Position Identification of Moving Agents in Networks

Anna Černá, Jan Černý

Faculty of Mangement, University of Economics, Prague
Jindřichův Hradec, Czech Republic
cerna@fm.vse.cz, cerny@fm.vse.cz

Abstract: The paper introduces a group of problems connected with the position identification of an agent moving through a given network and starting in an unknown vertex or edge (arrow). It presents a classification of the problems, describes the application fields and outlines further possible developments of the theory.

Keywords: Agent; position; identification; network

1 Introduction

Agent-based models represent a topic of wide and deep studies during the past two decades. They have been used in the simulation of road traffic [6], in grid calculation [4], in human systems [2], and in coalitions [7], and one can find books dealing with them, see e.g. [15]. Agents can possess different "cleverness", as one can see in the agent typology presented in a paper published in Acta Polytechnica Hungarica in 2006 [12]. Their learning capability is described in [18].

Mathematical models describing the motion of agents through networks are described in, for example, [8], [13], [20], although in [8] and [20], the authors use the notion "robot" instead of "agent". In contrast to [20], we shall deal with other approaches to "navigation". On the one hand, we shall suppose that the agent moves through a given network; on the other hand the agent is supposed to be able to deal with other observations. Our approach represents a generalization of the one from [5], [14], and [19]. They found methods leading the agent to the target vertex when the network fulfils some particular constraint of "synchronizing colouring". In that case, the "cleverness" of the agent may be rather low. We weaken the first feature and strengthen the second, achieving both in several different ways.

We shall present several methods of solving problems of agent position identification and we shall outline their possible connection to problems which have been known now for decades.

The problems of agent position identification in networks can be encountered in several branches of polytechnics, for example:

- We can meet objects moving through **telecommunication networks** in ways similar to the agents in our models. We can mention, for example, tokens in token rings, packets in packet mode networks and so forth, and identification of their positions is a natural issue.
- A robot inspecting a **pipeline network** from inside is another good example of application.
- A speleologist exploring a **cave network** can be regarded as an agent moving through it. However, our theory has a sense only in the case when the explorer cannot use a Tarry or Tremaux algorithm based on marking the ends of incident edges in vertices.
- **Transportation networks** represent probably the most important application field of the problems connected with the identification of and agent's positions.

The last-but-not-least thing to be mentioned is the fact that the models and methods used in automata theory studying synchronizing input sequences are closely related to the ones used in agent position identification theory. The bridge connecting these two branches can be seen in the problem of road colouring.

2 Types of Graphs Used

We expect the reader is familiar with the terminology of graph theory. We shall only specify several notions and denotations here.

First we mention that the **vertex set** V is always finite. If $v \in V$ and $w \in V$, then we form two kinds of pairs:

- **non-oriented pair** (v, w) without any order of v and w, i.e. $(v, w) = (w, v)$,
- **oriented pair** vw where v is the beginning and w is the end, i.e. $vw \neq wv$ for $v \neq w$.

The set of all non-oriented pairs (v, w) of vertices from V is denoted VV, and its elements are called **edges**. They are graphically expressed by abscissae connecting the symbols of vertices v and w, usually linear, exceptionally curvilinear.

The set of all oriented pairs of vertices vw from V is denoted $V \times V$ or V^2 and its elements are called **arrows**. They are graphically expressed by arrows starting in the symbol of the vertex v and ending in the symbol of w, usually linear, exceptionally curvilinear.

Note that in either $(v, w) \in E$ or in $vw \in A$ we allow the possibility $v = w$. In such a case we speak about **loops**, non-oriented or oriented respectively. Similarly, if we do not specify otherwise, we allow that a pair of vertices v, w is connected by more than one edge (v, w) or arrow vw. In such a case, either we distinguish between them by an auxiliary criterion, e.g. a colour or a position ("first to the left"), or we take an arbitrary one.

If either $(v, w) \in E$ or $vw \in A$, then we say that v and w are **incident** with (v, w) or vw, respectively.

Finally, we denote $2^V = \{W: W \subset V\}$ as the group of all subsets of V.

2.1 We speak about the **graph** $G = (V, E)$ if $E \subset VV$. The set E is said to be a **set of edges** of G. In other words, we speak briefly about a graph instead of "non-oriented graph".

2.2 We speak about the **digraph** $D = (V, A)$ if $A \subset V^2$. The set A is said to be a **set of arrows** of D. In other words, we speak briefly about a digraph instead of "oriented graph". Let $k > 1$. A vertex $v \in V$ is said to be k-**merging** if there exist k different arrows $u_1x, \ldots, u_kx$.

2.3 Let Γ be a finite set of **colours**. The graph $G = (V, E)$ or the digraph $D = (V, A)$ is said to be **coloured** if each $e \in E$ or $a \in A$ is assigned a colour $\gamma(e) \in \Gamma$ or $\gamma(a) \in \Gamma$, respectively. The mapping γ is said to be **colouring**. That is, speaking on colouring, we mean the colouring of edges or arrows. If we need to speak about the colouring of vertices, we will specify so. If we do not specify otherwise, we shall suppose that the colouring is a mapping **onto** the set Γ i.e. for any colour $c \in \Gamma$ there exist $x \in E$ or $x \in A$ such that $c = \gamma(x)$.

Let $D = (V, A)$ be a digraph, $\gamma \in \Gamma$ and $k > 1$. A vertex $v \in V$ is said to be k-**merging** for γ if there exist k different arrows $u_1x, \ldots, u_kx$ with the colour γ.

2.4 The digraph $D = (V, A)$ is said to be **derived** from the graph $G = (V, E)$ if exactly one of the relations $vw \in A$, $wv \in A$ holds for each $(v, w) \in E$. Moreover, if G is coloured, then D is coloured as well and $\gamma(vw) = \gamma((v, w))$ for each $vw \in A$. Note that D arises from G by the replacing of each edge by an arrow connecting the same vertices and it is of the same colour if G was coloured.

2.5 The digraph $DG = (V, AG)$ is said to be **equivalent** to the graph $G = (V, E)$ if $vw \in AG$ and $wv \in AG$ if and only if $(v, w) \in E$. Moreover, if G is coloured, then DG is coloured as well and $\gamma(vw) = \gamma(wv) = \gamma((v, w))$ for each $(v, w) \in E$. Note that DG arises from G by the replacing of all edges by pairs of counter arrows and they are of the same colour if G was coloured.

2.6 Let $D = (V, A)$ be a digraph and let $AD \subset VV \times VV$. Let $vw \in A$, $xy \in A$ if and only if $(v, x)(w, y) \in AD$. Then the digraph $DD = (VV, AD)$ is said to be the **double digraph** derived from D.

2.7 Let $D = (V, A)$ be a digraph and let $AT \subset 2^V \times 2^V$. Let $U \subset V$, $W \subset V$. Let $UW \in AT$ if and only if $W = \{w \in V$: there exists $u \in U$ such that $uw \in A\}$. Then the digraph $DT = (2^V, AT)$ is said to be the **total digraph** derived from D.

2.8 Let $D = (V, A)$ be a digraph, let $v \in V$ and let $AR \subset 2^{V-\{v\}} \times 2^{V-\{v\}}$. Let $U \subset V-\{v\}$, $W \subset V-\{v\}$. Let $UW \in AR$ if and only if $W = \{w \in V-\{v\}$: there exists $u \in U$ such that $uw \in A\}$. Then the digraph $DRv = (2^{V-\{v\}}, AR)$ is said to be the **reduced total digraph** derived from D and v.

2.9 Let C be a finite set called **set of commands**. Let $D = (V, A)$ be a digraph and let $\delta : A \times C \to A$ be a mapping. We say that the arc $\delta(a, c)$ **follows** after the arc a **due to the command** c. Let $A\delta = \{ab : a \in A, b \in A$ and there exists $c \in C$, $\delta(a, c) = b\}$. Let $D\delta C = (A, A\delta)$ be a coloured digraph such that $\delta(a, c) = b$ implies that the arrow ab has the colour c. Then the digraph $D\delta C$ is called the **transition digraph** corresponding to the digraph D and the set of commands C. We note that the vertices of $D\delta C$ are arrows of the former digraph D and the arrows of $D\delta C$ express the choice of next arrow in D due to the received command.

Remark: Commands from the set C serve as colours of arrows in the graph $D\delta C$.

2.10 A graph $G = (V, E)$ or a digraph $D = (V, A)$ is said to be **locally planar** if it is embedded in a smooth surface. This embedding determines cyclic ordering of edges or arrows incident with v for each $v \in V$. Since different embeddings of the same graph or digraph may determine different cyclic orderings, say "Given locally planar graph/digraph", we shall mean that it is given together with its embedding in a given surface, i.e. together with the cyclic orderings.

We point out that the cyclic ordering is normally meant in a counter clockwise sense. It is similar to the ordering of adjacent roads in a roundabout (continental European, not British, where the motion is clockwise).

2.11 Usually, for a motion of an agent through a graph or a digraph, it is important to note if there exists a path from v to w for each pair of vertices v, w. A graph $G = (V, E)$ having that property is said to be **connected**; a similar digraph $D = (V, A)$ is said to **strongly connected**. If in the digraph there exists a vertex v such that there exists a path from any $w \in V$ to v, then the digraph is said to be **weakly connected.** If we do not specify otherwise, we shall automatically suppose that the agent moves through a network which is connected in the non-oriented case and weakly connected in the oriented one.

