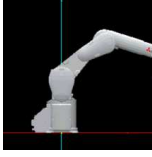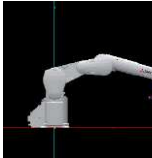
The color codes indicate speed. Red is high speed, which is 70% of the robot's maximum acceleration. Green is medium speed, which is 40% of its maximum acceleration. Blue is low speed, which is 20% of its maximum acceleration.

Table 4 shows an example dance sequence to accompany the song Mi Chica sang by Sarbel [34]. The sequence consists of 12 steps. For each step, there are three rows. The first row shows the simulation result, the second shows the dance with the humans, and the third uses the extended Labanotation proposed in this paper.

Conclusions

An extended Labanotation for an industrial robot arm was developed herein. The notation is simpler than the original Labanotation, yet very effective. The height and the extension information were added to the traditional notation to increase smoothness and effect. The color codes were added to simplify notation of rhythm. The new program, LabanRobot, has a user-friendly and easy-to-use interface, where a choreographer can design a complete dance sequence with no prior knowledge of the robot.

Table 3
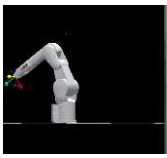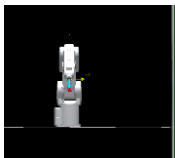Example for levels and extension levels

Height or level		Extension	
1 —		1 |	
2 =		2 ||	
3 ≡		3 |||	
4 ≣		4 ||||	
5 ≣		5 |||||	
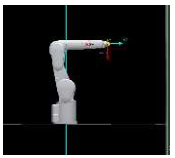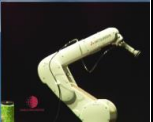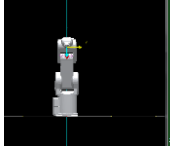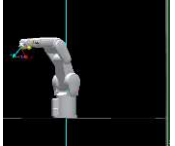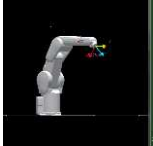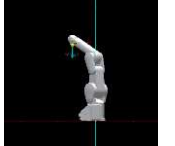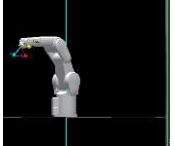
This research is unique in the sense that there is no other example of programming a dance for an industrial robot.

The algorithm was tested on an industrial robot on stage. The results were satisfactory. The next step of this research is to use the extended Labanotation on a multi-robot system and investigate human–robot interaction.

Acknowledgement

Table 4

Dance sequence for RV-7FL robot with the Mi Chica song sang by Sabel

References

[1]   E. Mirzabekiantz: Benesh Movement Notation for Humanoid Robots?, in Dance Notations and Robot Motion, Springer Tracts in Advanced Robotics, No. 111, J.-P. Laumond, N. Abe (eds), Switzerland, 2016, pp. 299-317

[2]   J. Challet-Haas: The Problem of Recording Human Motion, in Dance Notations and Robot Motion, Springer Tracts in Advanced Robotics, No. 111, J.-P. Laumond, N. Abe (eds), Switzerland, 2016, pp. 69-89

[3]   A. L. de Souza: Laban Movement Analysis-Scaffolding Human Movement, in Dance Notations and Robot Motion, Springer Tracts in Advanced Robotics, No. 111, J.-P. Laumond, N. Abe (eds), Switzerland, 2016, pp. 283-297

[4]   S. J. Burton, A.-A. Samadani, R. Gorbet and D. Kulić: Laban Movement Analysis and Affective Movement Generation for Robots and Other Near-Living Creatures, in Dance Notations and Robot Motion, Springer Tracts in Advanced Robotics, No. 111, J.-P. Laumond, N. Abe (eds), Switzerland, 2016, pp. 25-48

[5]   https://dance.osu.edu/research/dnb/laban-writer, accessed on March 10, 2016

[6]   T. Calvert: Approaches to the Representation of Human Movement: Notation, Animation and Motion Capture, in Dance Notations and Robot Motion, Springer Tracts in Advanced Robotics, No. 111, J.-P. Laumond, N. Abe (eds), Switzerland, 2016, pp. 49-68

[7]   A. La Viers, L. Bai, M. Bashiri, G. Heddy and Y. Sheng: Abstractions for Design-by-Humans of Heterogeneous Behaviours, in Dance Notations and Robot Motion, Springer Tracts in Advanced Robotics, No. 111, J.-P. Laumond, N. Abe (eds), Switzerland, 2016, pp. 237-262

[8]   H. Drewes: MovEngine-Developing a Movement Language for 3D Visualization and Composition of Dance, in Dance Notations and Robot Motion, Springer Tracts in Advanced Robotics, No. 111, J.-P. Laumond, N. Abe (eds), Switzerland, 2016, pp. 91-116

[9]   H. Peng, C. Zhou, H. Hu, F. Chao and J. Li: Robotic Dance in Social Robotics-A Taxonomy, IEEE Transactions on Human-Machine Systems, Vol. 45, No. 3, June 2015, pp. 281-293

[10]   P. Afsar, P. Cortez and H. Santos: Automatic Visual Detection of Human Behavior: A Review from 2000 to 2014, Expert Systems with Applications, Elsevier, Vol. 42, 2015, pp. 6935-6956

[11]   K. Hachimura, M. Nakamura, "Method of Generating Coded Description of Human Body Motion from Motion-captured Data," Proceedings of the IEEE International Workshop on Robot and Human Interactive Communication, 2001, pp. 122-127

[12]   K. Kojima, K. Hachimura and M. Nakamura: LabanEditor: Graphical Editor for Dance Notation, Proceedings of the 2002 IEEE International Workshop on Robot and Human Interactive Communication, Berlin, Germany, September 25-27, 2002, pp. 59-64

[13]   A. Nakamura, T. Niwayama, S. Tabata and Y. Kumo: Development of a Basic Dance Training System with Mobile Robots, Proceedings of the 2004 IEEE International Workshop on Robot and Human Interactive Communication, Kurashiki, Okayama, Japan, September 20-22, 2004, pp. 211-216

[14]   K. Suzuki and S. Hashimoto: Robotic Interface for Embodied Interaction via Dance and Musical Performance, Proceedings of the IEEE, Vol. 92, No. 4, April 2004, pp. 656-671

[15]   K. Kosuge and Y. Hirata: Human-Robot Interaction, Proceedings of the 2004 IEEE International Conference on Robotics and Biomimetics, August 22-26, 2004, Shenyang, China, pp. 8-11

[16]   Y. Choi, D. Kim, Y. Oh and B.-J. You: Posture/ Walking Control for Humanoid Robot Based on Kinematic Resolution of CoM Jacobian with Embedded Motion, IEEE Transactions on Robotics, Vol. 23, No. 6, December 2007, pp. 1285-1293

[17]   I.-W. Jeong, Y.-H. Seo and H. S. Yang: Effective Humanoid Motion Generation based on Programming-by-Demonstration Method for Entertainment Robotics, Proceedings of the 16th International Conference on Virtual Systems and Multimedia, Seoul, Korea, 2010, pp. 289-292

[18]   Y. Kekehashi, T. Izawa, T. Shirai, Y. Nakanishi, K. Okada and M. Inaba: The Trials of Hula Hooping by a Musculo-Skeletal Humanoid KOJIRO Nearing Dancing Using the Soft Spine, Proceedings of the 11th IEEE-RAS International Conference on Humanoid Robots, Bled, Slovenia, October 26-28, 2011, pp. 423-428

[19]   X. Dong, K. Wang, G. Wang, L. Yan and I-M. Chen: Design and Study of a Highly Articulated Mini Lion-Dance Robot, IEEE/ASME International Conference on Advanced Intelligent Mechatronics, Montréal, Canada, July 6-9, 2010, pp. 830-835

[20]   H. Wang, K. Kosuge: Control of a Robot Dancer for Enhancing Haptic Human-Robot Interaction in Waltz, IEEE Transactions on Haptics, Vol. 5, No. 3, July-September 2012, pp. 264-273

[21]   F. Auguliaro, A. Schoellig and R. D'Andrea: Dance of the Flying Machines, IEEE Robotics & Automation Magazine, December 2013, pp. 96-104

[22]   F. Zhou, F. De la Torre and J. K. Hodgins: Hierarchical Aligned Cluster Analysis for Temporal Clustering of Human Motion, IEEE Transactions on

Pattern Analysis and Machine Intelligence, Vol. 35, No. 3, March 2013, pp. 582-596

[23]   M.-B. Radac and R.-E. Precup: Optimal Behaviour Prediction Using a Primitive-Based Data-Driven Model-Free Iterative Learning Control Approach, Computers in Industry, Elsevier, Vol. 74, 2015, pp. 95-109

[24]   H. Wang and Q. Zou: B-Spline-Decomposition-Based Approach to Multiaxis Trajectory Tracking: Nanomanipulation Example, IEEE Transactions on Control Systems Technology, Vol. 22, No. 4, July 2014, pp. 1573-1580

[25]   F. A. Mussa-Ivaldi and S. A. Solla: Neural Primitives for Motion Control, IEEE Journal of Oceanic Engineering, Vol. 29, No. 3, July 2004, pp. 640-650

[26]   D. J. Grymin, C. B. Neas and M. Farhood: A Hierarchical Approach for Primitive-Based Motion Planning and Control of Autonomous Vehicles, Robotics and Autonomous Systems, Elsevier, Vol. 62, 2014, pp. 214-228

[27]   T. Okamoto, T. Shiratori, S. Kudoh, S. Nakaoka and K. Ikeuchi: Toward a Dancing RobotWith Listening Capability: Keypose-based Integration of Lower-, Middle-, and Upper-Body Motions for Varying Music Tempos, IEEE Transactions on Robotics, Vol 30, No. 3, June 2014, pp. 771-778

[28]   O. E. Ramos, N. Mansard, O. Stasse, C. Benazeth, S. Hak and L. Saab: Dancing Humanoid Robots, IEEE Robotics & Automation Magazine, December 2015, pp. 16-26

[29]   S. Yoshida, T. Shirokura, Y. Sigiura, D. Sakamato, T. Ono, M. Inami and T. Igarashi: RoboJockey: Designing an Entertainment Experience with Robots, IEEE Computer Graphics and Applications, January/February 2016, 62-69

[30]   W. Takano and Y. Nakamura: Real-time Unsupervised Segmentation of Human Whole-Body Motion and Its Application to Humanoid Robot Acquisition of Motion Symbols, Robotics and Automation Systems, Elsevier, Vol. 75, 2016, pp. 260-272

[31]   Mitsubishi Industrial Robot CR750-D/CR751-D/CR760-D Controller RV-4F-D/7F-D/13F-D/20F-D/35F-D/50F-D/70F-D   Series   Standard Specifications Manual, BFP-A8931-S, pp. 2-42, 2-44

[32]   M. Brandstötter, A. Angerer and M. Hofbaur: An Analytical Solution of the Inverse Kinematics Problem of Industrial Serial Manipulators with an Ortho-parallel Basis and a Spherical Wrist, Proceedings of the Austrian Robotics Workshop, Linz, Austria, 22-23 May, 2014, pp. 7-11

[33]   https://github.com/dtukel/RobotLaban

[34]   https://www.youtube.com/watch?v=8rsR-hljltA

Non-Destructive Analysis of Explanted Coronary Artery Stents

Dóra Károly[1], Lilla Asztalos[1], Tamás Micsík[2], Péter János Szabó[1]

[1] Department of Materials Science and Engineering, Faculty of Mechanical Engineering, Budapest University of Technology and Economics; Bertalan L. u. 7, H-1111 Budapest, Hungary; kdora@eik.bme.hu, lilla@eik.bme.hu, szpj@eik.bme.hu

[2] I. Department of Pathology and Experimental Cancer Research, Semmelweis University; Üllői út 26, H-1085 Budapest, Hungary; micsik.tamas@med.semmelweis-univ.hu

Abstract: Endovascular stents are used to effectively treat atherosclerosis during angioplasty. Nowadays, this surgery is performed on younger and younger patients, so the stent can be in the body for over 30 years. The stent cannot be removed from the patient, so the in vivo stability of the stent and the effects of the human organism on the stent can only be investigated on postmortem explanted stents. Therefore, this study deals with the non-destructive investigation of explanted coronary stents. Our goal was to determine the causes of visible damages and to measure the changes in the technical state.

Keywords: explanted coronary stent; non-destructive test; fracture; coating damage

1   Introduction

Over the last century, particularly in the last three decades, the occurrence of certain diseases in developed countries has increased significantly. These are known as civilization diseases, including a variety of cancers, cardiovascular diseases, hypertension and diabetes as well. These diseases occur mainly due to inappropriate lifestyle. The most important indicators are sedentary lifestyle, excessive stress, poor nutrition and a variety of addictions, e.g. alcohol and tobacco usage. Amongst these, ischemic heart disease (IHD) is the leading cause of death, accounting for around 11% of all deaths globally each year [1, 2].

The effective treatment of IHD is angioplasty. During the procedure usually a so called stent is also implanted into the vessel. A vascular stent is a small slotted

metal tube with a specific pattern. It is inserted into a vessel at the site of the narrowing (stenosis) to act as an internal scaffolding or support to the blood vessel and thus ensures the continuous flow of blood. Nowadays it is not uncommon that this surgery is performed on young and middle-aged patients, so the stent can be in the body for over 30 years [2].

Metal stents "last forever", they can't be removed from the body after implantation, therefore the investigation about the impact of the human organism to the stent has clinical relevance. Despite advances in polymer and drug technology, the metal alloy stent material remains a key determinant of clinical outcome [3].

During the life cycle of a stent, it has to withstand several mechanical loads. The stent is exposed to different kind of stresses during deployment, dilation and long-term usage. First the stent has to follow the vessel curvatures so it goes through a multiaxial bending. Due to the inflation of the balloon the stent has a uniform internal pressure and dilates. While expanding a stent we have to consider not only the tension effect of the balloon but also the impact of the plaque and the vessel curvature on the stent. In the body with every heart contraction the stent is exposed to the movement of the heart and the vessel. Assuming a 30 year long period, calculated with a heart rate of 60 beats per minute, nearly one billion contractions take place, so the stent bends the same number of times, which is a very high cycle fatigue load. The failure of the stent during deployment or usage (so during the useful life cycle) is clearly a negative impact on patients, so the discovery and the correction of any failures are very important [4, 5].

Most publications, which deal with stents after implantation, are from medical approach. Amongst these some pay attention on stent fractures. It has been already shown that the real occurrence of fractures are greater than reported in hospital records. One of the reasons can be the different interpretation of stent fracture (partial breakage or complete separation). Nakazawa et al. classified the fractures in five types; from one broken strut to the rupture of the stent. According to them, fractures occurred in 29% of all cases [6, 7].

There are only a few research papers dealing with explanted stents from the technical side. Halwani et al. investigated 9 excised vascular segment with 16 implanted stents. They found evidence of electrochemical and mechanically-induced corrosion on the stents. Vascular tissue surrounding corroded stents was shown to experience transfer of metallic elements [8]. These results, although coming from a limited number of specimens, encourage further investigation of the effect of in vivo corrosion on the structural integrity of stents.

This project covers the investigation of 13 coronary artery stents explanted from human cadavers (authorization number TUKEB 146/2012). All the explanted samples were subjected to several non-destructive tests (e.g. X-ray, CT) before the destructive ones. In our case non-destructive means that neither the deformations nor the material of the stents were damaged during the tests. Both bare metal

stents (BMS) and drug-eluting stent (DES) were explanted. The stents were excised atraumatically with the vessels, they were implanted in, to avoid any damage to the stents. During the preparation of the stents the aim was to find the fastest method to remove the vessels from the stents without damaging them (or with the least amount of damage). 20wt.% potassium hydroxide (KOH) was used for 7-9 days on the vessels to complete the dissolution [9].

Preparation process
- Explanted stent position in vessel (X-ray, CT)
- Effect of solvents on new stent (SEM, EDS)
- Dissolving carotid, coronary arteries (stereo microscope)
- Dissolving vessels from explanted stents (stereo microscope)

Non-destructive tests
➡ Deformations (stereo microscope)
➡ Fractures (stereo microscope, SEM)
➡ Material composition, smaller sized failures (SEM, EDS)
➡ Coating surface condition (SEM)
- Inner material failures (X-ray microscope)
- Surface roughness (AFM)

Destructive tests
- Coating adhesion, drug elution (in case of DES)
- Strut cross-section
- Metallographic analysis
- Radial force, radial strength
- Metallic surface area
- Flexibility
- Etc. (other functional properties)

Figure 1
Investigation process of explanted coronary stents (tests carried out are indicated by arrows)

This paper covers the non-destructive investigation of the samples after the appropriate preparation process. There isn't any standardized process which is dealing with the examination of explanted stents so an investigation protocol was specified. The location of these tests in the protocol is shown in Figure 1. Our aim was to determine the outer deformations and failures of the explanted stents, scan the material composition, define the fracture surfaces and the coating surface condition.

2 Materials and Methods

2.1 Preparation

The 9 vessel samples with 13 coronary artery stents were explanted at the I. Department of Pathology and Experimental Cancer Research, Budapest, Hungary. The vessels were 3-4 times longer than the stents inside them, to avoid any damage to the stents during explantation. Before the experiments, X-ray and computed tomography (CT) images were taken of the explanted stents to see if we cause any damage of the stents. The images were taken from several directions parallel and perpendicular to the longitudinal axis of the stents with a Dage XD6600 X-ray (tube voltage 100kV, tube power 0.90…0.95W) and a GE Phoenix

V\Tome\x M 300 CT. After the dissolution of the vessels from the stents, they were cleansed with ethanol to remove the solvent and any left vessel tissue. According to the input data, the X-ray images and the CT images, we could define the type of the explanted stents and give the main information about them.

2.2   Materials

More and more new stent materials appear including the promising bioabsorbable materials such as polylactic acid (PLA), but still physicians mostly use metal stents. Because of this and also because of the long-term follow-up of the previously implanted metal stents, our samples are made of the three most commonly used stent material; stainless steel (316LVM, X2CrNiMo18-14-3), cobalt-chromium alloy (L605, Co-Cr-W-Ni) and platinum-chromium alloyed steel (PtCr, Fe-Pt-Cr).

The summarized data about the investigated explanted stents that can be previously given is shown in Table 1.

Table 1
Previous information about the explanted stent samples

No.	Diameter* (mm)	Length* (mm)	Material	Type	Time spent in the body
(1)	3.50	12	PtCr	BMS	3 days
(2)	na	na	316LVM SS	DES	6 years
(3-1)	3.00	20	PtCr	BMS	3 days
(3-2)	2.50	13	L605 CoCr	BMS	3 days
(4)	2.50	20	PtCr	DES	2 years
(5)	4.00	12	PtCr	DES	1 year
(6-1)	na	na	L605 CoCr	DES	1 year
(6-2)	na	na	L605 CoCr	DES	1 year
(6-3)	na	na	Na	BMS	1 year
(7)	3.00	24	L605 CoCr	BMS	3 days
(8-1)	3.00	16	L605 CoCr	BMS	na
(8-2)	2.75	19	L605 CoCr	BMS	na
(9)	2.50	28	L605 CoCr	BMS	1 day

* size of stents according to manufacturer data

2.3   Methods

Several recordings were made of the available coronary stents from different angles with an Olympus SZX16 stereomicroscope. We examined the fractures and the deformations of the stents. On the samples, where the geometry is not significantly damaged, nominal diameter and length values were also measured to indicate the recoil of the stents inside the vessel. In the case of broken or damaged

stents the fracture surfaces were investigated with a Philips XL 30 scanning electron microscope (SEM). Images with both microscopes were also used to observe the quality of coating in the case of DES stents. EDAX Genesis and JEOL EX-54175JMU energy-dispersive X-ray spectroscopy (EDS) was used to determine the material composition of the samples (acc. volt. 15 keV, 20 keV).

3   Results

3.1   Stereomicroscopic Results

First of all, every stereomicroscopic images were compared to the X-ray images to see if any deformation happened during the dissolution of the vessels. We can say that every deformation occurred before the preparation process. The stereomicroscopic images about the 13 stents explanted from the 9 vessels can be seen in Figure 2.

In the case of sample no. (1), the stent is deformed due to a serious, punctual stenosis. The stent was taken on the curve of the vessel and the plaque layer. There are no fractures, but the stent spent only few days in the patient, so we cannot speak about long-term impacts.

In the case of sample no. (2), the stent is seriously deformed and fractured, separated to two pieces, some struts are broken off. This stent spent the most time in the body (6 years). In addition to the fractures, the stent preserved its diameter.

In the case of samples no. (3-1) and (3-2), the stents are used for bifurcation stenting (T-stent technique). Physicians have to use this bifurcation stenting if the narrowing involves a main vessel and a side branch as well. During this method (3-1) stent was implanted in the main vessel so that the side branch ostium is approximately in the middle of the stent. After that a (3-2) stent is placed and dilated into the branch vessel through the expanded cell of (3-1) stent in the main vessel. The angle between the stents is 65°. The materials of the stents are different (Co-Cr and Fe-Pt-Cr) which can be a problem with galvanic corrosion (electronegativity of the main elements: Co: 1.66, Cr: 1.88, Fe: 1.83, Pt: 2.28).

In the case of sample no. (4), a difference between the diameters (manufacturer's data and measured value) arises because the stent was dilated with 20 bar (instead of the 11 bar nominal pressure) according to the input data. Measurements also confirmed this.

In the case of sample no. (5), the stent is seriously deformed, which clearly shows that the stent has an open cell stent pattern. In this form the stent cannot properly carry out its function and in long-term the lumen of the vessel can decrease greatly.
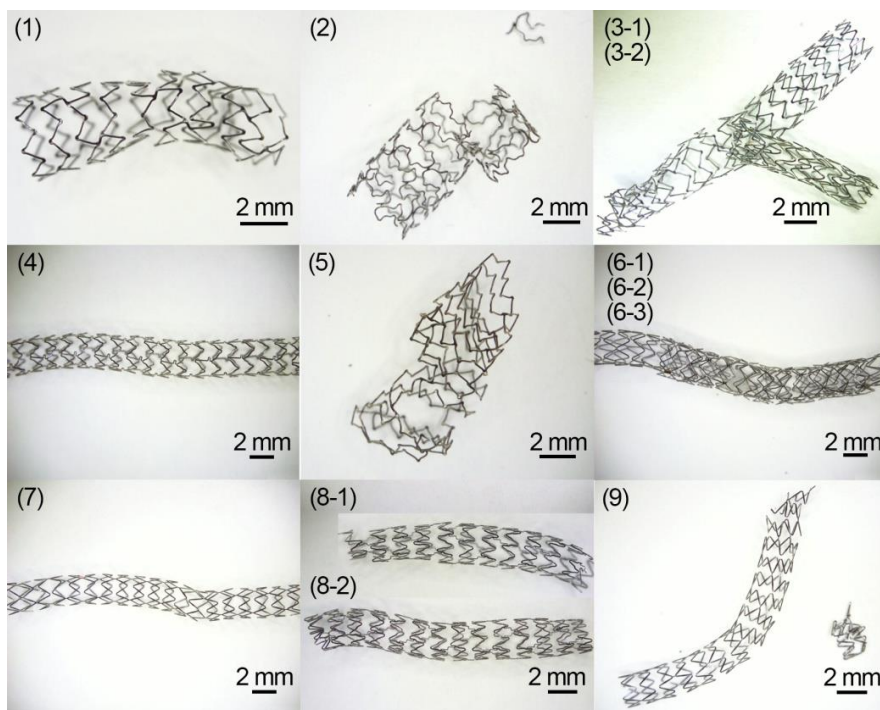
Figure 2
Stereomicroscope images about the explanted stents (with the numbered markings)

In the case of samples no. (6-1), (6-2) and (6-3), the three stents were overlapping in the vessel. (6-3) stent is over the other two stent and has markers to increase X-ray visibility. One of (6-3) stent's strut is broken, presumably due to the overexpansion of (6-1) and (6-2) stent. This fracture does not hamper the stent in filling its function. Some parts of the plaque is still between the struts because the struts are too dense and it could not be removed without damaging the stents. It should take into account from the difference of the stent pattern between stent no. (6-1) and (6-3) that the materials are also different.

In the case of sample no. (7), the stent is only mild deformed, but the stent spent only few days in the patient, so we cannot speak about long-term impacts.

In the case of samples no. (8-1) and (8-2), according to the description the stents have to be overlapping, but the X-ray images clearly show that they were not, the stereomicroscope images also indicate this. In the case of stent no. (8-1) some struts at the end of the stent are broken, which can be occurred during the explantation procedure. The image clearly shows that the diameters of the stents are similar, although the nominal diameters are different. The diameter decreased along the entire length of both stents. On average diameter of stent no. (8-1) reduced with 15%, diameter of stent no. (8-2) reduced with 10%.

In the case of sample no. (9), the stent is fractured to two separated parts and also seriously deformed at this part. As the stent spent a short time in the body, the fracture more likely the implication of a cutting during the explantation process. The broken struts are in a straight line which reinforces the idea that the stent was cut.

The summary of the stereomicroscope results can be seen in Table 2.

Table 2
Summary about quality of deformation and fractures of the explanted stents

No.	Deformation	Fractures
(1)	deformation in the middle, due to punctual plaque	No
(2)	serious deformation in the middle	several struts, in the middle
(3-1)	serious deformation at one end, bifurcation stenting (stent is in main vessel)	some struts, at the end
(3-2)	no deformation (stent is in side branch vessel)	no
(4)	no deformation	No
(5)	serious deformation	No
(6-1)	mild deformation, 3 stents overlapping	no
(6-2)	mild deformation	no
(6-3)	moderate deformation, over the other 2 stents	one strut, in the middle
(7)	mild deformation	No
(8-1)	mild deformation at the ends, 2 stents without overlapping	some struts, at the end
(8-2)	mild deformation at the ends	no
(9)	serious deformation at one end	some strut, at the end

3.2   SEM Results

Only the main results are shown according to SEM. The SEM images of stent no. (9) clearly show a cutting surface (Fig. 3 a-b). Based on the extensive regular planes and the grooves showing the edge of the cutting device, we can say that this damage occurred during the explantation process. In the case of stent no. (2) the SEM images show that two struts are facing each other (Fig. 3c). Fig. 3d shows the fracture surface of the stent no. (2). The classification of the surface is not clear. Ductile fracture can be ruled out, only brittle or fatigue fracture can be considered.