3 Moving Agents

3.1 An **agent** is defined as a finite discrete system $\mathcal{S}$ composed from two finite automata, namely an executor $\mathcal{A}$ performing the moving and a controller $\mathcal{A}'$ giving commands to $\mathcal{A}$. That is, from the point of view of the classification of agents, presented e.g. in [12], the agents we shall deal with belong to the class of **hysteretic** agents. Of course, in the further development of the theory, a role of "less clever" reactive agents or "more clever" deliberative and, maybe mainly, attentive ones will be studied as well.

We can imagine that the agent is specified and constructed by some external client, after which the agent enters the network in an "unknown" position, and from that moment it moves step by step in the network autonomously until the controller orders it to stop (usually in a precisely identified position).

3.2 The **executor** $\mathcal{A} = (Q, C, \delta)$ is without output. **Internal states** $q \in Q$ represent **positions** of the agent; **inputs** $c \in C$ represent **commands** coming from $\mathcal{A}'$; δ represents a **transition mapping**; that is, if the agent is in the position q and $\mathcal{A}$ receives a command c, then it moves to the new position $\delta(q, c)$.

We must admit that the executor $\mathcal{A}$ can only be **partially defined**, or in other words. the next position $\delta(q, c)$ need not be defined for some position $q \in Q$ and a command $c \in C$. Then the agent remains in the position q and waits for the next command. In other words, it behaves like $\delta(q, c) = q$ in that case like it was fully defined. This approach differs from the one of e.g. [16], where it is supposed that the input command c cannot come in such a case.

We suppose that exactly one agent moves through the given graph or digraph and its movement, initiated by a sequence of commands $\boldsymbol{s} = c_1, \ldots, c_n$, represents a discrete process $q_0, q_1, \ldots, q_n$, where q_0 is the initial position and for $i = 1, 2, \ldots, n$ we have $q_i = \delta(q_{i-1}, c_i)$. If the command is c = stop then $\delta(q, c) = q$.

Obviously, our approach possesses an extension to the case of more than one moving agent, but we shall not study it now.

3.3 The **transition mapping** δ can be **extended** and **modified** in several senses.

First, the previously-mentioned relations can be expressed also by the formulae:

$$q_n = \delta(q_0, \boldsymbol{s}) = \delta(q_0, c_1, \ldots, c_n) = \delta(\ldots \delta(\delta(q_0, c_1), c_2), \ldots, c_n)$$

$$q_0, q_1, \ldots, q_n = \delta^*(q_0, \boldsymbol{s}) = \delta^*(q_0, c_1, \ldots, c_n) =$$

$$= q_0, \delta(q_0, c_1), \ldots, \delta(q_0, c_1, \ldots, c_n)$$

The indexes of positions and commands need not be concordant; i.e. we can write

$$q_{m+n} = \delta(q_m, \boldsymbol{s}) = \delta(q_m, c_1, \ldots, c_n)$$

$$q_m, q_{m+1}, \ldots, q_n = \delta^*(q_m, \boldsymbol{s}) = \delta^*(q_m, c_1, \ldots, c_n)$$

3.4 The **controller** $\mathcal{A}' = (Q', X, C, \delta', \lambda', q_0')$ is a finite state transducer (= "complete" initial finite automaton – see 5.1) having the output set C. Its input set X represents the possible observations of some data in the current positions of the agent.

The simplest controller subsequently emits the "a priori" given sequence $c_1, \ldots, c_n$. It has the set of internal states $Q' = C$, the initial state $q_0' = c_1$, the input set $X = \emptyset$, and the output set $Y = C$, the transition mapping non depending on input: $\delta(c_i) = c_{i+1}$ for $i = 1, \ldots, n-1$ and $\delta(c_n) = c_n$. The output mapping λ is identical, i.e. $\lambda(c_i) = c_i$. This controller realizes the situation when the sequence of commands $\boldsymbol{s} = c_1, \ldots, c_n$ is known at the beginning and does not depend on the observations during the movement of the agent. Such a controller is called **trivial.**

More "smart" controller deduces the emitted commands as from observed current data x as from the previous steps "reflected" in the state q'.

3.5 Usually, if the agent moves in the digraph $D = (V, A)$ then we put $Q = V$, i.e. the positions are in vertices and arrows serve to transitions between vertices only. In this case we speak about "**vertex bound positions**".

On the other hand, if it moves in the graph $G = (V, E)$, then we model it in the equivalent digraph $DG = (V, AG)$ and we put $Q = AG$, i.e. the positions are in edges of G in a given direction of motion, and the vertices serve to transition between edges only. We speak about "**edge bound positions**" even though we model them as "arrow bound".

Exceptionally, these choices can be interposed, that is, we can put $Q = V$ (vertex bound) for the agent moving in $G = (V, E)$ and $Q = A$ (arrow bound) for the agent moving in $D = (V, A)$. However, these possibilities will not be in the centre of our focus.

3.6 The main problem we are focused on is the **identification of the agent's final state** by the controller. The controller manifests that fact via emitting the command "stop".

The simplest case is when it is "a priori" known that after emitting the sequence of commands $\boldsymbol{s} = c_1, \ldots, c_n$, the trivial controller will lead the executor to stop in the given position q independent of the initial position q_0. Then we say that $\boldsymbol{s} = c_1, \ldots, c_n$ is the **synchronizing sequence of commands**.

In more complicated cases, the controller observes the data of the passed vertices and edges or arrows and, on the basis of this, it identifies the current position and eventually decides to emit the command "stop".

3.7 Usually, there is given a **target** vertex or edge or arrow to the controller at the beginning. Then the controller compares the identifications with the target and in the case of their being equal emits "stop".

We have to note that this is exactly characterized if the target is of the same type as the position, e.g. both are vertices or both are arrows, etc. And, fortunately, this is the main case. The situation when the target is an arrow and positions are vertices is practically impossible. Thus, we must say what happens when the target is a vertex v but the agent's positions are in arrows, or in edges in given directions, which is nearly the same. Then the desired final positions of the agent are all of ending in the vertex v. For example, in the graph in Fig. 1, the target vertex x implies the desired final positions vx, yx and zx.

If the induced path passes or ends in the target for the given sequence $s = c_1, \ldots, c_n$ of commands and for any initial position, then we say that the sequence s is **target finding**.

An agent can **observe** some data characterizing its position.

The notion of observability in graphs was studied e.g. in [11]. Our point of view is the following:

3.8 In the sequel we shall distinguish the following data, assumed to be available (i.e. "observable" by the agent) for each vertex $v \in V$:

D0 Beginnings of outgoing edges or arrows and their colours (if coloured). Moreover, if it is not the first position of the agent, then the end of the edge/arrow the agent comes in the vertex v belongs here as well.

D1 Degree $d(v)$ of the vertex v in the case of (non-oriented) graph.

D2 Number $d(v, c)$ of incident edges of the vertex v with the colour c in the case of (non-oriented) graph.

D3 Out-degree $d^+(v)$ of the vertex v in the case of digraph.

D4 Ends of ingoing arrows, their colours (if coloured) and in-degree $d^-(v)$ of the vertex v.

D5 Number $d^+(v, c)$ of outgoing arrows from the vertex v with the colour c in the case of digraph.

D6 Number $d^-(v, c)$ of ingoing arrows to the vertex v with the colour c in the case of digraph.

D7 Counter clockwise or clockwise cyclic ordering of incident edges or arrows with a vertex in the case of planar or locally planar graphs.

D8 Mark $\mu(v)$ of the target vertex v.

3.9 As concerns edges or arrows, we shall distinguish the following data, assumed to be available for each edge $e \in E$ or each arrow $a \in A$:

D9 Colour $\gamma(e)$ or $\gamma(a)$ if it is coloured.

D10 Orientation (= the allowed direction of moving) of a.

D11 Mark $\mu(e)$ of the target edge e or $\mu(a)$ of the target arrow a.

3.10 If positions of the agent are vertex bound then each position may be characterized by data selected from D1-D8. If they are edge bound then we add D9, D11 and for arrow bound we add D9, D10 and D11 to the previous ones. We say "add", not "replace", because the agent, moving through an edge or arrow, is able to observe data characterizing the final vertex before taking the decision concerning the next position.

Among the data D0-D10, we shall emphasise the ones that are needed in each case of agent motion we can meet. We call them **basic data**. If we deal with graphs and vertex bound agent positions, then D0, D1 and D2 belong to them; if it is edge bound then we add D9. If we deal with digraphs and vertex bound agent positions, then D0, D3 and D5 belong to them; if it is arrow bound then we add D9 and D10.

A bit beyond this frame are D4 and D6 for digraphs. If we add them to D0, D3 and D5, we speak about **extended basic data**. We note that in some applications it is not possible to observe that in a vertex there are ends of incoming arrows. Of course, extended basic data are observable, for example, in the major part of transportation networks.

On the other hand, in transportation networks and in some electric flex networks the assumption of local planarity is quite natural and then we can observe the data D7 there.

Finally, we can imagine many possible marks applied to vertices, edges or arrows. Their technical principles may vary very widely depending on the essence of graph or digraph in practice.