After the fracture of stent no. (2), the tip of the strut was exposed to abrasive impact, because the surface is smoothed on many places. The different crushes and small cracks on the side of the strut also indicates that the opposing struts were collided several times thus creating plastic deformation at the end of the strut.
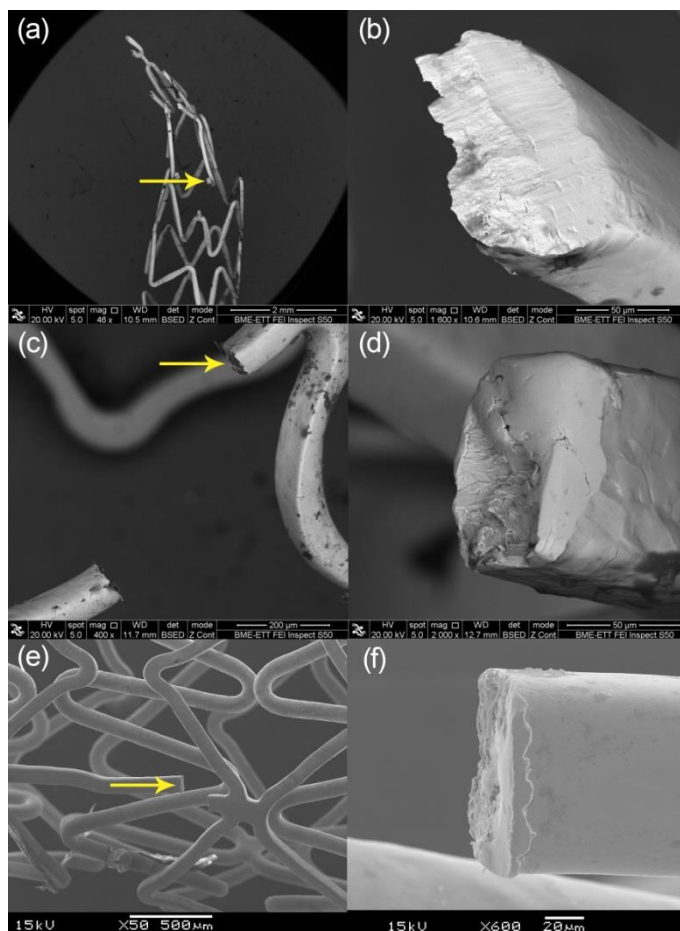
Figure 3

SEM image of a) stent no. (9) with marking the investigated strut, b) cut surface of (9) stent,
c) stent no. (2) with marking the investigated strut, d) fracture surface of (2) stent,
e) stent no. (6-3) overlapping on the other 2 stent, f) brittle fracture of (6-3) stent

Figure 3e and 3f show the fracture surface of stent no. (6-3), this is a brittle fracture most likely due to the over dilation.

After seeing the fractures of stents we thought it would be interesting to carry out a magnetic experiment. We used a simple permanent magnet to see if it attracts the stent or not, and the results show that most fractured stent are ferromagnetic. That is a huge problem if someone with an implanted stent needs radiology with magnetic resonance imaging (MRI), because the systems operate with a strong magnetic field (1-7 Tesla). The magnetic properties of fractured stents will be investigated with a vibrating sample magnetometer (VSM).

According to the available SEM images we cannot drawn exact conclusions about the causes of the damages in the coating. Damages can be occurred during the crimping process, the expansion, the time spent in the body or the preparation process of the stents. On average the detachment of coating is started at the tip of the struts (Fig. 4). Although stent no. (5) was in the body for 1 year, other stents which were exposed to the body only for 1-3 days show also the same coating damages at the tip of the struts. It means that the coating damaged in the beginning of the treatment, and that is a very serious problem. The function of the coating is to form a border surface between the "foreign material" and the tissue to decrease the probability of in-stent restenosis and thrombosis. If it is damaged the metal surfaces further induce the inflammatory reactions.
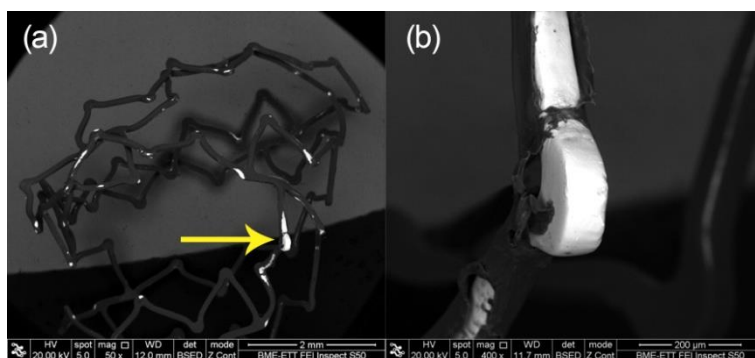
Figure 4
a) Coating of stent no. (5), b) detachment of coating at the tip of the strut

3.3   EDS Results

Our measurements show that there are no significant differences between the material standards (Table 3), the manufacturers' data and the measured material composition. The Fe-Pt-Cr alloy was developed by Boston Scientific, and currently only stents are made of this material, therefore there is no standardized composition for this material. We were interested especially in the quantity of Ni in the stent surface, because Ni is allergenic, but further investigation is needed to determine the aleatory Ni dissolution. In some cases we found a negligible amount of potassium, which is likely to be some residues remained from the vessel dissolution.

The material of stent no. (6-3) was unknown, we found that it is made of 316 LVM stainless steel, so our assumption is true that (6-3) stent is made of a different material than (6-1) and (6-2). The material of the markers is gold. Figure 5 shows the borderline of the marker and the base material, the assimilation of Au is satisfying and there aren't any signs of corrosion on the surface.

Table 3

Summary about material composition according to the standards and the manufacturer data

Alloy	ISO 5832-1	ISO 5832-5 Wt.%	Fe-Pt-Cr*
C	<0.03	-	0.003-0.023
Si	<1.00	0.40	<0.10
Mn	<2.00	1.00-2.00	<0.05
P	<0.025	0.40	<0.01
S	<0.01	-	-
N	<0.100	-	<0.01
Cr	17.00-19.00	19.00-21.00	17.50-18.50
Mo	2.25-3.00	-	2.43-2.83
Ni	13.00-15.00	9.00-11.00	8.50-9.50
Cu	<0.50	-	<0.02
W	-	14.00-16.00	-
Fe	x	3.00	x
Pt	-	-	32.50-33.50
Co	-	X	-

* manufacturer data
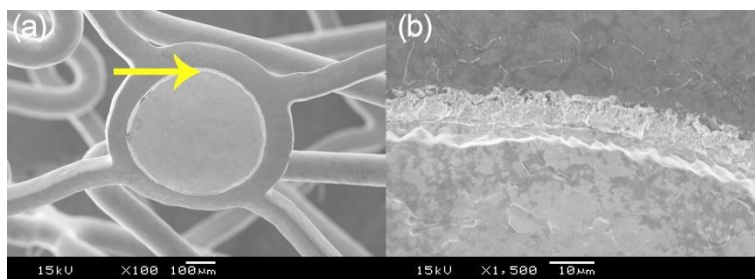
Figure 5

a) Gold marker on stent no. (6-3), b) borderline between the marker and the base material

Due to the position of the overlapping (6-1), (6-2), (6-3) and the bifurcation (3-1), (3-2) stents the examination of wear and corrosion can only be carried out by destructive tests (stent separation, maybe cutting).

Conclusions

During this series of experiments 13 explanted stents were investigated. Non-destructive test were performed; the outer deformations and failures of the stents were determined, the fracture surfaces were defined and the material composition was scanned. There is only limited information concerning the in vivo stability of explanted coronary artery stents, but still there are millions of metal stents implanted in people, therefore the investigation of explanted stents are well-

established and important. Future lines of research will clearly be the continuation of the investigation protocol, beside this testing the formation time and causes of the coating damage and exploring the magnetic properties of stents is also justified.

References

[1]   A. N. Nowbar, J. P. Howard, J. A. Finegold, P. Asaria, D. P. Francis, 2014 Global Geographic Analysis of Mortality from Ischaemic Heart Disease by Country, Age and Income: Statistics from World Health Organization and United Nations, Int. J. of Cardiol. 174:2 (2014) 293-298

[2]   D. L. Mann, D. P. Zipes, P. Libby, R. O. Bonow, Braunwald's Heart Disease: A Textbook of Cardiovascular Medicine. Ch. 1 Fundamentals of Cardiovascular Disease, Saunders/Elsevier, 10th edition, Philadelphia, 2014

[3]   E. Charpentier, A. Barna, L. Guillevin, J.-M. Juliard, Fully Bioresorbable Drug-Eluting Coronary Scaffolds: A Review. Arch. of Card. Dis. 108:6-7 (2015) 385-397

[4]   R. V. Marrey, R. Burgermeister, R. B. Grishaber, R. O. Ritchie, Fatigue and Life Prediction for Cobalt-Chromium Stents: A Fracture Mechanics Analysis. Biomaterials 27 (2006) 1988-2000

[5]   F. J. H. Gijsen, F. Migliavacca, S. Schievano et al, Simulation of Stent Deployment in a Realistic Human Coronary Artery. BioMed. Eng. OnLine 7:23 (2008)

[6]   D. Alexopoulos, I. Xanthopoulou, Coronary Stent Fracture: How Frequent It Is? Does It Matter? Hellenic J Cardiol 52 (2011) 1-5

[7]   G. Nakazawa, A.V. Finn, M. Vorpahl et al, Incidence and Predictors of Drug-Eluting Stent Fracture in Human Coronary Artery. J Am Coll Cardiol 54:21 (2009) 1924-1931

[8]   D. O. Halwani, P. G. Anderson, B. C. Brott et al, Surface Characterization of Explanted Endovascular Stents: Evidence of in VIVO corrosion. J Biomed Mater Res B Appl Biomater 95B:1 (2010) 225-238

[9]   D. Károly, D. Charalambous, B. Pogácsás, et al, Preparation of Explanted Coronary Stents for Investigation of Material Properties. Materials Today: Proceedings 3 (2016) 997-1002

Managing Big Data Using Fuzzy Sets by Directed Graph Node Similarity

Marko Jocic[*], **Endre Pap**[**,***], **Anikó Szakál**[**], **Djordje Obradovic**[*], **Zora Konjovic**[***]

[*]Faculty of Technical Sciences/Department of Computing and Automation, Trg DositejaObradovića 6, 21000 Novi Sad, Serbia

[**]Óbuda University, Bécsi út 96/b, H-1034 Budapest, Hungary

[***]Singidunum University, Danielova 29, 11000 Belgrade, Serbia

m.jocic@uns.ac.rs, epap@singidunum.ac.rs,.aniko@uni-obuda.hu, obrad@uns.ac.rs, zkonjovic@singidunum.ac.rs

Abstract: This paper proposes a novel algorithm for discovering similar nodes in very large directed graphs, with millions of nodes with billions of connections, which is based on the fuzzy set theory. The required input is a sample of representative nodes that are highly affiliated with some feature. This approach is practically verified on Twitter social network case study to discover influential Twitter users in the field of science.

Keywords: big data; directed graph; node similarity; Twitter

1 Introduction

The explosive growth of social media and massive participation in social networks is reflected in the countless number of contributions that are constantly posted and discussed on social sites such as Facebook, Twitter, Instagram, Pinterest and others. One aspect of special interest is the popularity and status of some members of these social networks measured by the level of attention they receive in terms of followers [1]. The other aspect is the influence that these individual's wield, which is determined by the propagation of their content through the network [2].

Despite an abundance of solutions providing influence measurement, there is still a need for improvements that cope well with the vagueness and uncertainty inherent to data describing influence, while keeping computational efficiency and accuracy of the output results.

In this paper we propose a novel algorithm aimed at discovering similar nodes, in very large directed graphs, that is both efficient and accurate with respect to real-world problems. The algorithm is based on the fuzzy set theory, and its practical application could be, among others, finding influential people in a network.

The paper is organized in the following way. Following this introductory section, the second section presents related works, while the third describes the proposed algorithm. Fourth section brings the Twitter social network case study discovering influential Twitter users in the field of science by which the algorithm is verified. The fifth section gives conclusions and directions for further research on this subject.

2   Related Works

This section presents some popular tools for determining the influence of social media users, and scientific papers using graphs as a model for social networks as well as for determining social network user similarities.

The tool *Socialbakers* [3] tracks, analyses, and benchmarks over 8 million social profiles across all the major social platforms including Facebook, Twitter, YouTube, LinkedIn, Instagram, Google+ and VK. They have statistics that are free of charge and available to everyone with daily updates and historical data up to 3 months. *Simply Measured* [4] is the leading social media analytics platform, providing complete measurement and reporting for serious marketers in all major social platforms including Facebook, Twitter, Google+, Instagram, YouTube, Vine, LinkedIn, and Tumblr. This service delivers profile analytics and audience insights including influence and sentiment analysis. It is available only in paid version, and only samples of data are shown for free. *Trackur* [5], which is also a paid service, allows full monitoring of all mainstream social media including Twitter, Facebook and Google+, but also news, blogs, reviews and forums. This service delivers executive insights including influence scoring. All results are delivered almost in real-time, as many sources are updated every 30 minutes. *Klout* [6] uses Twitter, Facebook, LinkedIn, Wikipedia, Instagram, Bing, Google+, Tumblr, Foursquare, YouTube, Blogger, WordPress, Last.fm, Yammer and Flickr data to create Klout user profiles that are assigned a "Klout Score" ranging from 1 to 100. Higher score corresponds to a higher ranking of the breadth and strength of one's online social influence. Klout is free for influencers, but paid for business users. What is common to all these services is that they are mostly paid, and only just outline algorithms and methods used to infer influence of a certain individual.

Graphs have been used a lot in order to describe social networks and analyze them. A myriad of techniques and approaches to social network analysis, data mining, graph mining and representation of social networks as graph structures

can be found in [7]. There are also papers about determining important network characteristics [8], [9] and graph properties [10] that have a potential for practical implementations in various domains including social networks. Similarity measures between objects are of central importance for various data analysis techniques including the special case of similarity measures related to graphs. A number of measures for such purpose have been proposed, especially for calculation of similarity of graphs nodes [11], [12], [13], [14], [15], [16], [20]. These methods have been successfully applied in several domains like ranking of query results [11], synonym extraction [13], database structure matching [17], construction of phylogenetic trees [12], analysis of social networks [18], [19], recommending trustworthy agents in a trust network [20], etc. There is a vast number of researches that try to determine the measure of influence in various social media and social networks [2], [21], [22]. This leads to many measures of influences, where some of these are correlated and some are not.

What is important for realistic modeling of influence in social media/networks is to capture the temporal dynamics (new influential users appearing over time, some users losing their influence over time), and the fact that attributes determining the measure of influence may be subjective, with values which are often imprecise or even vague. These imprecise data, in the context of social networks, and more specifically relations between users, their interests and influence can be modeled by fuzzy graphs [23], [24].

This encouraged the authors of this paper to adopt a fuzzy-like approach to the problem in order to model imprecise knowledge about influence within a social network.

3   Construction of the Algorithm

In this section we present the proposed algorithm (the basic one, and the one with reduced computational complexity) preceeded by preliminaries aimed mainly at introducing basic notions, and the notation that is used throughout the paper.

3.1   Preliminaries

We denote a directed graph by G , i.e. ordered pair $G = (N, E)$, where N is a set of all nodes in the graph, and E is a set of all edges in the graph. Edge $e = (a, b)$ is an ordered pair of two nodes and is directed from node a to node b, where a is the source node, and b is the terminating node. Another notion is that b is a direct successor of a, and a is said to be a direct predecessor of b. We denote by |N| the cardinality of a graph (the total number of nodes), and by |E| a graph size (the total number of edges/ connections between nodes). For each node in a directed

graph, there is an *in-degree* – the number of edges coming "into" certain node, and *out-degree* – the number of edges coming "out" of certain node. A *degree* is the sum of in-degree and out-degree. Individual elements of all sets will be denoted with lowercase letters corresponding to the name of the set, along with subscript index, e.g., set N has elements $n_i, i \in [1, |N|]$, $i \in \mathbb{N}$, where i represents an index of i-th element in set N. All direct predecessors of some node a are noted as the set DP(a). This means that the in-degree of node a can be represented as the cardinality of DP(a). Obviously, DP(a) $\subset$ N. Also, individual elements of DP(a) will be denoted by $dp_i(a)$, i.e.,

$$DP(a) = \{z | (z, a) \in E \; \wedge \; z, a \in N\} \tag{1}$$

$$indegree(a) = |DP(a)| \tag{2}$$

All direct successors of some node a are noted as a set $DS(a)$. This means that the out-degree of node a can be represented as a cardinality of $DS(a)$. Also, $DS(a) \subset N$. Individual elements of $DS(a)$ will be denoted $ds_i(a)$, i.e.,

$$DS(a) = \{z | (a, z) \in E \; \wedge \; z, a \in N\} \tag{3}$$

$$outdegree(a) = |DS(a)| \tag{4}$$

Following the fuzzy set theory [25], there is a fuzzy set described by its membership function μ_R, which here corresponds to the affiliation with some feature. Every node a in graph G has a value in this membership function $\mu_R(a)$,

$$\mu_R(a) \in [0,1] \wedge \mu_R(a) \in \mathbb{R}, a \in N \tag{5}$$

Here, a value of 0 means that no affiliation with certain feature exists, a value of 1 means utter affiliation with certain feature, and the values between 0 and 1 render partial association with a certain feature - the higher the value, the greater the affiliation. It can also be said that two nodes with similar values of the membership function for a particular feature, also exhibit some similarity between each other regarding that feature.

It is important to note that this directed graph G is large in a sense that it has a huge number of nodes and an even larger number of directed connections, where the number of nodes is at least a million and the number of connections is on the order of several billion. This introduces difficulties from the computational resources point of view. The proposed algorithm is designed to work with a (parts of) large directed graphs, where average *indegree* of nodes exceeds the average *outdegree* of nodes by several orders of magnitude.

3.2 Basic Algorithm

The algorithm starts by selecting representative nodes in a graph that have high association with some desired feature. These nodes should be selected by expert or experts in the particular domain in which this feature is exposed. This set of

representative nodes will be denoted as $R, R \subset N$, where each node in R will be denoted $r_i, i \in [1, |R|]$. Based on expert's knowledge, every selected node r_i is assigned its membership function value $\mu_R(r_i)$, which depicts its affiliation with the desired feature. Now, the procedure should be able to find nodes in a graph that are similar to the selected nodes – similar in their membership function values μ_R. The creation of this representative set and assigning membership function values can be regarded more formally as obtaining of a training set for a supervised machine learning algorithm. After selecting representative nodes, for each of these nodes, the algorithm obtains all its direct predecessors. The result is a union (set) of all direct predecessors for all representative nodes; this set will be denoted as DPR, and its elements will be denoted as dpr_i, i.e.,

$$DPR = \bigcup_{i=1}^{|R|} DP(r_i) \tag{6}$$

Many of the representative nodes have some common direct predecessors, which imply that many of these predecessors in DPR will have common direct successors among the representative nodes in R. As these direct successors in R are subset of all direct successors, a new set is introduced, $DSR(a) = DS(a) \cap R$. This intersection $DSR(a)$ results in only those direct successors of certain node a that are in representative set R, too. Now, for each of these nodes in a set DPR a value v is calculated by summing membership function values of nodes in R which are direct successors to a certain node:

$$\forall dpr_i \in DPR => v(dpr_i) = \sum_{dsr \in DSR(dpr_i)} \mu_R(dsr) \tag{7}$$

The calculated values are refined by taking in an account the out-degree for each node in DPR. More precisely, the calculated value v of a node in DPR is divided with its out-degree, resulting in a new value vr:

$$vr(dpr_i) = \frac{v(dpr_i)}{|DS(dpr_i)|} \tag{8}$$

When all nodes in DPR have had their values vr calculated, the algorithm can proceed to the next step. For each of the nodes in DPR, all its direct successors are obtained, i.e. the nodes that it is directly connected to. The result of this is a union (set) of all direct successors for all nodes in DPR. This set will be denoted as DSU, and its elements will be denoted as dsu_i, i.e.

$$DSU = \bigcup_{i=1}^{|DPR|} DS(dpr_i) \tag{9}$$

It can be noted here that DSU is a superset of R, $DSU \supset R$, because it will certainly contain all the representative nodes, but also other nodes that weren't marked as representative by the expert. After obtaining DSU, for each node in DSU, a value s is calculated by summing all of the already calculated values vr of its direct predecessors in DPR. This summed value can be explained as a similarity measure between certain node in DSU and nodes in R, i.e.,

$$\forall dsu_i \in DSU => s(dsu_i) = \sum_{dp \in DP(dsu_i)} vr(dp) \tag{10}$$

One final step must be taken in order to complete the calculation. If one takes two arbitrary nodes from DSU, where both of these have the same calculated value, but have different in-degree (number of all direct predecessors), how can that be interpreted? The node with smaller in-degree should expose more similarity with representative nodes in R than the node with larger in-degree. The explanation lies in absolute representation of this similarity, as it is just merely a sum, while some relative representation might contain more information. Thus, each of the calculated values is recalculated by dividing it with the node's in-degree, resulting in value sr, i.e.,

$$sr(dsu_i) = \frac{s(dsu_i)}{|DP(dsu_i)|} \tag{11}$$

First, the value sr is calculated for all of the representative nodes, in order to examine if calculated similarity values correspond to the originally assigned values by the expert. Here, it was noticed that the calculated values are actually smaller than the assigned values, which is a consequence of representative set being smaller than average in-degree of nodes in the representative set. In some cases, the calculated values are just two times smaller, but in larger graphs with lots of connections, this ratio might be a much larger value. In order to correct the calculation, new value is introduced - a correction coefficient c. This coefficient is actually a mean ratio between the assigned value and the calculated value, i.e.,

$$c = \frac{\sum_{r \in R} \mu_R(r)/sr(r)}{|R|} \tag{12}$$

The final measure of similarity denoted by $srf(a)$, or the value of membership function is then calculated by multiplying the calculated value sr with the correction coefficient c, i.e., $srf(a) = c * sr(a)$. The resulting value srf is basically membership function value for each node in DSU, where nodes that have the smallest values of this number shouldn't manifest similarity with the feature exposed among nodes in R, and vice versa. The calculated similarity measure is actually a value of membership function that represents affiliation with the nodes in the representative set $\mu_R(a) = srf(a)$. Now the expert can select the nodes from DSU with the highest calculated similarity and examine if the desired feature is present, as it was in representative set R.

The above described algorithm is more concisely represented by Pseudo-code 1.

<div align="center">
Pseudo-code 1:

Basic algorithm
</div>

```
choose set of representative nodes
foreach node r in R:
   assign membership function value for r

init DPR = empty set
```

```
foreach node r in R:                    (equation 6)
  DP = fetch_direct_predecessors(r)
  DPR = find_union(DP, DPR)

foreach node dpr in DPR:        (equations 7 and 8)
  DS = fetch_direct_successors(dpr)
DSR = intersection(DS, R)
  init v = 0
  foreach node dsr in DSR:
    v = v + membership(dsr, R)
  vr(dpr) = v / size(DS)
  init DSU = empty set
foreach node dpr in DPR:                (equation 9)
  DS = fetch_direct_successors(dpr)
  DSU = find_union(DS, DSU)
```

Pseudo-code 1:

Basic algorithm (continuation)

```
foreach node dsu in DSU:      (equations 10 and 11)
  DP = fetch_direct_predecessors(r)
  init s = 0
  foreach node dp in DP:
    s = s + vr(dp)
  sr(dsu) = s / size(DP)

init c = 0
foreach node r in R:                    (equation 12)
  c = c + sr(r)
c = c / size(R)

foreach node dsu in DSU:
  srf(dsu) = c * sr(dsu)
```

3.3   Algorithm with Reduced Computational Complexity

Computational complexity is an important issue of any implementation that intends to reach performance satisfying real-life demands. The algorithm that is described in previous subsection is not exception to this. This means that performance indicators on quality characteristics like precision should be balanced with performance indicators like resource's consumptions.

There are few things that one needs to consider while determining computational complexity of the proposed algorithm.

- First, the running time of a union operation in equation (6) depends on sum of in-degree values for each node in representative set, thus the running time of this part is $O(\sum_{r \in R} indegree(r)) = O(|DPR|)$.

- Second, running time of calculating values $v(dpr_i), vr(dpr_i)$ from equations (7) and (8), respectively, depends on a number of nodes in DPR set, $n_{DPR} = |DPR|$, and number of nodes in DSR set, $n_{DSR} = |DSR|$, which results in running time for this part of $O(n_{DPR} * n_{DSR})$. But since $DSR \subset R \Rightarrow |DSR| < |R|$, and $|R| \ll |DSR|$, running time can be approximated as $O(n_{DPR}) = O(|DPR|)$

- Third, the running time of the union operation in equation (9) depends on the sum of out-degree values for each node in DPR set, which makes running time of this part $O(\sum_{dpr \in DPR} outdegree(dpr))$. Again, since this algorithm works with parts of graphs which usually have in-degree values with at least two orders of magnitude larger than out-degree values, running time for this part can be approximated as $O(|DPR|)$ as well.

- Fourth, running time of calculating values $s(dsu_i), sr(dsu_i)$ from equations (10) and (11), respectively, depends on total number of nodes in DSU set, and total number of direct predecessors for each node in DSU set. But, because this part of the algorithm iterates through direct predecessors that are already in the DPR set, one can say that running time of this part is also O(|DPR|).

Finally, by summing all these running times, we get $O(4 * |DPR|) = O(|DPR|)$, which leads to the conclusion that computational complexity of the entire algorithm depends mostly on the number of taken direct predecessors of nodes in the representative set.

Since the defined (part of) large directed graph G typically has nodes with high in-degree values, this makes the whole approach computationally expensive. This indicates that, in order to reduce the running time, one could try to take smaller portion of these direct predecessors, i.e. set $SDPR \subset DPR$, which contains randomly sampled elements of DPR. Here the cardinality of $SDPR$ is defined as $|SDPR| = \rho * |DPR|, \rho \in [0,1]$, and parameter ρ denotes the portion of randomly sampled elements.

By adopting this approach in the paper we created an algorithm modification, which is presented by the Pseudo-code 2 where the bolded content is for modifications.

<div align="center">Pseudo-code 2:</div>

<div align="center">Algorithm modification for reduction of computational complexity</div>

```
choose set of representative nodes

foreach node r in R:
  assign membership function value for r
```

```
init DPR = empty set
foreach node r in R:
  DP = fetch_direct_predecessors(r)
  DPR = find_union(DP, DPR)

SDPR = random_sample(DPR, rho)

foreach node dpr in SDPR:
DS = fetch_direct_successors(dpr)
  DSR = intersection(DS, R)
  init v = 0
  foreach node dsr in DSR:
    v = v + membership(dsr, R)
  vr(dpr) = v / size(DS)
init DSU = empty set
```

Pseudo-code 2:

Algorithm modification for reduction of computational complexity (continuation)

```
foreach node dpr in SDPR:
  DS = fetch_direct_successors(dpr)
  DSU = find_union(DS, DSU)

foreach node dsu in DSU:
  DP = fetch_direct_predecessors(r)
  init s = 0
  foreach node dp in DP:
    s = s + vr(dp)
  sr(dsu) = s / size(DP)

init c = 0
foreach node r in R:
  c = c + sr(r)
c = c / size(R)

foreach node dsu in DSU:
  srf(dsu) = c * sr(dsu)
```

4   Twitter Case Study

In this section we present a case study aimed at demonstrating the algorithm, and to show that the proposed modification of the basic algorithm performs better in terms of resources consumption, retaining, at the same time, acceptable precision.

Since the proposed algorithm was verified on the Twitter social network, in the first subsection of this section some notions are given about this social network and its data. The second subsection presents application of our algorithm to Twitter.

4.1   Twitter and Its Data

Twitter is one of the biggest social networks with more than 300 million active users per month, which makes it second largest social network, after Facebook. Twitter users follow others or are followed. Unlike on most online social networking sites, such as Facebook or MySpace, the relationship of following and being followed requires no reciprocation. A user can follow any other user, and the user being followed doesn't need to follow back. Being a follower on Twitter means that the user receives all the messages (called tweets) from those the user follows. Common practice of responding to a tweet has evolved into well-defined markup culture: RT stands for retweet, '@' followed by a user identifier addresses the user, and '#' followed by a word represents a hashtag. This well-defined markup vocabulary combined with a strict limit of 140 characters per posting conveniences users with brevity in expression. The retweet mechanism empowers users to spread information of their choice beyond the reach of the original tweet's followers [26].

Twitter, and social networks in general, can easily be represented as a graph, where Twitter is an example of a directed graph.

Twitter REST API provides programmatic access to read and write Twitter data [27]. The REST API identifies Twitter applications and users using OAuth; responses are available in JSON. In order to perform the proposed algorithm, one should be able to obtain: profiles of certain Twitter users (profile contains basic information, as well as the number of followers and number of following), followers of certain user, and users that certain user is following. All of these data are obtained via Twitter REST API. The base URI for all Twitter REST API calls is https://api.twitter.com/1.1.