4 Problems Connected with the Identification of the Agent's Position

4.1 Main Problem

4.1.1 As we see from the title of the paper, our main task is to formulate and solve various identification problems of the current position of agents moving in networks. However, in connection with this, some "auxiliary" problems arise, for example, how to choose the orientation of a given graph not yet oriented, how to choose the colouring of a given graph or digraph if not yet coloured, etc. Even though these problems are interesting as well, we shall start with the **general position identification problem** – in short, **GPIP**: Disregarding the (unknown at the beginning) initial position of the agent, to identify its current position after performing a given sequence of commands. This identification can be

- **full**, i.e. the current position is unambiguously determined as a concrete $q_n \in Q$,

- **partial**, i.e. a subset $Q_n \subset Q$ is found (as small as possible) such that the current position belongs to it.

4.1.2 GPIP can be divided into the following sub-problems:

GPIP-1: To find the partial or full identification of the initial position q_0 of the agent at the beginning of its movement.

GPIP-2: If the initial position is not fully identified, i.e. $q_0 \in Q_0$, $|Q_0| > 1$, to find such sequence $\boldsymbol{s} = c_1,\ldots,c_n$ of input commands that the position $q_m \in \delta(\ldots\delta(Q_0, c_1),\ldots,c_m)$ is unambiguously identifiable for some $m \leq n$.

GPIP-3: Among the solutions $\boldsymbol{s}$ of **GPIP-2** to find the shortest one.

4.2 X/Y/Z Classification of Systems

4.2.1 The solution of an agent position identification problem strongly depends on the type of system, consisting of the agent and the network (i.e. graph or digraph). We shall classify them using three symbols *X/Y/Z* expressing the following features:

X) **type of network and position of agent**: Da – digraph, position in arrows (= "arrow bound"), Dv – digraph, position in vertices, Ne/Nv – graph (non-oriented), position in edges/vertices respectively,

Y) possibility of **additional observations** beyond basic data (see 3.8-3.10): B – basic data only, E – extended basic data in digraphs, L – locally planar graph/digraph with cyclic ordering of edges in vertices, Tv – target vertex marking, Te – target edge marking,

Z) occurrence of **edge colouring**: C – coloured, N – non-coloured.

For instance, the class Ne/Tv/C contains full or partial identification of marked target vertex in non-oriented coloured graphs with positions of the agent in edges.

4.2.2 Dv/B/C represents a deeply explored and well known case of synchronizing colouring [5], [14], [19]. It concerns digraphs with proper colouring, i.e. no outgoing arrows from the same vertex have equal colours. The command "c" determines the colour of the arrow to the next position, hence $C = \Gamma$. The trivial controller subsequently emits the "a priori" given sequence $\boldsymbol{s} = c_1, \ldots, c_n$. The **sequence** $\boldsymbol{s}$ is called **synchronizing** if $|\delta(V, \boldsymbol{s})| = 1$. If there exists a synchronizing sequence then the **colouring** is called **synchronizing** as well.

These notions are important mainly for the case when each vertex has the same outdegree $|\Gamma|$ i.e. for each vertex and each colour there exist exactly one outgoing arrow having that colour.

4.2.3 We admit more than one symbol in the second position, e.g. Dv/BTv/C means that both basic data and target vertex mark are observable.

4.3 Auxiliary Problems

It may happen that the given network is modified before the agent starts its motion in it. In our case, the goal of such modifications is to facilitate the consecutive identification of moving agent's position.

We need to emphasise that the problems mentioned henceforth will be fully formulated and able to be solved only if the *X/Y/Z* type of system is specified.

4.3.1 One of the auxiliary problems is the following: Given a locally planar connected graph $G = (V, E)$. To find a digraph $D = (V, A)$ derived from G such that it is synchronizing with respect to the commands of the type "c" where c is a positive integer and the command means "when you come to the vertex w by the arrow vw, then continue through the c^{th} outgoing arrow, counting in clockwise sense from q".

4.3.2 **Another** auxiliary problem starts with a digraph $D = (V, A)$ and a set of colours Γ. Then one has to find a synchronizing colouring of D in the system Dv/B/C. This is the well known **road colouring problem** – see e.g. [1], [19] studied mainly in the case of a constant outdegree of vertices in D.

5 Applications and Connection with Other Theories

5.1 Finite Automata

5.1.1 As we remember from the "classic" publications, a **finite automaton** (of the Mealy type) is defined as a quintuple $\mathcal{A} = (Q, X, Y, \delta, \lambda)$ where Q is a finite set of internal states, X is a finite set of inputs, Y is a finite set of outputs, further $\delta: Q \times X \to Q$ is a transition mapping determining the next state, i.e. $q_{n+1} = \delta(q_n, x_{n+1})$, $\lambda: Q \times X \to Y$ is an output mapping determining the next output, i.e. $y_{n+1} = \lambda(q_n, x_{n+1})$. We note that in more recent papers this notion is often named the "**finite state transducer**".

5.1.2 The automaton $\mathcal{A}$ can be represented by a "coloured" digraph $D = (Q, A)$ with the "colour" set $\Gamma = X$ and auxiliary marking with the mark set Y. The arrow $pq \in A$ with the colour $\gamma(pq) = x$ and the mark $\mu(pq) = y$ if and only if $q = \delta(p, x)$ and $y = \lambda(p, x)$.

We see that the colouring of the digraph D is proper and the outdegree of each vertex equals to the number of colours $|X|$.

Note that D may be a multigraph; there may exist more than one arrow connecting two adjacent vertices p, q. Of course, they are distinguished by their colours.

5.1.3 We can say that the question of final state identification is almost as old as the finite automata theory itself. As a "black box", the automaton does not allow for the observation and to identification of the internal state directly. In late 50's, e.g. in [9], there were studied so called **experiments**, i.e. observations of output sequence $\boldsymbol{o} = y_1, \ldots, y_n$ corresponding to a given input sequence $\boldsymbol{s} = c_1, \ldots, c_n$ in order to identify the final state. Saying by the "moving agent through network" language, the agent receives the command sequence $\boldsymbol{s} = c_1, \ldots, c_n$, passes through a path observing the marks $\boldsymbol{o} = y_1, \ldots, y_n$ and identifies the final vertex. If there is a target final vertex then we say the sequence $\boldsymbol{s} = c_1, \ldots, c_n$ is target finding.

5.1.4 A decade later, in [3], the notion of synchronizing sequence $\boldsymbol{s} = c_1, \ldots, c_n$ was introduced using the name "directing word". There $|\delta(Q, \boldsymbol{s})| = 1$, i.e. the final state is unambiguously determined without any observation. That means, the automaton is thought as a triplet $\mathcal{A} = (Q, X, \delta)$, the corresponding coloured digraph has no marks and the synchronizing property enables identification of the final state anyway. The minimal length of the synchronizing sequence $\boldsymbol{s}$ was also studied and estimated, as seen e.g. in [10] and [17].

5.2 Transportation

The authors are convinced that **transportation networks** represent the most important application field of the problems connected with identification of an agent's position in networks. Of course, the moving agents are persons in this case. Primarily, we shall consider road networks, bicycle path networks and pedestrian or hiking path networks. One can be sure that these studies have practical sense in spite of GPS devises for several reasons. Somewhere or sometimes there may be problems with signal, some area is not depicted well in GPS and mainly, many agents (drivers, cyclists, hikers) do not possess GPS.

Dealing with diverse transportation networks we can meet many variants of our main problem described in 4.1 and not less auxiliary problems presented in 4.3. The systems we shall deal with may belong to many classes in our triple X/Y/Z classification.

5.2.1 Example: Let $G = (V, E)$ be the graph from Fig. 1 representing a very simple network of cycle paths. Let us suppose that the moving agent's name is Clark. If it is a system of the type Nv/B/N, even though Clark can distinguish the vertices in according to the degrees, $d(v) = d(z) = 2$, $d(x) = d(y) = 3$, he has **no chance to identify the current state;** neither will he follow the simple rule represented by a sequence of commands $\boldsymbol{s} = c_1, \ldots, c_n$ nor a more complex rule represented by a controller $\mathcal{A}' = (Q', X, C, \delta', \lambda', q_0')$. The vertices v and z are undistinguishable observing basic data only. The same remains true even in the cases Nv/BL/N and Ne/BL/N as one can verify easily.

5.2.2 Example continuation: On the other hand, the systems Ne/BLTv/N and Ne/BLTe/N enable to find "target finding" sequence $\boldsymbol{s} = c_1, \ldots, c_n$.

In the graph $G = (V, E)$ from Fig. 1 we have $Q = \{vx, vy, xv, xy, xz, yv, yx, yz, zx, zy\}$. Since the graph is locally planar the command set may be $C = \{\text{stop}, 1, 2, 3\}$ where $m = 1, 2, 3$ means: "take the m^{th} exit anticlockwise. If the marked vertex is x then the target x finding sequence of commands is 1, 1, 1:

- The initial positions *vx, yx* and *zx* have the observable marked target vertex *x* immediately **at the end of the edge**.
- The initial positions *vy, yv* and *yz* have the observable marked target vertex *x* **at the end of the next position after applying the command** "1".
- The initial positions *xv, xy* and *zy* have the observable marked target vertex *x* **at the end of the reached position after applying the subsequence** "1,1", e.g. $xy \to yz \to zx$.
- The initial position *xz* has the observable marked target vertex *x* **at the end of the reached position after applying the sequence** "1, 1, 1": $xz \to zy \to yv \to vx$.

We have just shown that regardless of the initial position, applying the sequence "1, 1, 1" certainly enables the agent to pass through the target vertex *x*.