It is also important to mention that the calls to REST API are rate limited – all methods allow only a limited number of calls within a 15 minute window. Rate limiting is primarily considered on a per-user basis, or more accurately, per access token in your control. If a method allows for 15 requests per rate limit window, then it allows you to make 15 requests per window per leveraged access token [28]. This rate limiting raises practical difficulties when trying to download big

amounts of data from Twitter because it simply takes a lot of time – one should carefully orchestrate calls to the REST API within these 15 minutes windows in order to make the most use of Twitter REST API and its data.

For fetching Twitter users' profiles, HTTP request GET https://api.twitter.com/1.1/users/lookup is called, which returns fully-hydrated user objects for up to 100 users per request, as specified by comma-separated values passed to the user_id and/or screen_name parameters. Rate limit for this endpoint is 180 calls per 15 minute window.

In order to obtain followers of a certain Twitter user, an HTTP request GET https://api.twitter.com/1.1/followers/ids is called, which returns a cursored collection of user IDs for every user following the specified user. Results are given in groups of 5,000 user IDs and multiple "pages" of results can be navigated through using the next_cursor value in subsequent requests. Rate limit for this endpoint is 15 calls per 15 minute window.

4.2 Discovering Similar Influential Users on Twitter

The proposed algorithm was used to discover similar influential users in Twitter social network.

The process starts with selecting a group of representative users. In this experiment a group of 163 Twitter users was selected, based on their influence in the category of science. Influential science users engage their followers with news and interactive tweeting in many spheres of science. For each of these users, an expert assigns a membership function value, which determines how much certain user belongs to this representative group.

Some users of this representative group of Twitter users influential in science domain are shown in the Table 1.

Table 1
A sample of influential Twitter users in science

Name	Twitter screen name	Membership function value
NASA	@nasa	0.9
Scientific American	@sciam	0.85
New Scientist	@newscientist	0.85
WIRED Science	@wiredscience	0.7
Neil deGrasse Tyson	@neiltyson	0.8
CERN	@cern	0.9
National Geographic	@natgeo	0.7
Curiosity Rover	@marscuriosity	0.8
Phil Plait	@badastronomer	0.85
Richard Dawkins	@richarddawkins	0.7

After selecting a group of representative users, all their followers must be downloaded from the Twitter REST API. This is the most resource (time and storage) consuming part of the whole process, as most of these users have more than 1 million followers, while some of them have even more than 10 million followers. This implies that for most users, in order to get all of their followers, the Twitter REST API must be called at least 200 times (200 x 5K followers = 1M followers). Bearing in mind that for fetching followers only 15 requests per 15 minute window is allowed by the Twitter REST API, the estimated needed time to download 1 million followers with one Twitter access token is 3 hours and 20 minutes. Also, each request returns 5,000 IDs of followers, where ID is a 64-bit number. This results in need of at least 8B x 1M followers = 8,000,000 B ~ 7,5MB of storage for single user's followers, and that is only for users with 1 million followers.

Note that needed resources (time and storage) increase linearly with number of users' followers. However, this still imposes technical and practical difficulties on huge graphs with a large number of connections, like the Twitter social network. It is also important to note that lots of these followers are not unique for certain representative user. It is very likely that two Twitter users that promote science and that have more than 1 million followers will have some (or many) common followers. This implies that many of these followers will follow more than just one user from the representative group.

After downloading all the representative users' followers, for each of these followers a relative value is calculated that describes his/her "interest" in this representative group of users, in this case specifically – in science. The way this is done is trying to utilize relevant data (the number of users that promote science followed by a certain user as well as representative users' membership functions), and, at the same time, neutralize what's called *aggressive following*[1] on Twitter.

In order to calculate this value, the total number of friends (users that a certain user follows) has to be downloaded. This is done by obtaining the whole Twitter profile via Twitter REST API for all followers, but storing just total number of friends. After obtaining the total number of friends, a value that should give more insight about certain user interests is calculated.

Following the calculation of followers' interest in the representative group, by examining the data it can be seen that many of the users with the highest calculated value of interest really are engaged a lot in the Twitter scientific community, which means they often tweet, retweet and favorite scientific articles, facts, news, etc.

When all followers are downloaded and their values of interest are calculated, the process proceeds with downloading all of the followers' friends – users that

---

[1] *Aggressive following* is defined as indiscriminately following hundreds of accounts just to garner attention.

followers follow. Unlike downloading followers, downloading friends is much less resource consuming because according to Twitter, every account can follow 2,000 users in total. Once someone has followed 2,000 users, there are limits to the number of additional users can follow. This number is different for each account and is based on user's ratio of followers to following; this ratio is not published. Follow limits cannot be lifted by Twitter and everyone is subject to these limits, even high profile and API accounts. That said, just one request to the Twitter REST API should suffice in order to obtain certain user's friends.

When the process of downloading followers' friends is complete, for each of these friends a value is calculated by summing all values of interest already calculated for their followers. This new value could be interpreted as how much certain user's followers are interested in science, or more generally whatever feature the representative group is affiliated to. We call this value *absolute similarity*. Dividing this absolute similarity by the number of followers results in a kind of "per follower influence", which we call *relative similarity*.

Now, for each of the representative users, the calculated value of similarity can be compared to the originally assigned value of membership function. What occurred in our experiment was that the calculated values were much smaller than the assigned values, with mean ratio between the assigned and the calculated value $c = 2.57$. Therefore, the final similarity value is then calculated by multiplying the previously calculated relative similarity by the correction coefficient c. This final similarity is in regards to Twitter users in the representative group, or more specifically influential Twitter users in the field of science.

Here the algorithm stops and the expert inspects the calculated data most probably by looking at the users with the highest calculated value of influence.

Of course, Twitter users that were not in the representative group are especially interesting for examination because they are new influential Twitter users not known as such previously. In addition to discovering new users, the results might show that certain Twitter users that were originally put in the representative group are actually not that influential at all. This gives expert an opportunity to review and revise the data (for example, newly discovered influential users can be added to the representative group) and the algorithm can be re-run hopefully yielding even more discovery.

The results of discovering similar influential Twitter users in science show that the proposed algorithm indeed works as intended.

When provided with 163 representative influential Twitter users in the category of science, a total of 72 new users were discovered with a similarity measure above 0.5.

Some discovered users are shown in Table 2.

Table 2

A sample of discovered influential Twitter users in science

Name	Twitter screen name	Description	Similarity
Robert Simpson	@orbitingfrog	Astronomer	0.84
Stuart Clark	@drstuclark	Astronomer, journalist	0.71
CaSE	@sciencecampaign	The campaign for Science & Engineering	0.73
Chemistry World	@chemistryworld	Chemistry magazine	0.6
Armed with Science	@armedwscience	US Defense Depart. science and technology blog	0.68
NASA's Juno Mission	@nasajuno	NASA's mission to Jupiter	0.78
ESA Science	@esascience	European Space Agency science blog	0.81

Table 3 shows a comparison of the calculated values of membership function of the representative users with those assigned by the expert. Many of the calculated values don't deviate much from the assigned values. The root error of the mean square error for all 163 representative users is 0.09.

Table 3

A comparison between assigned and calculated value of membership function of sampled

representative users

Name	Twitter screen name	Assigned value	Calculated value
NASA	@nasa	0.9	0.82
Scientific American	@sciam	0.85	0.71
New Scientist	@newscientist	0.85	0.82
WIRED Science	@wiredscience	0.7	0.77
Neil deGrasse Tyson	@neiltyson	0.8	0.84
CERN	@cern	0.9	0.79
National Geographic	@natgeo	0.7	0.66
Curiosity Rover	@marscuriosity	0.8	0.69
Phil Plait	@badastronomer	0.85	0.69
Richard Dawkins	@richarddawkins	0.7	0.83

4.3   Performance Improvement

The previous subsection has shown that it is possible to find similar Twitter users by feeding the proposed algorithm with a representative group of Twitter users assuming download of all her/his followers. This section attempts to find out how

many followers per Twitter user are enough to be downloaded, but still be able to get results comparable with those when all followers were downloaded.

Let's say that the entire process of discovering 72 users with a similarity measure above 0.5 from 163 representative users in the previous experiment was performed in some reference time *T*. One could expect that, by taking a smaller portion of all followers, a smaller number of users with a similarity measure bigger than 0.5 will be found. So, let's say that the discovered 72 users with a similarity measure bigger than 0.5 is the highest precision possible with the provided data, because all followers are taken, and this number of 72 users is considered as a reference value for precision, which will be denoted *D*. Hence, the precision when all followers are taken is *p* = 1, when *D* = 72.

In order to test if the algorithm performs well even if fewer followers are taken, it was run again by downloading 50%, 25% and 10% of all followers for each representative Twitter user. This portion of followers taken is denoted as ρ, where $\rho \in [0,1] \land \rho \in \mathbb{R}$. However, it is important to note that by taking only a portion of followers, the calculated similarity measure is expected to be proportionally smaller. Therefore the calculated similarity is multiplied by corresponding multiplier, i.e.:

$$\mu_R(dsu_i) = \frac{sr(dsu_i)}{\rho} \tag{13}$$

Table 4 shows the number of discovered users with similarity measure above 0.5, the precision relative to the highest precision possible, the time needed for the whole process, precision to time ratio and the root mean square error (RMSE) for the representative users.

Table 4
Algorithm performance improvement results comparison

Portion of followers taken (ρ)	Discovered users (D)	Precision (p)	Time (T)	p/T	Correction coefficient (c)	RMSE
1	72	1	1	1	2.57	0.09
0.5	67	0.93	0.54	1.72	6.13	0.11
0.25	58	0.8	0.28	2.86	12.33	0.13
0.1	51	0.71	0.12	5.92	27.65	0.18

As expected, even by taking less followers, comparable results are obtained in much less time. By taking 10 times less followers (10%), the algorithm discovered 71% of the users that were discovered by taking all followers, but consuming almost 10 times fewer resources. The root mean square error was only doubled by taking 10% of the followers.

The main advantage of running the algorithm by taking less number of followers is, of course, the less use of resources, primarily time. This allows an expert to run the process by taking just 10% of followers, thus saving almost 90% time, but sacrificing only 30% precision. Also, the discovered users can be added to the representative set and then the algorithm can run again iteratively, discovering even more users.

Conclusions

This paper proposes a novel algorithm for finding similar nodes in directed graphs. The algorithm needs to be provided with the representative set of nodes that expose some feature which should be discovered among other nodes in graph. The representative set of nodes is described by values of their membership function, which corresponds to the affiliation of the node with the desired feature. The calculated measure of similarity with the representative nodes correlates to the true value of the membership function. The algorithm is verified on the Twitter social network case study by discovering influential Twitter users in the field of science. A set of 163 representative influential Twitter users is fed to the algorithm, which results in discovery of new 72 influential users in science.

Also, a proposal is made aimed at improving the algorithm's efficiency in terms of time and storage complexity, which relies upon assumption that many of the members in the representative group share many common followers (which evidently holds for the presented case study). The preliminary results indicate that by reducing the number of downloaded followers to only 10% of the original set the algorithm yielded comparable results with a tolerable increase in error. Bearing in mind that influence in social media is very dynamic and vague concept, the algorithm can be refined by a series of measures: user popularity (number of followers), user activity (number of tweets), followers engagement on user activity (number of retweets, favorites), combination of user popularity and activity, etc. Also, each of these measures change over time, and these changes could be tracked in order to provide some valuable trend information. By taking some of these measures into account to certain Twitter user's influence, it is reasonable to assume that the whole algorithm could yield better results.

References

[1]   D. M. Romero, W. Galuba, S. Asur, and B. A. Huberman, "Influence and Passivity in Social Media," in Machine Learning and Knowledge Discovery in Databases, Springer, 2011, pp. 18-33

[2]   M. Cha, H. Haddadi, F. Benevenuto, and P. K. Gummadi, "Measuring User Influence in Twitter: The Million Follower Fallacy," ICWSM, Vol. 10, pp. 10-17, 2010

[3]   "Social Media Marketing, Statistics & Monitoring Tools," Socialbakers.com. [Online]. Available: http://www.socialbakers.com/. [Accessed: 28-Dec-2014]

[4]   "Simply Measured | Easy Social Media Measurement & Analytics," Simply Measured. [Online]. Available: http://simplymeasured.com/. [Accessed: 28-Dec-2014]

[5]   "Social Media Monitoring Tools & Sentiment Analysis Software," Trackur. [Online]. Available: http://www.trackur.com/. [Accessed: 28-Dec-2014]

[6]   "Klout | Be Known For What You Love," Klout. [Online]. Available: https://klout.com/home. [Accessed: 28-Dec-2014]

[7]   D. F. Nettleton, "Data mining of social networks represented as graphs," Comput. Sci. Rev., Vol. 7, pp. 1-34, Feb. 2013

[8]   A. Rusinowska, R. Berghammer, H. De Swart, M. Grabisch, "Social Networks: Prestige, Centrality, and Influence (Invited paper)". de Swart. RAMICS 2011, Springer, pp.22-39, 2011, Lecture Notes in Computer Science (LNCS) 6663. <hal-00633859>

[9]   V. Halasz, L. Hegedus, I. Hornyak, B. Nagy: Solving Application Oriented Graph Theoretical Problems with DNA Computing. In Proceedings of Seventh International Conference on Bio-Inspired Computing: Theories and Applications (BIC-TA 2012), AISC 201 (Springer) 2012, pp. 75-85

[10]   T. Réti, I. Felde "On Some Properties of Pseudo-Semiregular Graphs," Acta Polytechnica Hungarica, Vol. 13, No. 6, pp. 45-65, 2016

[11]   J. Kleinberg, "Authoritative Sources in a Hyperlinked Environment," J. ACM JACM, Vol. 46, No. 5, pp. 604-632, 1999