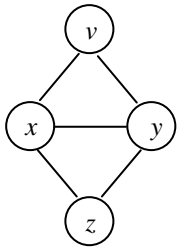


Figure 1
Original graph

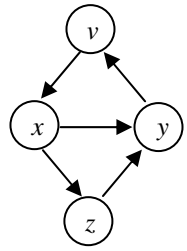


Figure 2
Derived digraph

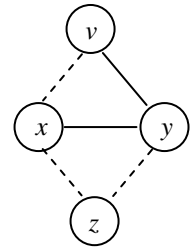


Figure 3
Coloured graph

If the marked target position is *vx* then the the target finding sequence of commands is 1, 1, 1, 2, 1, 1, 1, 2, 1, 1 as can be easily seen.

Real world interpretation. Let us suppose that the graph from Fig. 1 represents a cycle path network in a forest. Clark knows that his friend Xanthia is waiting for him sitting at a vertex (junction) *x*. He comes to the forest from outside and finds a path not knowing where he is and which direction to go in. Following the command sequence "1, 1, 1" he is sure that he will find Xanthia at the latest at the third passed vertex.

5.2.3 Example. We can ask whether the given graph $G = (V, E)$ may be replaced by a digraph $D = (V, A)$ derived from the graph G (see 2.4) in such a manner that D is synchronizing as a locally planar graph. That is, we consider the agent moving in D as a system Da/BL/N.

For the graph G from Figure 1 the answer is positive, the derived digraph D is in Figure 2. It is easy to see that the sequence $\boldsymbol{s}$ = 2, 1, 2, 2, 2, 1, 2, 2, 1, stop synchronizes the digraph D into the position $q = vx$. E.g. xz 2→ zy 1→ yv 2→ vx 2→ xy 2→ yv 1→ vx 2→ xy 2→ yv 1→ vx stop.

Real world interpretation. Let us suppose that the supervisor of the cycle network is considering converting all paths to a one-way type. Existence of a synchronizing sequence of commands may be a welcome feature.

That was "a priori" a given sequence $\boldsymbol{s}$ applied by a trivial controller. Let us consider a non-trivial controller $\mathcal{A}' = (Q', X', C, \delta', \lambda', q_0')$ where $Q' = \{p, r\}$, $X' = \{1, 2\}$, $C = \{1, \text{stop}\}$, $\delta'(p, 1) = p$, $\delta'(p, 2) = \delta'(r, 1) = \delta'(r, 2) = r$, $\lambda'(p,1) = 1$, $\lambda'(p, 2) = \lambda'(r, 1) = \lambda'(r, 2) = \text{stop}$, $q_0 = p$. When the agent's position is uw, then the current input x' of the controller is equal to the value of outdegree $d^+(w)$. Hence, the final position vx is reached by the agent in at most 3 steps; e.g. $q_0 = xy$, $q_0' = p$, $x_1' = d^+(y) = 1$, $q_1' = \delta'(p, 1) = p$, $c_1 = \lambda'(p,1) = 1$, $q_1 = \delta(xy, 1) = yv$, $x_2' = d^+(v) = 1$, $q_2' = \delta'(p, 1) = p$, $c_2 = \lambda'(p,1) = 1$, $q_2 = \delta(yv, 1) = vx$, $x_3' = d^+(x) = 2$, $q_3' = \delta'(p, 2) = r$, $c_3 = \lambda'(p,2) = \text{stop}$, $q_3 = \delta(vx, \text{stop}) = vx$ (repeated, i.e. the agent stops there).

We note that the vertex y is 2-merging (see 2.2). Obviously, existence of such a vertex is a necessary condition for the synchronizing property.

5.2.4 Example. We can put another question then in 5.2.3. We shall ask whether the given graph $G = (V, E)$ may be coloured in such a way that it becomes synchronizing as a locally planar graph. That is, we consider the agent moving in coloured G as a system Ne/BL/C.

Real world interpretation. Let us suppose that the supervisor of the cycle network is considering "colouring" the paths in the network, distributing coloured spur stones. Existence of synchronizing sequence is welcome as well.

For the graph G from Fig. 1 the answer is positive, the coloured graph G is in Fig. 3. Dashed lines represent the colour "b" (blue), solid lines the colour "r" (red). The possible commands we shall use are the following: 1r (first red to the right, i.e. in a counter clockwise direction), 2r, 1b (first blue to the right) and 2b. The set of all positions is $Q = \{vx, vy, xv, xy, xz, yv, yx, yz, zx, zy\}$. It is easy to prove that $\delta(q_0, \boldsymbol{s}) = xy$ for the sequence $\boldsymbol{s}$ = 1r, 1b, 1r, 1b, 1r, 1r, 1b, 1r and for each $q_0 \in Q$. E.g. $q_0 = xz$ 1r→ xz 1b→ zy 1r→yv1b→ vx 1r→ xy 1r→ yv 1b→ vx 1r→ xy.

We can note that having the desired final vertex x, the sufficient synchronizing sequence is $\boldsymbol{s}$ without the last "1r", i.e. 1r, 1b, 1r, 1b, 1r, 1r, 1b.

Similarly as in example 5.2.3, it is possible to look for some non-trivial controller here as well. Its construction can result from the fact that all vertices are distinguishable by their cyclic ordering of the colours of incident edges: $v \sim$ b-r, $x \sim$ b-b-r, $y \sim$ b-r-r, $z \sim$ r-r. Hence we can put $X = \{v, x, y, z\}$ as observable input data for the controller $\mathcal{A}' = (Q', X, C, \delta', \lambda', q_0')$. The other components of $\mathcal{A}'$ can be defined in many ways, e.g. $Q' = \{q_t, q_u, q_v, q_x, q_y, q_z\}$, $C = \{1r, 1b, stop\}$, $q_0' = q_u$. The transition and output mappings are defined only partially; the other "values" are irrelevant: $\delta'(q_t, y) = q_t$, $\lambda'(q_t, y) = stop$, $\delta'(q_u, v) = q_v$, $\lambda'(q_u, v) = 1b$, $\delta'(q_u, x) = q_x$, $\lambda'(q_u, x) = 1r$, $\delta'(q_u, y) = q_y$, $\lambda'(q_u, y) = 1b$, $\delta'(q_u, z) = q_u$, $\lambda'(q_u, z) = 1b$, $\delta'(q_v, x) = q_x$, $\lambda'(q_v, x) = 1r$, $\delta'(q_x, y) = q_t$, $\lambda'(q_x, y) = stop$, $\delta'(q_y, z) = q_z$, $\lambda'(q_y, z) = 1b$, $\delta'(q_z, x) = q_x$, $\lambda'(q_z, x) = 1r$. The digraph representing the partially defined controller is in Fig. 4. We shall show how the agent works if the starting position (= initial state of $\mathcal{A}$) is $q_0 = xz$:

1° $\mathcal{A}'$, starting in q_u, receives input z, remains in q_u and emits the command 1b. $\mathcal{A}$ receives it and transits to zy.

2° $\mathcal{A}'$ receives input y, transits to q_y and emits 1b. $\mathcal{A}$ receives it and transits to yz.

3° $\mathcal{A}'$ receives input z, transits to q_z and emits 1b. $\mathcal{A}$ receives it and transits to zx.
4° $\mathcal{A}'$ receives input x, transits to q_x and emits 1r. $\mathcal{A}$ receives it and transits to xy.
5° $\mathcal{A}'$ receives input y, transits to q_t and emits stop. $\mathcal{A}$ receives it and stops in xy.
$\mathcal{A}'$ receives input y remains in q_t and emits stop etc. The agent stops in the position xy "for ever".

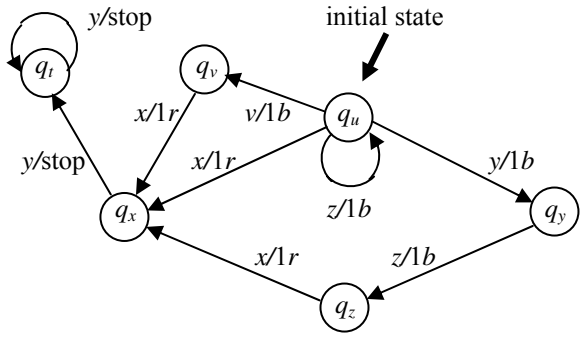


Figure 4
Finite transducer $\mathcal{A}'$

5.2.5 Example – continuation of 5.2.3. We have seen that the digraph in Fig. 2 possesses a synchronizing sequence if it is considered as a system Da/BL/N, i.e. when the agent's positions are arrow bound and each vertex has a cyclic ordering of outgoing arrows. Now we can ask whether there exists a colouring of arrows such that the coloured digraph becomes synchronized even in the system Dv/B/C. Yes, it exists (see Fig. 5); the synchronizing colouring exists with vertex bound

positions without any cyclic ordering in vertices. Of course, the set of commands $C = \{b, r\}$, "b" means blue, in Fig. 5 dashed lines, "r" means red, in Fig. 5 solid. The synchronizing sequence is $\boldsymbol{s} = r, b, b, r$. We must not forget that if the position is for example z, the command is b and from there no outgoing blue arrow exists, the agent remains in the same position. Hence his behaviour is identical as if there was a blue loop as in Fig. 6, where the loops were added and the digraph has constant outdegree 2 in each vertex. Colouring of such digraphs is the purpose of "road colouring" as described e.g. in [1] or [19].