[12]   M. Heymans and A. K. Singh, "Deriving Phylogenetic Trees from the Similarity Analysis of Metabolic Pathways," Bioinforma. Oxf. Engl., Vol. 19 Suppl 1, pp. i138-146, 2003

[13]   V. D. Blondel, A. Gajardo, M. Heymans, P. Senellart, and P. Van Dooren, "A Measure of Similarity between Graph Vertices: Applications to Synonym Extraction and Web Searching," SIAM Rev., Vol. 46, No. 4, pp. 647-666, Jan. 2004

[14]   L. A. Zager and G. C. Verghese, "Graph Similarity Scoring and Matching," Appl. Math. Lett., Vol. 21, No. 1, pp. 86-94, Jan. 2008

[15]   M. Nikolic, "Measuring Similarity of Graph Nodes by Neighbor Matching," Intell. Data Anal., Vol. 16, No. 6, pp. 865-878, 2012

[16]   L. Kovács, G. Szabó, "Conceptualization with Incremental Bron-Kerbosch Algorithm in Big Data Architecture" Acta Polytechnica Hungarica, Vol. 13, No 2. pp. 139-158, 2016

[17]   S. Melnik, H. Garcia-Molina, and E. Rahm, "Similarity Flooding: A Versatile Graph Matching Algorithm and its Application to Schema Matching," in Data Engineering, 2002. Proceedings. 18th International Conference on, 2002, pp. 117-128

[18]   E. A. Leicht, P. Holme, and M. E. Newman, "Vertex Similarity in Networks," Phys. Rev. E, Vol. 73, No. 2, p. 26120, 2006

[19]   A. Anderson, D. Huttenlocher, J. Kleinberg, and J. Leskovec, "Effects of User Similarity in Social Media," in Proceedings of the fifth ACM international conference on Web search and data mining, 2012, pp. 703-712

[20]   C.-W. Hang and M. P. Singh, "Trust-based Recommendation Based on Graph Similarity," in Proceedings of the 13th International Workshop on Trust in Agent Societies (TRUST). Toronto, Canada, 2010. [Online]. Available: https://www.csc2.ncsu.edu/faculty/mpsingh/papers/mas/aamas-trust-10-graph.pdf

[21]   M. Gomez-Rodriguez, J. Leskovec, and A. Krause, "Inferring Networks of Diffusion and Influence," ACM Trans. Knowl. Discov. Data, Vol. 5, No. 4, pp. 1-37, Feb. 2012

[22]   J. Leskovec, A. Singh, and J. Kleinberg, "Patterns of Influence in a Recommendation Network," in Advances in Knowledge Discovery and Data Mining, Springer, 2006, pp. 380-389

[23]   M. S. Sunitha and S. Mathew, "Fuzzy Graph Theory: a Survey," Ann. Pure Appl. Math., Vol. 4, No. 1, pp. 92-110, 2013

[24]   J. Vascak, L. Madarasz, "Adaption of Fuzzy Cognitive Maps - a Comparison Study" Acta Polytechnica Hungarica, Vol. 7, No. 3, pp. 109-122, 2010

[25]   L. A. Zadeh, "Fuzzy Sets," Inf. Control, Vol. 8, pp. 338-353, 1965

[26]   H. Kwak, C. Lee, H. Park, and S. Moon, "What is Twitter, a Social Network or a News Media?," in Proceedings of the 19th International Conference on World Wide Web, 2010, pp. 591-600

[27]   "REST APIs," Twitter Developers. [Online]. Available: https://dev.twitter.com/rest/public. [Accessed: 26-Aug-2015]

[28]   "API Rate Limits," Twitter Developers. [Online]. Available: https://dev.twitter.com/rest/public/rate-limiting. [Accessed: 26-Aug-2015]

Reproducibility Analysis of Scientific Workflows

Anna Bánáti[3], Péter Kacsuk[1,2], Miklós Kozlovszky[1,3]

[1] MTA SZTAKI, H-1518 Budapest, Pf. 63., Hungary

[2] University of Westminster, 115 New Cavendish Street, London W1W 6UW

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

peter.kacsuk@sztaki.mta.hu, {banati.anna,kozlovszky.miklos}@nik.uni-obuda.hu)

Abstract: Scientific workflows are efficient tools for specifying and automating compute and data intensive in-silico experiments. An important challenge related to their usage is their reproducibility. In order to make it reproducible, many factors have to be investigated which can influence and even prevent this process: the missing descriptions and samples; the missing provenance data about the environmental parameters and the data dependencies; the dependencies of executions which are based on special hardware, changing or volatile third party services or random generated values. Some of these factors (called dependencies) can be eliminated by careful design or by huge resource usage but most of them cannot be bypassed. Our investigation deals with the critical dependencies of execution. In this paper we set up a mathematical model to evaluate the results of the workflow in addition we provide a mechanism to make the workflow reproducible based on provenance data and statistical tools.

Keywords: scientific workflows; reproducibility; analytical model; provenance; evaluation; gUSE

1 Introduction

In large computational challenges scientific workflows have emerged as a widely accepted solution for performing in-silico experiments. In general, these in-silico experiments consist of series of particularly data and compute intensive jobs and in most cases their executions require parallel and distributed infrastructure (supercomputers, grids, clusters, clouds). The complexity of workflows and the continuously changing nature of the environment make it hard or even prevent to reproduce or share the results in the scientist's community. The different users for different purposes may be interested in reproducing the scientific workflow

(SWf). The scientists have to prove its results, other scientists would like to reuse the results and reviewers intend to verify the correctness of the results [13]. A reproducible workflows can be shared in repositories and can become useful building blocks that can be reused, combined or modified for developing new experiments. The workflows have to be reproducible in order to be shared or reused. Unfortunately experiences have showed that many workflows failed on occasion of a later re-execution. Zhao et al. [23] [11] investigated the main purposes of the so-called *workflow decay*, which means that year by year the ability and success of the re-execution of any workflow significantly reduces. They found four main causes which have prevented the re-execution: 1. the missing environmental parameters, 2. missing third party resources; 3. missing descriptions about the workflows; 4. the missing samples of the experiments or the inputs and outputs of the workflows.

By incorporating these results into our previous paper [2] we have deeply investigated the requirements of the reproducibility and we have given a taxonomy of the different dependencies of the execution which can interfere with a later re-execution. To sum up our conclusions, in order to reproduce an in-silico experiment the scientist community and the system developers have to face three important challenges:

1) More and more meta-data has to be collected and stored pertaining to the infrastructure, the environment, the data dependencies and the partial results of an execution in order to make us capable of reconstructing the execution in a later time even on a different infrastructure. The collected data – called provenance data – help to store the actual parameters of the environments, the partial and final data results and system variables. Concerning the provenance, the challenge is what, where and how to store the captured information.

2) Descriptions and samples have to be stored together with the workflows which are provided by the user (scientist).

3) Some services or input data can change or become unavailable during the years. For example, third party services, special local services or continuously changing databases. Scientific workflows which are established on them can become instable and non-reproducible. In addition there are computations based on random generated values (for example, in case of image processing) thus, their executions are not deterministic so these computations cannot be repeated to provide the same result in a later time. These factors – we *call dependencies of the execution* - can especially influence the reproducibility of the scientific workflows, consequently, we have performed a deeper analysis.

The first issue can be solved by capturing detailed provenance information. The second one is the responsibility of the user (scientist), however, the scientific workflow management systems (SWfMS) can and should support the scientist to

provide detailed descriptions and samples. We have dealt with this issue in one of our previous paper [1].

In this paper I deal with the third issue. Based on a provenance database we introduce a so-called descriptor-space referred to the jobs of the workflow which contains all the parameters required to reproduce the jobs. The elements of the descriptor-space we call descriptors. Every descriptor has a name, a value and a so-called decay-parameter which refer to the fluctuation of the descriptor value. A workflow can be reproducible if all the descriptor values are known and storable. However, there are descriptors which cannot be stored (for example too big input), can become unavailable in later time (for example volatile third party resources), can vary in time (for example input originated from continuously changing database). Additionally, the descriptors can be either unknown if they are based on random generated values or other operation-related system-calls. In this case, the full reproducibility is very challenging task.

By our research, we intend to make the scientific workflow reproducible by extending the scientific workflow management system (SWfMS) with an analyzer tool. With the help of the expressions of the descriptors and the decay-parameters we can perform a pre-analysis before the execution. During this phase, we can examine the jobs of a given workflow and determine whether they are reproducible or not. If not, we determine the tools and the methods which can help to reproduce the job. According to the decay-parameter, the jobs can be grouped into four groups and executed in different ways. After the execution, based on provenance data a post-analysis can be performed by the application of statistical tools. An evaluation can be computed to replace the non-reproducible parts of the workflow.

In order to achieve our goal, on one hand we have analyzed [2] the criteria of the reproducibility on the other hand we have collected and have categorized all the necessary information which are required to reproduce the scientific workflows [1]. Finally, in this paper we set up a mathematical model to formalize the problem and determine certain statistical methods to predict, evaluate or simulate the results of the jobs and the re-executed workflows. We defined the descriptor space, the decay parameter of the descriptors and the reproducible job and workflow. Based on these definitions, we set up a mathematical model of the reproducibility analysis to formalize the problem and to give our solution.

The ultimate goal of our research is to make the workflows either reproducible by eliminating the dependencies or simulating the non-reproducible jobs of the scientific workflows.

Our paper is organized as follows. In the next subsection (1.1) we introduce the WS-PGARDE/gUSE system, in which we would like to test our results. Chapter 2 gives a brief summary about the related works. Chapter 3 represents our model and the components of the reproducibility analysis. In Chapter 4 we introduce the process and the phases of the analysis. Finally, we sum up our results in 5 and in

Chapter 6 we conclude our research with a brief provisioning of possible future research directions.

1.1. WS-PGRADE/gUSE

gUSE (grid and cloud user support environment) is a well-known and permanently improving open source science gateway framework developed by the Laboratory of Parallel and Distributed Systems (LPDS) that enables users the convenient and easy access to grid and cloud infrastructures. It has been developed to support a large variety of user communities. It provides a generic purpose, workflow-oriented graphical user interface to create and run workflows on various Distributed Computing Infrastructures (DCIs) including clusters, grids, desktop grids and clouds. [20]

The WS-PGRADE Portal [21] [10] is a web based front end of the gUSE infrastructure. The structure of WS-PGRADE workflows are represented by directed acyclic graphs.

The nodes of the graph, namely the jobs are the smallest units of a workflow. They represent a single algorithm, a stand-alone program or a web-service call to be executed. Ports represent input and output connectors of the given job node. Directed edges of the graph represent data dependency (and corresponding file transfer) among the workflow nodes. This abstract workflow can be used in the second step to generate various concrete workflows by configuring detailed properties (first of all the executable, the input/output files where needed and the target DCI) of the nodes representing the atomic execution units of the workflow.

A job may be executed if there is a proper data (or dataset in case of a collector port) at each of its input ports and there is no prohibiting programmed condition excluding the execution of the job. The execution of a workflow instance is data driven forced by the graph structure: A node will be activated (the associated job submitted or the associated service called) when the required input data elements (usually file, or set of files) become available at each input port of the node.

2 State of the Art

The researchers dealing with the reproducibility of scientific workflows have to approach this issue from two different aspects. First, the requirements of the reproducibility have to be investigated, analyzed and collected. Secondly, techniques and tools have to be developed and implemented to help the scientist in creating reproducible workflows.

2.1. Requirements

Researchers of this field agree on the importance of the careful design [8],[15], [16], [17], [22] which on one hand means the increased robustness of the scientific code, such as modular design and detailed description about the workflow, about the input/output data examples and consequent annotations [7]. On the other hand, the careful design includes the careful usage of volatile third party or special local services.

Groth et al. [10] based on several use cases analyzed the characteristics of applications used by workflows and listed seven requirements in order to enable the reproducibility of results and the determination of provenance. In addition, they showed that a combination of VM technology for partial workflow re-run along with provenance can be useful in certain cases to promote reproducibility.

Davison [7] investigated which provenance data have to be captured in order to reproduce the workflow. He listed six vital areas such as hardware platform, operating system identity and version, input and output data etc.

Zhao et al. [23] in their paper investigated the cause of the so called workflow decay. They examined 92 Taverna workflows submitted in the period between 2007 and 2012 and found four major causes: 1) Missing volatile third party resources 2) Missing example data 3) Missing execution environment (requirement of special local services) and 4) Insufficient descriptions about workflows. Hettne et al. [11] in their papers listed ten best practices to prevent the workflow decay.

2.2. Techniques and Tools

There are existing available tools, VisTrail, ReproZip or PROB [5], [9], [14] which allow the researcher and the scientist to create reproducible workflows. With the help of VisTrail [9], [12] reproducible paper can be created, which includes not only the description of scientific experiment, but all the links for input data, applications and visualized output. These links always harmonize with the actually applied input data, filter or other parameters. ReproZip [5] is another tool, which stitches together the detailed provenance information and the environmental parameters into a self-contained reproducible package.

The Research Object (RO) approach [3], [6] is a new direction in this research field. RO defines an extendable model, which aggregates a number of resources in a core or unit. Namely a workflow template; workflow runs obtained by enacting the workflow template; other artifacts which can be of different kinds; annotations describing the aforementioned elements and their relationships. Accordingly to the RO, the authors in [4] also investigate the requirements of the reproducibility and the required information necessary to achieve it. They created ontologies, which help to uniform these data. These ontologies can help our work and give us a basis

to perform our reproducibility analysis and make the workflows reproducible despite their dependencies.

Piccolo et al [18] collected the tools and techniques and proposed six strategies which can help the scientist to create reproducible scientific workflows.

Santana-Perez et al [19] proposed an alternative approach to reproduce scientific workflows which focused on the equipment of a computational experiment. They have developed an infrastructure-aware approach for computational execution environment conservation and reproducibility based on documenting the components of the infrastructure.

To sum up the results mentioned above, we can conclude that the general approach is that the scientist has to create reproducible workflows with careful design, appropriate tools and strategies. But none of them intended to solve the problem related to the dependencies rather they suggested to bypass them. Moreover, they did not deal with the following question: How an existing workflow can be made reproducible?