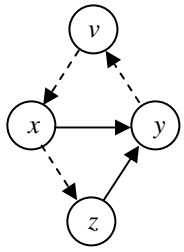


Figure 5
Synchronizing colouring

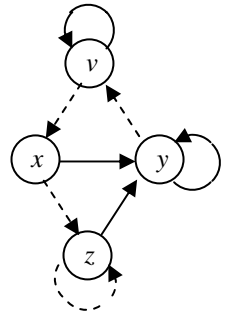


Figure 6
Constant outdegree

It is easy to see that the colouring from Fig. 3 is also synchronizing if applied to Fig. 2. The synchronizing sequence is very simple – b, b, b.

6 Target Reach in Systems Ne/BLTe/N and Ne/BLTv/N

As we have mentioned in 4.2.2, the case of oriented and coloured Dv/B/C systems have been deeply explored by many authors.

Now we turn our focus to the yet "untouched" types Ne/BLTe/N and Ne/BLTv/N. On one hand, they need not be coloured and oriented, but on the other hand, they must fulfil stronger constraints of local planarity and marked targets.

We shall present two theorems. The first one proves the existence of the desired sequence of commands leading to the target. The second one shows how to minimize the length of such a sequence.

6.1 Existence of Target Reaching Sequence of Commands

6.1.1 Theorem: Let $G = (V, E)$ be a finite connected locally planar graph with at least two vertices and with a marked edge (u, v). Let $DG = (V, AG)$ be the equivalent digraph with a marked arrow uv. Let $C = \{1, \ldots, d_{\max}\}$ be the set of commands, where $d_{\max} = \max\{d(w): w \in V\}$. Let us have $\delta(xy, c) = yz$ where yz is the c^{th} outgoing arrow from the vertex y for each $c \in C$ and $q = xy \in AG$. Then there exists a sequence of commands $\boldsymbol{s} = c_1, \ldots, c_n$ such that for each $q_0 = xy \in Q = AG$ the sequence (see 3.3) $\delta^*(xy, \boldsymbol{s})$ starts in uv or ends in uv or passes through uv.

Proof: First, we introduce the denotation $\boldsymbol{sr} = c_1, \ldots, c_n, c_{n+1}, \ldots, c_{n+m}$ for $\boldsymbol{s} = c_1, \ldots, c_n$, $\boldsymbol{r} = c_{n+1}, \ldots, c_{n+m}$ as the concatenation of sequences $\boldsymbol{s}$ and $\boldsymbol{r}$.

Since G is connected, the digraph DG is strongly connected and therefore there exists a path in DG from xy to uv for each $xy \in AG$ and therefore there exists a sequence of commands $\boldsymbol{s} = c_1, \ldots, c_n$ such that $\delta(xy, \boldsymbol{s}) = uv$.

1° Let us denote $Q_1 = AG - \{uv\}$. If $Q_1 = \varnothing$ we have finished, even for the empty sequence Λ we have $q \in AG \Rightarrow \delta(q, \Lambda) = uv$. If $Q_1 \neq \varnothing$ then we choose an element $q_1 = x_1y_1 \in Q_1$. Let $\boldsymbol{s}_1$ be the sequence of commands with the property $\delta(x_1y_1, \boldsymbol{s}_1) = uv$.

2° Let us denote $Q_2 = AG - \{uv, x_1y_1\}$. If $Q_2 = \varnothing$ we have finished since $\delta^*(q, \boldsymbol{s}_1)$ starts or ends in uv for all $q \in AG$. If $Q_2 \neq \varnothing$ then we choose an element $q_2 = x_2y_2 \in Q_2$. Let $\boldsymbol{s}_2$ be the sequence of commands with the property $\delta(\delta(x_2y_2, \boldsymbol{s}_1), \boldsymbol{s}_2) = \delta(x_2y_2, \boldsymbol{s}_1\boldsymbol{s}_2) = uv$.

3° Let us denote $Q_3 = AG - \{uv, x_1y_1, x_2y_2\}$. If $Q_3 = \varnothing$ we have finished since $\delta^*(q, \boldsymbol{s}_1\boldsymbol{s}_2)$ starts or ends in uv or passes through uv for all $q \in AG$. If $Q_3 \neq \varnothing$ then we choose an element $q_3 = x_3y_3 \in Q_2$. Let $\boldsymbol{s}_3$ be the sequence of commands with the property $\delta(\delta(\delta(x_3y_3, \boldsymbol{s}_1), \boldsymbol{s}_2), \boldsymbol{s}_3) = \delta(x_3y_3, \boldsymbol{s}_1\boldsymbol{s}_2\boldsymbol{s}_3) = uv$.

… etc. By induction we find a sequence $\boldsymbol{s} = \boldsymbol{s}_1\boldsymbol{s}_2\ldots\boldsymbol{s}_m$ having the desired property.

6.1.2 If we have a marked target vertex v instead of target arrow uv, because of connectivity of G there certainly exists an arrow $uv \in AG$ and we can use the sequence $\boldsymbol{s}$ from the theorem 6.2.1.

6.2 Minimization of Target Reaching Sequence of Commands

6.2.1 Theorem: Let $G = (V, E)$, $DG = (V, AG)$, uv and C be defined as in 6.1.1. Let $D\delta C = (AG, AG\delta)$ be the **transition digraph** corresponding to the digraph DG and the set of commands C (see 2.9) and let $D\delta CRuv = (2^{AG-\{uv\}}, AR)$ be the **reduced total digraph** derived from $D\delta C$ and uv (see 2.8). Let $\boldsymbol{p}$ be the shortest path from the vertex $AG-\{uv\}$ to the vertex $\varnothing$ in the graph $D\delta CRuv$ and let $\boldsymbol{s} = c_1, \ldots, c_n$ be the sequence of colours of the arrows generating the path $\boldsymbol{p}$. Then $\boldsymbol{s}$ is the

shortest sequence of commands such that for each $q_0 = xy \in Q = AG$ the sequence $\delta^*(xy, \boldsymbol{s})$ starts in uv or ends in uv or passes through uv.

Proof: 1) Let $\boldsymbol{s} = c_1, \ldots, c_n$ be such a sequence of commands that for each $q_0 = xy \in Q = AG$ the sequence $\delta^*(xy, \boldsymbol{s})$ starts in uv or ends in uv or passes through uv. Since

$$\delta^*(q_0, c_1, \ldots, c_n) = q_0, \delta(q_0, c_1), \delta(\delta(q_0, c_1), c_2), \ldots, \delta(\ldots(\delta(\delta(q_0, c_1), c_2), \ldots, c_n)$$

the sequence $\delta^*(xy, \boldsymbol{s})$ represents a path containing the vertex uv in the coloured transition digraph $D\delta C$.

Let the arrows of a path $\boldsymbol{r}$ starting in the vertex $AG-\{uv\}$ in the reduced total digraph $D\delta CRuv$ have the colours $c_1, \ldots, c_n$. Then the next vertices of the path are

$$\delta(AG-\{uv\}, c_1) = \{\delta(xy, c_1): xy \in AG-\{uv\}\}$$

$$\delta(AG-\{uv\}, c_1, c_2) = \{\delta(xy, c_1): xy \in AG-\{uv\}\}$$

… etc.

and at least one of them (and all after it, if exist) must be $\varnothing$, because for each $xy \in AG-\{uv\}$ the sequence$\delta^*(xy, c_1, \ldots, c_n)$ contains uv. Hence $\boldsymbol{r}$ is a path from the vertex $AG-\{uv\}$ to the vertex $\varnothing$ in the digraph $D\delta CRuv$.

2) Conversely, if $\boldsymbol{r}$ is a path from the vertex $AG-\{uv\}$ to the vertex $\varnothing$ in the graph $D\delta CRuv$ and $c_1, \ldots, c_n$ are the colours of its arrows, then the sequence $\delta^*(xy, c_1, \ldots, c_n)$ must contain uv.

This concludes the proof; the shortest path $\boldsymbol{p}$ represents the shortest sequence of colours = commands $\boldsymbol{s} = c_1, \ldots, c_n$ of the desired properties.

6.2.2 Example. Let us return to the example 5.2.1 (the graph from Fig. 1), to the second part, when the marked target position is vx. The target finding sequence of commands is 1, 1, 1, 2, 1, 1, 1, 2, 1, 1 (the stop at the end is not important). In the digraph $D\delta CRuv$ we have

$AG-\{vx\} = \{vy, xv, xy, xz, yv, yx, yz, zx, zy\}$ $1\rightarrow$ $\{vy, xv, xy, yv, yx, yz, zx, zy\}$ $1\rightarrow$

$1\rightarrow$ $\{vy, xv, xy, yv, yx, yz, zx\}$ $1\rightarrow$ $\{vy, xv, xy, yx, yz, zx\}$ $2\rightarrow$ $\{xv, xz, yv, yz, zy\}$ $1\rightarrow$

$1\rightarrow\{vy, yv, zx, zy\}$ $1\rightarrow\{xy, yv, yx\}$ $1\rightarrow\{xv,yz\}$ $1\rightarrow\{zy\}$ $1\rightarrow\{yv\}$ $1\rightarrow \varnothing$

Real world interpretation. The supervisor of the network can use theorem 6.2.1 for two purposes:

- ascertaining that the network posses a synchronizing sequence
- finding the shortest one.

6.2.3 Theorem 6.2.1 solves the problem of marked target finding in the case Ne/BLTe/N when the target is in "arrow" uv. As concerns the system Ne/BLTv/N with a marked target vertex v, it is enough to replace the set $AG-\{uv\}$ by the set $AG-\{xv: xv \in AG\}$ in the reduced total digraph.

Conclusion, Further Perspectives

We have shown that the position identification of agents moving in networks possesses several important features:

- It represents a specific and interesting issue within the theory of agent based systems.
- It has interesting practical applications. Among these, mainly the transport ones are promising.
- It demonstrates interesting links between three branches of discrete systems – finite transducers, coloured graphs and transportation.

Moreover, the authors hope that this paper can stimulate further studies oriented mainly to the following directions:

D1 Generalizations, namely

- towards higher cleverness of agents,
- concerning random environment - e.g. the colouring of roads perhaps partially damaged, etc.,
- extending the number of agents moving in the same network.

D2 New methods and algorithms solving concrete versions of the main problem or of the auxiliary one.

References

[1] Adler, R., Goodwin, I., Weiss, B.: Equivalence of Topological Markov Shifts, in Israel. J. Math 27 (1977), pp. 49-63

[2] Bonabeau, E.: Agent-based Modeling: Methods and Techniques for Simulating Human Systems. In Proc.National Academy of Sciences 99 (2001) No. 3, pp. 7280-7287

[3] Černý, J.: Poznámka k. homogénnym experimentom s konečnými automatmi, Mat. fyz. čas SAV 14 (1964), pp. 208-215 (in Slovak)

[4] Cicortas, A., Iordan, V.: A Multi-Agent Framework for Execution of Complex Applications. Acta Polytechnica Hungarica 3 (2006), No. 3, pp. 97-119

[5] Culik, K. II, Karhumäki, J., Kari, J. A.: Note on Synchronized Automata and Road Coloring Problem, in: Lecture Notes in Computer Science 2295 (2002), pp. 175-185

[6] Erol, K.: A Study of Agent-based Traffic Simulation", Final Report, FHWA, US DOT, 1998

[7] Frankovič, B., Dang,, Tung-T., Budinská, I.: Agents' Coalitions Based on a Dynamic Programming Approach. Acta Polytechnica Hungarica 4 (2007), No. 2, pp. 5-21

[8] Fukuda, T., Hosokai, H., Otsuka, M.: Path Planning and Control of Pipeline Inspection and Maintenance Robot, in IECON'88, Proc. 14th Annual Conf. (1988) pp. 38-43

[9] Ginsburg, S.: On the Length of the Smallest Uniform Experiment which Distinguishes the Terminal States of a Machine, in J. Assoc. Comput. Mach 5 (1958) pp. 266-280

[10] Göhring, W.: Minimal Initializing Word: A Contribution to Cerny's Conjecture, in Journal of Automata, Languages and Combinatorics 4 (1997) pp. 209-226

[11] Jungers, R. M., Blondel, V. D.: Observable Graphs, in arXiv:cs/0702091v1 [cs.MA] 16 Feb 2007

[12] Kelemen, J.: Agents from Functional-Computational Perspective, in Acta Polytechnica Hungarica 3 (2006) No. 4, pp. 37-54

[13] Kranakis, E., Santoro, N., Sawchuk, C. and Krizanc, D.: Mobile Agent Rendezvous in a Ring, Proceedings of the 23rd International Conference on Distributed Computing Systems, May 19-22, 2003, p. 592

[14] Mateescu, A. and Salomaa, A: Many-valued Truth Functions, Cerny's Conjecture and Road Coloring, in EATCS Bulletin 68, 1999, pp. 134-150

[15] North, M. J., Macal, C. M.: Managing Business Complexity: Discovering Strategic Solutions with Agent-based Modeling and Simulation, Oxford University Press: New York, NY, 2007 ISBN 0195172116

[16] Roman, A., Foryś, W.: Lower Bound for the Length of Synchronizing Words in Partially Synchronizing Automata, in Sofsem 2008 (W. Geffert et al. eds.), Springer Verlag Berlin, Heidelberg 2008, pp. 448-459

[17] Rystsov, I. K.: Reset Words for Commutative and Solvable Automata in Theoret. Comput. Sci. 172 (1997), pp. 273-279

[18] Sebestyénová, J.: Case-based Reasoning in Agent-based Decision Support System. Acta Polytechnica Hungarica 4 (2007), No. 1, pp. 127-138

[19] Trahtman, A. N.: The Road Coloring Problem, in arXiv:0709.0099 [es.DM] 21 Dec 2007

[20] Vaščák, J.: Navigation of Mobile Robots Using Potential Fields and Computational Intelligence Means. Acta Polytechnica Hungarica 4 (2007) No. 1, pp. 63-74

Developing Stability Control Theories for Agricultural Transport Systems

Tamás Szakács

Institute of Mechatronics and Autotechnics
Bánki Donát Faculty of Mechanical Engineering and Security Technology
Óbuda University
Népszínház u. 8, H-1081 Budapest, Hungary
E-mail: szakacs.tamas@bgk.uni-obuda.hu

Abstract: The design and management of agricultural transport systems is facing a series of challenges. The goals of increasing vehicle mobility and reducing soil stress have always been in the center of attention. Involving trailer wheels in the production of pulling force could be beneficial in both cases. In this way, the mobility of the transport system could be secured even in heavy soil conditions, and the soil damage could be reduced as well. Despite the potential advantages, trailer protrusion is not currently practical. The reason for this is the lack of proper safety measures. The origin of this problem is the force generated by the trailer protrusionacting on the drawbar, which can bring the tractor into an unstable state, causing the tractor to roll over or the vehicle train to jack-knife. To avoid such accidents, a control system must be developed which can recognize the beginning of unstable vehicle behavior and either by warning or intervention help to maintain the stable state of the vehicle.

The goals of this paper can be summarized as follows:

- *To introduce a computer simulation model of an agricultural transport vehicle-park consisting of different dynamical models of tractors and trailers. The aim of this model-park is to perform stability analysis of tractor – propelled axle trailer vehicle combinations.*
- *To introduce stability criteria equations for future stability control programs which can determine different unstable states, and by interaction can re-stabilize the vehicle combinations.*

Keywords: Vehicle-train; stability; agricultural transport systems

1 Introduction and Motivation

The subject of this paper is the stability analysis of transport vehicle combinations consisting of agricultural tractors and propelled axle trailers, and the reduction of accidents related to the trailer drive. Special attention has been paid to the accidents caused by the trailer protrusion. In order to achieve the stability increase, I have developed a computer model of a transport vehicle park, which is capable of performing dynamical simulations of transport vehicle trains and which can be the basis for developing stability control theories.

Propelled axle agricultural trailers, and the related problems:

Tractors involved in heavy agricultural transporting tasks are ballasted by extra weights in order to make it possible to produce the pulling force which is required to move the trailer in heavy soil conditions. In my opinion instead of ballasting the tractors, it could be more beneficial to involve the wheels of the trailer in the force development to move the train. This could secure the mobility of the transport system in heavy soil conditions, and the soil damage could be reduced as well.

Involving the wheels of the trailer in the production of the pulling force has the following possible advantages:

- Despite the general tendency towards increasing power in agricultural machinery, the weight of the tractor can be reduced.
- By means of optimized weight and driving torque distribution, soil damage, environmental pollution and operation costs can be reduced.
- By making the transport system more independent from soil conditions, harvesting losses can also be reduced.

Despite the potential advantages, because of the lack of safety measures, trailer protrusion is not part of practical life. The origin of the problem is that the trailer can generate a side force on the rear wheels of the tractor, which under certain conditions can bring the tractor into an unstable state. The trailer can cause the tractor to roll over or to jackknife.

The conclusion can be drawn, that the tractor-trailer vehicle combination can only be used when it is equipped with a control system that can avoid the trailer drive originated accidents to happen.

2 The Models Used for Developing the Stability-Programs

In this chapter the models used for the stability analysis of the agricultural transport vehicle trains and the methods of instability-determinations are briefly introduced. The models I have developed in Matlab/Simulink environment are created for this work. The models are developed in a modular system so that they can be used for the simulation of other types of vehicles in addition to agricultural transport systems. During the development phase, special expectations of agricultural transport systems have been taken into consideration.

2.1 The Vehicle Model Park Developed for the Task

In order to model agricultural transport systems, I have developed dynamic models of different tractors and trailers. From the model, vehicle combinations can be created as in the reality.

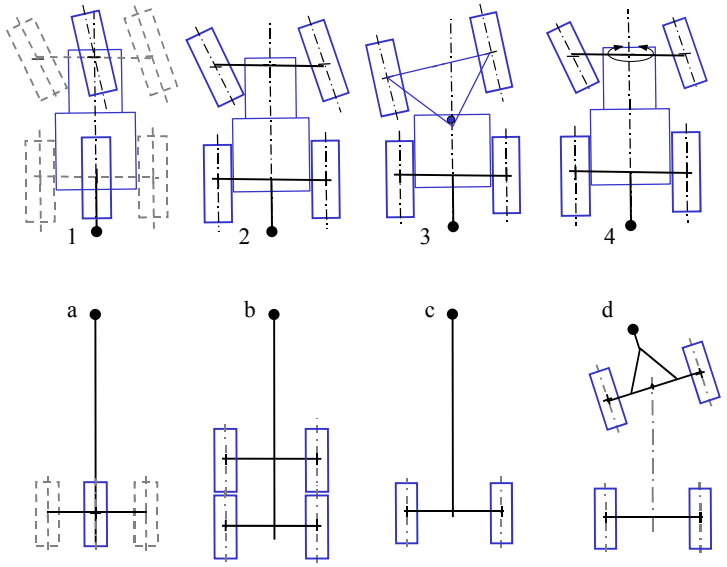


Figure 1
Vehicle model park

Figure 1 shows the modelled agricultural transport vehicle park. Available tractors are a single-track tractor (1), a system-tractor (2), a joint-tractor (3), and a common ploughing tractor (4). The trailers available are a single-track trailer (1), a semi-suspended tandem axle trailer (2), a semi-suspended single axle trailer (3), and a pulled trailer (4).