2.3. Reproducibility Support in WS-PGRADE/gUSE System

In the WS-PGRADE/gUSE system with the help of the "RESCUE" feature the user has the possibility to re-execute a job which does not own all the necessary inputs but the provenance data is available from the previous executions. (Fig. 1)
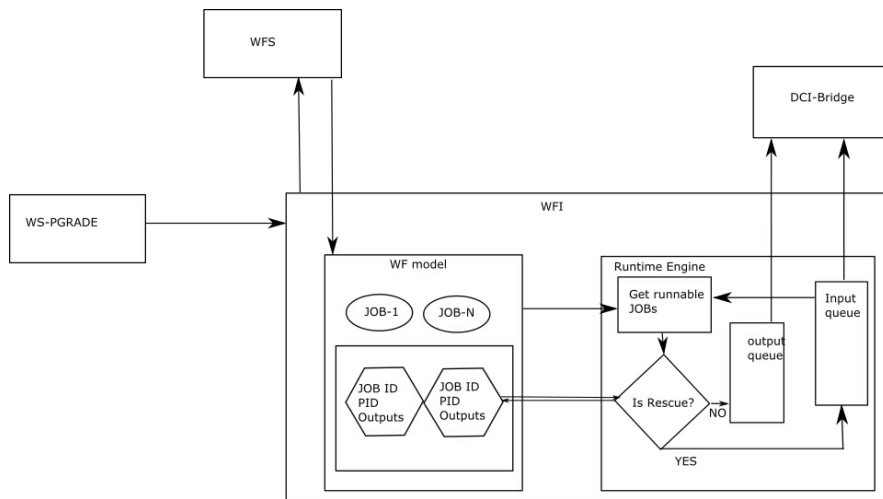

Figure 1

Operation of the *Rescue* feature in the WS-PGRADE/gUSE system

When submitting a job which has the identifier originated from the previous execution, the workflow instance (WFI) queries the description file of the workflow. This XML file includes the jobs belonging to the workflow. Their input

and output ports, their relations and the identifiers of the job instances executed previously with their outputs. After processing the XML file, a workflow model is created in the memory representing the given workflow during its execution. At this point the Runtime Engine (RE) takes over the control to determine the "ready to run" jobs then it examines whether these jobs have already stored outputs originated from previous executions. Concerning the answer the RE puts the job in the input or in the output queue.

3   Reproducibility Analysis

In this section, we introduce our mathematical model of reproducibility analysis. Next we give a method to handle the influence factors of the reproducibility of a job and to make the non-reproducible job reproducible under certain conditions or by a given probability. Finally we deal with jobs applying random generated values in an independent subsection.

3.1. The Model

In order to formalize the problem let us introduce the following notations and definitions:

- The scientific workflow (SWf) can be represented by a directed acyclic graph, where the vertices denote the jobs and the edges denote the dataflow between jobs.

    $V = \{J_1, \dots, J_N\}$, where $N \in$ N; the number of the job of a given workflow

    $E = \{(J_i, J_j) \in V \times V | i \in [1, 2, \dots N - 1]; j \in [2, 3, \dots, N] \text{ and } i \neq j\}$

- The job J_i is *exit job* (exit node) in the graph, if $\nexists J_j \in V: (J_i, J_j) \in E$; Notation: J_{exit}

- The job J_i is *entry job* (entry node) in the graph, if $\nexists J_j \in V: (J_j, J_i) \in E$; Notation: J_{entry}

- The job, which is neither exit nor entry job, is an *inside job*.

- The *forward sub-workflow of a job J_i* is the part of the workflow, which contains all the successor jobs (nodes) and the edges between them.

- From our point of view the SWf is a function: $\mathbf{SWF}(t_0, J_1, J_2, \dots, J_N) = \mathbf{Y}$, where t_0 is a given time of the submission of the workflow and Y is the result of the workflow.

- Assuming that the workflow was successfully executed at least once and provenance database is available a descriptor-space $\boldsymbol{D_{J_i}}$ can be created to store all the necessary parameter needed to re-execute the job.

- The job J_i ($i = 1, 2, \ldots, N$) has K_i descriptors: $V_1, V_2, \ldots, V_{Ki}$, which are necessary to reproduce the workflows. The values of descriptors are: $\boldsymbol{D_{J_i}} = \{\boldsymbol{v_{i1}, v_{i2}, \ldots, v_{iK_i}}\}$

- With the help of the descriptors every job can be written as a function: $\boldsymbol{J_i(t_0, v_{i1}, v_{i2}, \ldots, v_{iK_i}) = Y_i}$

- For every descriptor we have defined a so called *decay-parameter* which indicates how the descriptor's value changes in time. There are four cases:

  1. The availability and the value of the descriptor is not changing in time. In this case the decay parameter is 0.

  2. The availability of the descriptor is changing in time. There are two cases: the probability distribution function of the descriptor's availability is known or not. In the first case, the decay parameter can be determined by the given distribution function and in the second one, the descriptor value is infinite.

  3. The value of the descriptor is changing in time. Similarly to the second item, the change of the value can be known or unknown. According to the actual case the value of decay parameter is a function describing the change or it is infinite.

  4. The value of the descriptor is not constant, but both its availability and change is unknown. For example a random generated value, which is used during the execution but it is not known. In this case the value of the decay-parameter is infinite.

In formal:

$$decay(v_i) = \begin{cases} \boldsymbol{0}, & \text{if the value of the descriptor is not changing in time} \\ \boldsymbol{\infty}, & \textbf{if the value of the descriptor is unknown} \\ \boldsymbol{F_i(t)}, & \textbf{distribution function} \text{ of the availability of the given value} \\ \boldsymbol{Vary_i(t, v_i)}, & \textbf{if the value of the descriptor is changing in time} \end{cases}$$

With help of these expressions we can define the reproducibility as the following way:

Definition (D1): The J_i job is *reproducible*, if the descriptor space $D_{J_i} = \{v_{i1}, v_{i2}, \dots, v_{iK_i}\}$ of job – which contains all the inputs and environmental parameters - is known and can be stored, in other words all the decay parameters are zero.

Notation: J_i^{repro}; $JOB_i^{repro}(v_{i1}, v_{i2}, \dots, v_{iK_i}) = Y_i$

Corollary: A reproducible job is invariable in time (time-independent):

$$JOB_i^{repro}(t_0, v_{i1}, v_{i2}, \dots, v_{iK_i}) = JOB_i^{repro}(t_0 + \Delta t, v_{i1}, v_{i2}, \dots, v_{iK_i}) = Y_i \quad \text{for}$$
every Δt.

Definition (D2): The *scientific workflow* is *reproducible*, if its exit jobs and the $SubWF_{J_{exit}}^{back}$ of the exit jobs is reproducible.

It can be easily proven, that if and only if every job of a SWf is reproducible, then the SWf is also reproducible.

We introduce other properties, namely the substitutional and the approximative reproducibility referring to that case, in which the decay parameter of one of its descriptors changes in time and this variation is known. There is two option: the first one is that the variation of result can be described with a function determined by the variation function of the descriptor; the second one is that the variation of result can be estimated. In formal:

Definition (D2): The J_i job is *reproducible by substitution*, if the descriptor space $\{v_{i1}, v_{i2}, \dots, v_{iK_i}\}$ of job and $\exists k \in [1, 2, \dots, K_i]: Vary_{ik}(\Delta t, v_{ik})$ is known, and based on vary function a $Vary_i^*(\Delta t, Y_i)$ can be unambiguously determined.

If
$$JOB_i(t_0, v_{i1}, v_{i2}, \dots, v_{iK_i}) = Y_i$$

Then
$$JOB_i(t_0 + \Delta t, v_{i1}, v_{i2}, \dots, vary_{ik}(\Delta t, v_{ik}), \dots, v_{iK_i}) = Vary_i^*(\Delta t, Y_i)$$

Notation: $JOB_i^{vary}(v_{i1}, v_{i2}, \dots, vary_{ik}(\Delta t, v_{ik}), \dots v_{iK_i})$

Definition (D3): The J_i job is *approximately reproducible*, if J_i is reproducible under condition that $\exists k \in [1, 2, \dots, K_i]: Vary_{ik}(\Delta t, v_{ik})$ is precisely known, and in accordance this function a $Vary_i^*(\Delta t, Y_i)$ can be estimated with an acceptable accuracy:

$$JOB_i(t_0 + \Delta t, v_{i1}, v_{i2}, \dots, vary_{ik}(\Delta t, v_{ik}), \dots, v_{iK_i}) \approx Vary_i^{appro}(\Delta t, Y_i) = \widetilde{Y_i}$$

Notation**: $JOB_i^{appro}(v_{i1}, v_{i2}, \dots, vary_{ik}(\Delta t, v_{ik}), \dots v_{iK_i})$**

3.2. Pre-Analysis

The first step of the process of the reproducibility analysis is to create the descriptor space of all the jobs belonging to the given scientific workflow. The descriptors and their decay parameters can originate from three different sources: from the users, from the provenance database and it can be automatically generated by the SWfMS. [1]

Analyzing the decay parameters of the descriptors we have separated those, which can influence the reproducibility of the workflow in other words which have non-zero decay parameters. Four groups have been created:

1. With the help of additional resources or tools this dependency of execution can be eliminated. For example, in case of random generated values we are going to implement an operating system level tool, which captures the return value of the random generator, and stores it in the provenance database (see subsection 3.4)

2. With the help of approximation tools the value of the descriptor can be evaluated or even replaced. (see subsection 3.3)

3. A time interval can be given during which the descriptor is available by a given probability *p*.

4. There is no method to make the workflow reproducible.

3.3. Evaluation

In this subsection we investigate the case when one decay parameter of the job's descriptors is changing in time.

In case of the presented methods we assume two essential conditions:

1. The availability of the whole descriptor's space of the job in a given SWf, which means all the necessary information to reproduce the job.

2. The availability of a provenance database which contains the provenance information about the previous executions of a given SWf. For example, descriptor values, partial and final results of the jobs etc.

Based on provenance database a sample set can be defined which contains provenance data originated from *s* (where *s* is a natural number) previous executions:

$$
S = \left\{
\begin{array}{c}
J_i\left(t_0, v_{i1}^0, v_{i2}^0, \dots, v_{iK_i}^0\right) = Y_i^0 \\
J_i\left(t_0, v_{i1}^1, v_{i2}^1, \dots, v_{iK_i}^1\right) = Y_i^1 \\
\dots \\
J_i\left(t_0, v_{i1}^{s-1}, v_{i2}^{s-1}, \dots, v_{iK_i}^{s-1}\right) = Y_i^{s-1}
\end{array}
\right\}
\tag{1}
$$

If a v_{ij} ($i = 1 \dots N$; $j = 1 \dots K_i$) descriptor's value is not changing in time its decay parameter is 0, thus, in the sample set the given elements are equals:

$$v_{ij}^{0} = v_{ij}^{1} = \dots = v_{ij}^{s-1}$$

Assuming that only one descriptor value is changing in time the sample set related to the job J_i can be written in a simpler form:

$$S = \{(t_0, v_{ij}^0, Y_i^0), (t_0, v_{ij}^1, Y_i^1), \dots, (t_0, v_{ij}^{s-1}, Y_i^{s-1})\} \tag{2}$$

Based on the sample set $S = i$ the correlation can be investigated between the variables v_{ij}^{k} an Y_i^{k} ($k = 1, 2, \dots$, s-1) with the? help of the following expression:

$$corr(v, Y) = \frac{\sum_{k=0}^{S-1}(v_{ij}^k - \widetilde{v_{\iota J}})(Y_i^k - \widetilde{Y_\iota})}{\sqrt{\sum_{k=0}^{S-1}(v_{ij}^k - \widetilde{v_{\iota J}})^2 \sum_{k=0}^{S-1}(Y_i^k - \widetilde{Y_\iota})^2}}, \tag{3}$$

where $\widetilde{Y_\iota}$ and $\widetilde{v_{\iota J}}$ are the empirical (sample) mean of the adequate variables.

In addition, based on provenance data we can determine the coverage of a given descriptor, which contains every job influenced by this descriptor. We can compute the correlation matrix of the $v_{i,j}$ descriptor and the results of all the successors of the job J_i.

$$\mathbf{R}_{(p+1)\times(p+1)} = \begin{pmatrix} cor(v_{i,j}, v_{i,j}) & cor(v_{i,j}, Y_1) & \dots & cor(v_{i,j}, Y_p) \\ cor(Y_1, v_{i,j}) & cor(Y_1, Y_1) & & cor(Y_1, Y_p) \\ & & & \\ cor(Y_p, v_{i,j}) & cor(Y_p, Y_1) & & cor(Y_p, Y_p) \end{pmatrix} \tag{4}$$

where $Y_1, Y_2 \dots, Y_p$ is the results of the successors of job J_i. The R matrix is symmetric and the values in the diagonal are 1.

The coverage of the given descriptor can be determined based on the first row of the correlation matrix. The non-zero values, which are close to 1 can show which J_i, i = 1, 2, …, p belong to the coverage zone.

Concerning to the value of the expression (3) we can differentiate two cases:

1. The result is close to 1, which means that the two variables are bounded? up with each other thus the result Y_i. can be evaluated by applying some approximation. For example, the linear regression consequently, the result Y_i. can be written as a linear combination of the changing descriptor.

    $$Y_i = \beta_0 + \beta_1 v_{ij} \tag{5}$$

    where the β_0 and β_1 are the linear coefficients.

    In this way, $Y_i^t = \beta_0 + \beta_1 vary_{ij}(t, v_{ij})$, where t is arbitrary.

    If the result of (3) is closer to 0.5 then to 1, nonlinear regression or other curve fitting method can be used.

Storing the approximation and the final results in the repository makes it possible that during the re-execution of a workflow, the non-reproducible job can be replaced by these approximated or simulated results.

we call The scientific workflows associated to this group reproducible by substitution or approximately reproducible scientific workflows.