The models are created in a Matlab/Simulink environment and built in a modular system. This ensures that different sub-models can be changed within models, and new model units (like a trailer behind a car, a plough behind an agricultural tractor, or a different tractor) can be inserted in the vehicle model if the simulation goals require it.

I have created an animation-window, which helps via simplified schemas of the vehicles to follow the motion of the vehicle train. It has indications for the state of the stability control system, and helps to create animations for demonstration goals.

2.2 Determination of the Pulling Angle

Pulling angle: (Also tow angle (ProPride Inc.)) The horizontal component of the angle between the pulling and the pulled vehicle. Notation: γ (^{o} or rad)

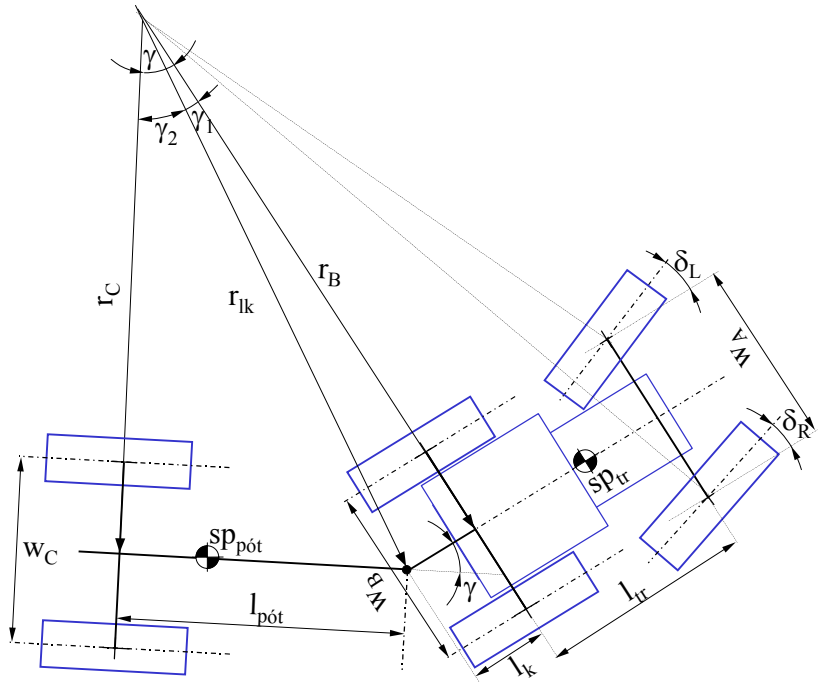


Figure 2
The extended Ackermann-condition of vehicle trains

One method of determining the state of the vehicle is based on measuring the pulling angle and comparing its value to the expected one. To determine the expected value of the pulling angle, I have introduced the extended Ackermann conditions of steered wheel angles to vehicle trains. *Figure 2* shows a tractor and a

trailer attached to it. The angles of the steered wheels (δ_L, δ_R) are calculated in the conventional manner of Ackermann's method (Ackermann, 1990).

The Ackermann condition of vehicle train is fulfilled when not only the axles of the wheels of the tractor but also the wheels of the trailers are pointing in the theoretical turning center (momentan centrum). The γ_{stat} pull-angle in the steady curving can then be calculated using the notations of Figure 2 as follows:

$$\gamma_{stat}=\gamma_1+\gamma_2 \tag{1}$$

where

$$\gamma_1 = \arctan\left(\frac{l_k}{r_B}\right) \text{ and} \tag{2}$$

$$\gamma_2 = \arcsin\left(\frac{l_{pót}}{r_{lk}}\right) = \arcsin\left(\frac{l_{pót}}{\sqrt{(l_k^2 + r_B^2)}}\right). \tag{3}$$

Equations (2) and (3) substituted in eq. (1):

$$\gamma_{stat} = \arctan\left(\frac{l_k}{r_B}\right) + \arcsin\left(\frac{l_{pót}}{\sqrt{(l_k^2 + r_B^2)}}\right) \tag{4}$$

The turning radius of the rear axle is:

$$r_B = \frac{l_{tr}}{\tan(\delta_L)} + \frac{w_A}{2} \tag{5}$$

The calculated steady state γ_{stat} pulling angle is equivalent only in steady curving with the actual pulling angle. The reason for this is that in contrast to the δ_L, δ_R steering angles of front wheel of the tractors, which immediately occur as the steering wheel is turned, the pulling angle is continuously changing as the vehicle moves. To reach its steady state value, the vehicle has to travel a certain distance. My goal was to set up a model or equation which describes the actual value of the pulling angle, not only in steady but also in transient state.

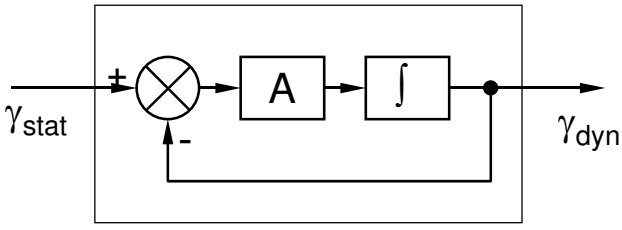


Figure 3

Determination of γ_{dyn} actual pull-angle based on the γ_{stat} steady-state pull-angle

Figure 3 shows the block scheme realization of the transient state pull-angle determination. In this case, the matrix *A* is simplified to scalar, and its value is the reciprocal of the time required by the rear axle of the trailer to reach the position of the rear axle of the tractor.

In equation:

$$A = \frac{v_x^{tr}}{l_{pót} + l_k}, \tag{6}$$

where v^{tr}_x is the speed of the tractor (in m/s), $l_{pót} + l_k$ is the distance of the axles of the tractor and the trailer (in meters).

The actual value of the pulling angle in a differential equation form:

$$\frac{1}{A}\frac{d\gamma_{din}}{dt} + \gamma_{din} = \gamma_{stat} \tag{7}$$

The solution of the equation depends on the input function (γ_{stat}). Using a step input function with the amplitude of γ_{stat}, the equation of γ_{din} output will be:

$$\gamma_{din} = \gamma_{stat}\left(1 - e^{-At}\right). \tag{9}$$

2.3 Stability Determination

The stability determining functions are logical criteria functions of which the value is 0 if the vehicle behaves as expected, and 1 if instability is detected.

The general form of the criteria equations is:

$$stab = \begin{cases} 0, if\ \exp ected < \exp ected_{\min} \vee measured \in \left[\exp ected_{low} \ldots \exp ected_{high}\right] \\ 1, if\ \exp ected > \exp ected_{\min} \wedge measured \notin \left[\exp ected_{low} \ldots \exp ected_{high}\right] \end{cases} \tag{10}$$

where:

- expected is the calculated value of the stability determining parameter.
- $expected_{min}$ is the threshold value of the expected parameter. If measured lower is expected, then no stability validation is done,
- $expected_{low}$ is the lower limit of the reference band used in the comparison of expected and measured values,
- $expected_{high}$ is the upper limit of the reference band used in the comparison of expected and measured values.

The expected values and the stability determining method can then be embedded in vehicle and Fuzzy logic controller design, and its application as an embedded system can be considered as well. (Farkas, Halász, 2006)

3 Validations by Field Measuring

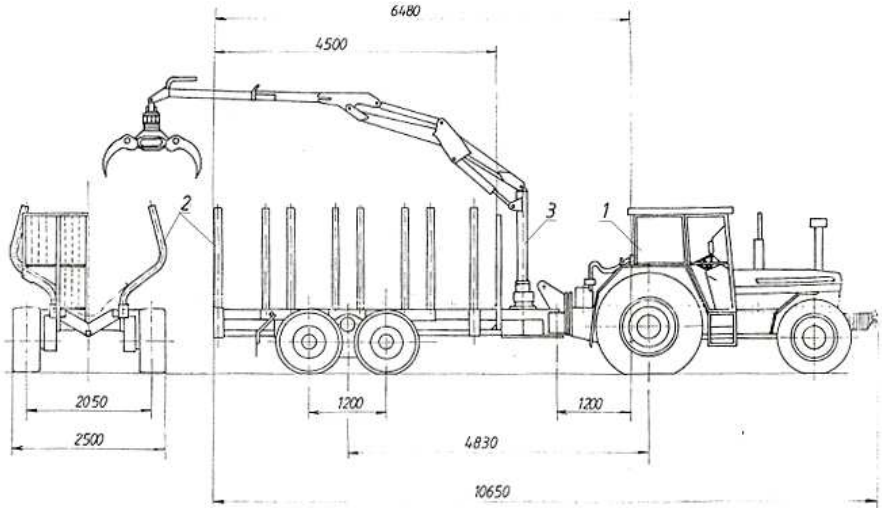


Figure 4
Schematic drawing of the measured tractor-trailer combination (Horváth, 1996)