2. The result of the correlation coefficient is close to 0, which means that the descriptor v_{ij} does not influence the result Y_i. In this case, the analysis has to be continued and the correlations between the results of the successor jobs have to be investigated.

3.4. Random Based Dependency

Many jobs use applications and computations which are based on random generated values (RGV). For example, the image processing applications, the different simulators and workflows which simulate some physical or chemical phenomena or even cryptographic algorithms. In this case, during the execution a system call is performed which returns a random generated value but this result is stored only in the memory. Consequently, provenance information does not get into the provenance database. We have designed a tool which operates at the operating system level and it captures the return values of the system call. Next, it stores the given value in the provenance database or on a predefined location. With the help of this tool the random RGVs can be stored together with the workflow in a repository. In/on? occasion of a later re-execution, the SWfMS uses the originally stored value instead of the newly generated random value.

4   The Process of Reproducibility-Analysis

Based on the decay-parameter (DP) the pre-analyzer performs a classification of the jobs of the given SWf. Depending on the classification, the job can be executed in three ways:

1. Standard execution, if all the decay parameters are zero.

2. Replacing the execution with evaluation, if there are changing descriptor values or the availabilities are defined by a probability distribution function (PDF).

3. Execution with random value capture (RVC) tool, if the execution of the job is based on random generated value.

In all cases updating the Provenance Database (PDB) is performed occasionally by extra provenance information (for example a random value).

Based on the PDB the post-analyzer creates a sample set. The evaluator module computes the evaluated output of the given job. Figure 2 shows the flow-chart about the process and Figure 3 presents the block-diagram.
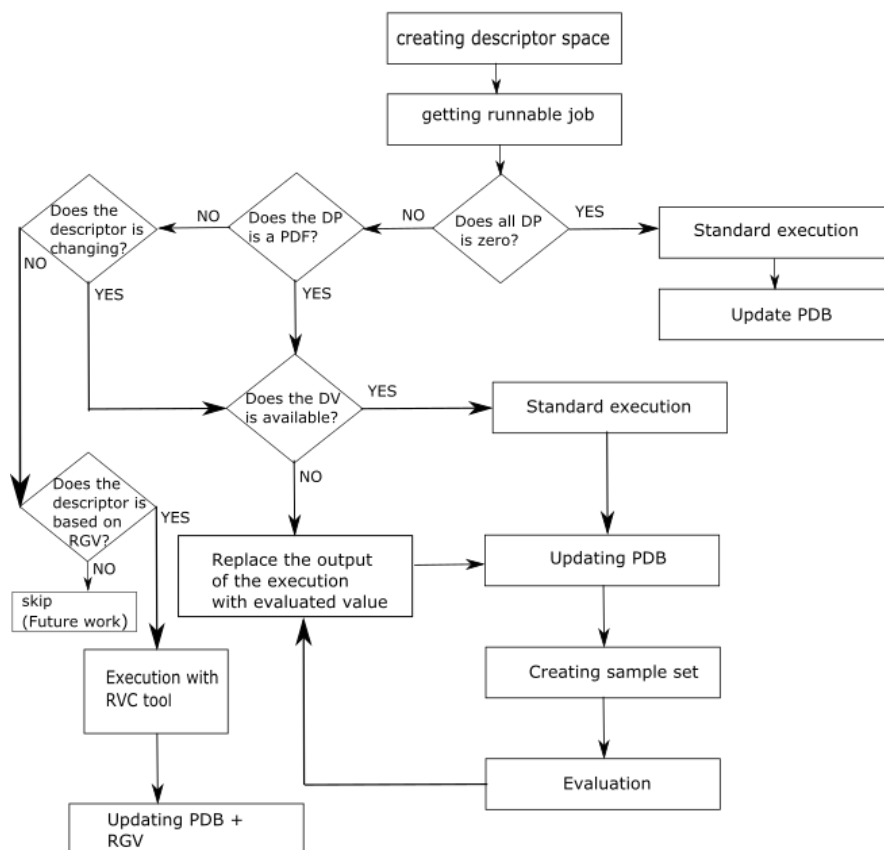

Figure 2

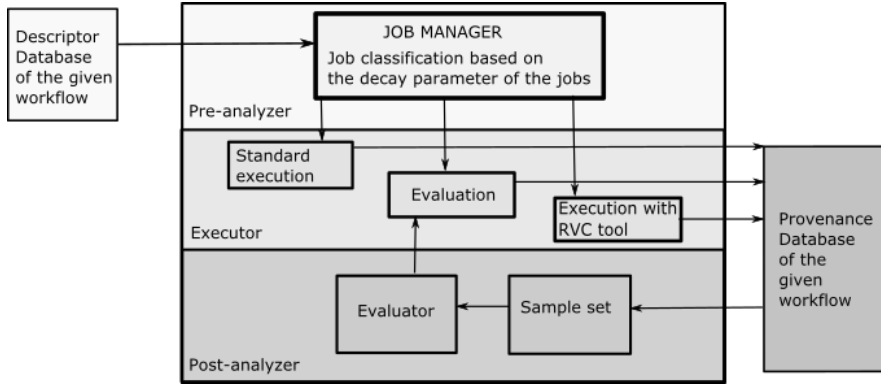The flow-chart of the process of the reproducibility analysis

Figure 3

The block-diagram of the system of the reproducibility analysis

5 Results and Implementation

Since this investigation is based on the descriptor-space and the descriptor-space is based on the provenance database, the first step toward the implementation of an evaluating tool is the implementation of a provenance framework. The implementation of a Provenance manager (PROV-man) framework is already finished. It provides functionalities to create and manipulate provenance data in a consistent manner and ensures its permanent storage. It also provides a set of interfaces to serialize and export provenance data into various data format, serving interoperability [24], [25], [26]. Three main components constitutes the PROV-man framework:

- A set of methods to build and manipulate provenance data, while preserving full compliance with the PROV specifications

- A set of interfaces for provenance data sharing and interoperation. These interfaces covers serialization to formats of the PROV family of documents (e.g. XML, RDF, DC, etc.) and other specifically required format (e.g. Graphviz, PDF, JPG, etc.)

- A relational database that serves as a main repository for storing provenance data, reflecting the PROV-man data model

Additionally, the "rescue" feature and the tool which captures the RGVs and stores them in the PDB or in a predefined location are implemented in the WS-PGRADE/gUSE system. Furthermore, we intend to extend it with an evaluating tool in case the job cannot run because of the missing or unavailable inputs. This solution makes us able to apply not only the previously stored results but an evaluated or a simulated output of a previously executed job.

Conclusions

We analyzed the requirements of the reproducibility and the critical, continuously changing or non-deterministic descriptors of the scientific workflows to make them reproducible. To formalize the problem, we set up a mathematical model and gave definitions of the reproducible jobs and workflows. Based on the model, we worked out a reproducibility analysis process which involves three phases. The first is a pre-analysis based on the descriptor's space to determine the reproducible parts of the workflow and to classify the jobs according to their decay-parameter. The jobs in the different classes are executed in different ways. In the post-analysis phase, assuming that provenance data is available about the previous executions a sample set is created and the determination of the evaluating algorithm is performed. This information is stored together with the workflow in the repository or in the provenance database. On occasion of a re-execution of the workflow and in case of a non-reproducible job, instead of the standard execution we evaluate the outputs based on the stored sample set and on the evaluating parameters.

The presented framework is theoretical in the sense that the time is limitless. If the probability distribution function of the availability referring to a descriptor is given or can be estimated based on provenance, the limit of the function as time approaches infinity is 1, and the time – during the descriptor is available – can be determined, theoretically. If an estimating method can be determined for a changing descriptor, it is also "time-limitless" and the method can be applied at any time, maybe if the appropriate resources is out of time. However, the experience actually shows, that the technological development can be prevent the re-execution and can shorter the theoretical time given by our analysis.

In addition, we particularly have dealt with the job executions based on random generated values and we have developed a mechanism which is able to capture the return values of the system calls and to store it for a later re-execution.

In our future work, we would like to develop other tools to be able to handle more special dependencies of the workflow execution. Also we intend to explore other procedures to find a more general solution for the evaluating problems when many descriptors' values change, simultaneously. Furthermore, we plan to implement our methods and tools within the gUSE framework.

Acknowledgement

References

[1]   Bánáti A., Kacsuk P., Kozlovszky M.: Minimal Sufficient Information about the Scientific Workows to Create Reproducible Experiment. In: IEEE 19th International Conference on Intelligent Engineering Systems (INES), Slovakia, 2015, pp. 189-194

[2]   Bánáti A., Kacsuk P. Kozlovszky, M.: "Four Level Provenance Support to Achieve Portable Reproducibility of Scientific Workflows" In Information and Communication Technology, Electronics and Microelectronics (MIPRO) 2015 38th International Convention on (pp. 241-244) IEEE

[3]   Bechhofer S., De Roure D., Gamble M., Goble C., Buchan I.: „Research Objects: Towards Exchange and Reuse of Digital Knowledge" In: he Future of the Web for Collaborative Science, 2010

[4]   Belhajjame K., Zhao J., Garijo D., Gamble M., Hettne K., Palma R., Goble C.: „Using a Suite of Ontologies for Preserving Workow-Centric Research Objects" In: Web Semantics: Science, Services and Agents on the World Wide Web, 2015

[5]   Chirigati F., S., Shasha D., Freire J.: „ReproZip: Using Provenance to Support Computational Reproducibility" Presented as part of the 5th USENIX Workshop on the Theory and Practice of Provenance, 2013

[6]   Corcho O., Garijo Verdejo D., Belhajjame K., Zhao J., Missier P., Newman D., Goble C.: „Workflow-Centric Research Objects: First Class Citizens in Scholarly Discourse" In: Proceedings of Sepublica 2012, pp. 1-12, 2012

[7]   Davison, A. „Automated Capture of Experiment Context for Easier Reproducibility in Computational Research", Computing in Science & Engineering, Vol 14/ 4, pp. 48-56, July 2012

[8]   De Roure D., Belhajjame K., Missier P., Gmez-Prez J. M., Palma R., Ruiz J. E., Hettne K., Roos M., Klyne G., Hekkelman M. L.: „Towards the Preservation of Scientific Workows". In Proceedings of 8th International Conference on Preservation of Digital Objects (iPRES 2011) 2011

[9]   Freire J., Koop D., Chirigati F. S, and Silva C. T.,: "Reproducibility Using VisTrails", Implementing Reproducible Research 33, 2014, OnlineAvailable:
        http://citeseerx.ist.psu.edu/viewdoc/download doi:10.1.1.369.9566

[10]   Groth, P; Deelman E., Juve G., Mehta G., and Berriman B., „Pipeline-Centric Provenance Model", in Proceedings of the 4th Workshop on Workflows in Support of Large-Scale Science, 2009, p. 4

[11]   Hettne, K. Wolstencroft, K. Belhajjame, Goble C. A., Mina E., Dharuri H., D. De Roure, Verdes-Montenegro L., Garrido J., and M. Roos, „Best Practices for Workflow Design: How to Prevent Workflow Decay", in SWAT4LS, 2012

[12]   Koop D., Freire J., and Silva C. T., „Enabling Reproducible Science with VisTrails", arXiv preprint arXiv:1309. 1784, 2013

[13]   Koop, D., Santos, E., Mates, P., Vo, H. T., Bonnet, P., Bauer, B., ... and Freire, J. (2011) "A provenance-based Infrastructure to Support the Life Cycle of Executable Papers". Procedia Computer Science, 4, 648-657

[14]   Korolev, A. Joshi, V. Korolev, M. A. Grasso, A. Joshi, M. A. Grasso, D. Dalvi, S. Das, V. Korolev, Y. Yesha, and others, „PROB: A Tool for Tracking Provenance and Reproducibility of Big Data Experiments.", Reproduce'14. HPCA 2014, Vol. 11, pp. 264-286, 2014

[15]   Mesirov J. P., „Accessible Reproducible Research", Science, Vol. 327/ 5964, pp. 415-416, January 2010

[16]   Missier P., Woodman S., Hiden H., and P. Watson, „Provenance and Data Differencing for Workflow Reproducibility Analysis", Concurrency and Computation: Practice and Experience, 2013

[17]   Peng, „Reproducible Research in Computational Science", Science, vol. 334/ 6060, pp. 1226-1227, 2011

[18]   Piccolo S. R., Lee A. B., Frampton M. B.: „Tools and Techniques for Computational Reproducibility". In: bioRxiv, Vol. 022707, 2015

[19]   Santana-Perez I., Prez-Hernndez M. S.: „Towards Reproducibility in Scientific Workows: An Infrastructure-based Approach". In: Scientific Programming, Vol. 2015, p. 11, 2015

[20]   SZTAKI LPDS: User's Guide. Http://guse.hu/about/home

[21]   SZTAKI LPDS: User's Guide
        Http://sourceforge.net/projects/guse/_les/3.7.4/Documentation

[22]   Woodman, Hiden H., Watson P., and Missier P., „Achieving Reproducibility by Combining Provenance with Service and Workflow Versioning", in Proceedings of the 6th Workshop on Workflows in Support of Large-Scale Science, 2011, pp. 127-136

[23]   Zhao J., Gomez-Perez J. M., Belhajjame K., Klyne G., Garcia-Cuesta E., Garrido A., Hettne K., Roos M., De Roure D., and Goble C., „Why Workflows Break—Understanding and Combating Decay in Taverna Workflows", in E-Science (e-Science), 2012 IEEE 8th International Conference on, 2012, pp. 1-9

[24]   http://nl.sharp-sys.com/PROV-man.html

[25]   Kiss, Tamás, et al. "Ws-pgrade/guse in European Projects." Science Gateways for Distributed Computing Infrastructures. Springer International Publishing, 2014, 235-254

[26]   Benabdelkader, A., Antoine AHC van Kampen, and Silvia D. Olabarriaga. PROV-Man: A PROV-Compliant Toolkit for Provenance Management. No. e1347. PeerJ PrePrints, 2015