3.1 Combined Braking and Steering Test

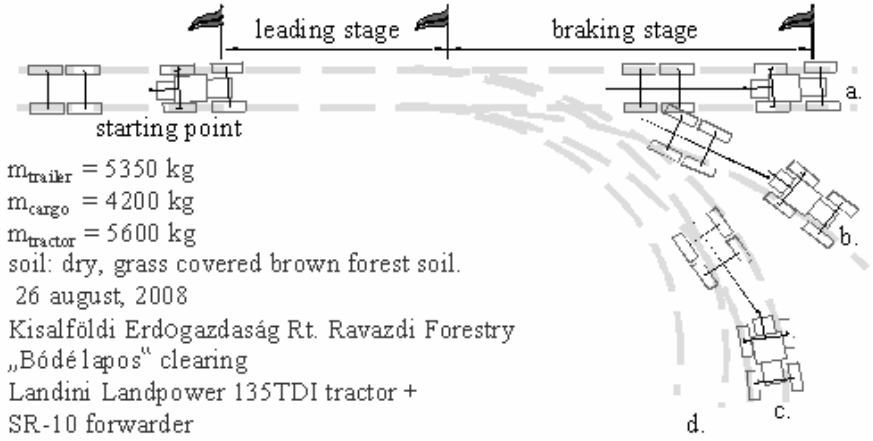


Figure 5
Measuring stages of the combined braking and steering test

I have done the validations using a vehicle combination consisting of a Hungarian made SR-10 type propelled axle forwarder attached to a Landini Landpower agricultural tractor.

The measurements were performed on the lands of the forest of Kisalföldi Erdőgazdaság Rt. at the Ravazdi Erdészet location. The measuring track is shown in the Figure 5. The steady-state stage is 30 metres long, and the braking stage is also 30 meters. The steering angles are from a to d 0, 7, 15, and 28 degrees (measured on the front left wheel).

The measurements were performed on a dry sunny day, with temperatures in the range of 20-23^{o}C. The soil was a brown forest soil, permanently maintained, grass covered clearing.

During the taking of the measurement, data were partially manually and partially electronically recorded. The measuring sets became a code for further pairing of the corresponding manually and instrumentally recorded data.

The vehicle-train, after a pre-stage travels, at the second stage either proceeded straight ahead or took a turn with the steering wheel turned to 7, 15 or 18 degrees (a-d variations in the Figure 5), while at the same time decelerating to still-stand by the smooth application of the brakes. The measurement is continued till the still-stand of the vehicle. During the measurement, the steering-wheel angle, the pulling angle, and the acceleration were recorded by a data recording system, and the number of the rotations of the wheels, the time required to pass stages 1 and 2, the speed of the tractor, and a trigger signal were recorded manually. Having completed the first series of measurements, two similar sets of measurements followed. In this way three repetitions were performed.

The goal of the test is to determine the accelerations, wheel slips and curving stability during braking and combined braking and steering.

3.2 Roundabout Test

One of the stability programs is based on measuring the pulling angle between the tractor and the trailer. The roundabout test is to validate the determination of the expected value of the pull angle in steady and transient states. The first stage is a lead up, while the second stage is a roundabout at the maximum steering angle of the tractor. Figure 6 shows the plan of the test and details. In this test only the trailer drive was applied.

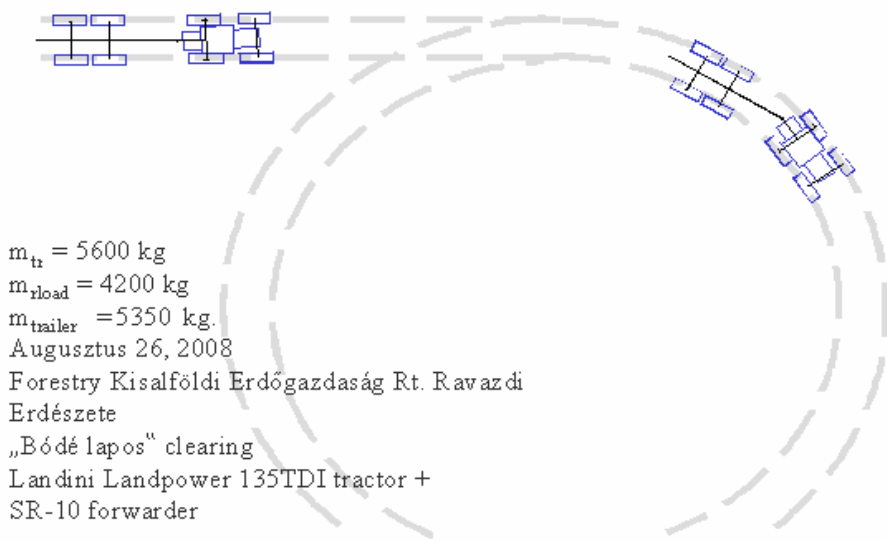


Figure 6
Plan and data of the roundabout test

3.3 Validation Test

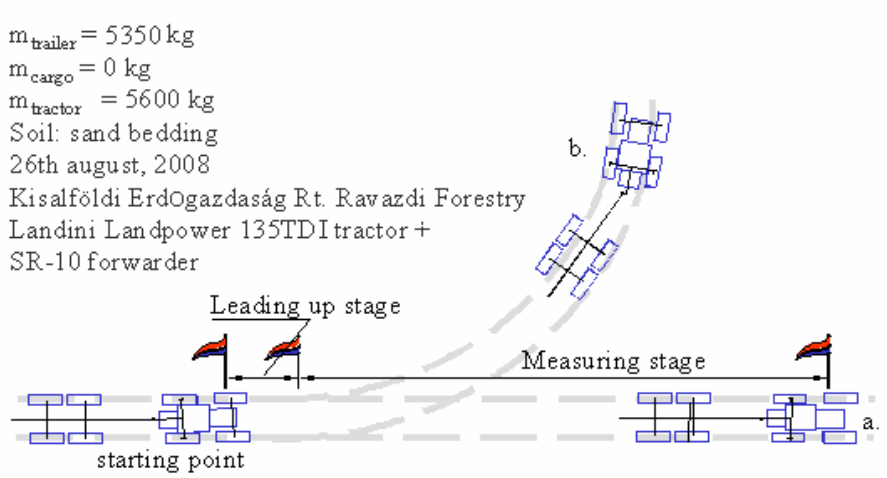


Figure 7
Plan and data of the validation test

The validation tests were performed in an empty barn, which was under construction. The barn had fine sand and a level bedding, which appeared ideal and easily reproducible. In the bedding I marked a short stage to achieve a near steady-state running of the vehicle combination, followed by either a 20-meter-long straight track, or a left turning one. The same data sets were recorded as in case of the combined braking and steering test. The goal of the measuring was to check the functionality of the measuring instruments, and to perform measurements to validate the pull-angle and yaw-rate determinations.

4 Processing the Measuring Data and Results

4.1 Evaluation of the Braking and Steering Test

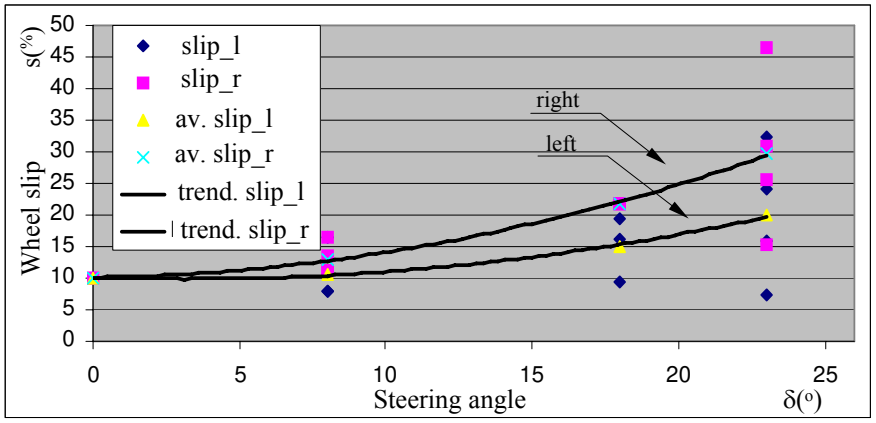


Figure 8
Wheel slips as function of steering angle

Figure 8 shows the dependency of the slip of the driven wheels as a function of the steering angle. The continuous line is a tendency curve drawn over the average calculated slip at the set steering wheel angles. It can be seen that the slip differentiates between inner and outer wheels in the respect of turning direction. The slip of the inner wheel is proportionally greater than the outer, depending on the steering angle. The averages of the slip values rise proportionally with the steering angle. The reason for this is that the travelling losses increase with the steering angle, and the increased losses indirectly cause increased slip values.

4.2 Evaluation of the Roundabout Test

The goal of the roundabout test is to bring the vehicle train to an extreme state in order to validate the stability program and the calculation method of the pulling angle. During the test only the trailer drive was used with the steering angle set at 28 degrees to the right.

Data recorded during the test were:

- Numbers of rotations of the trailer tires (n_b, n_j)
- Accelerations of the trailers body (a_{A1x}- a_{A4y})
- Steering angle (δ)
- Pulling angle (γ)
- Travel speed of the train (v_h)

Figure 5 summarizes the result of the validation of the pulling angle calculation.

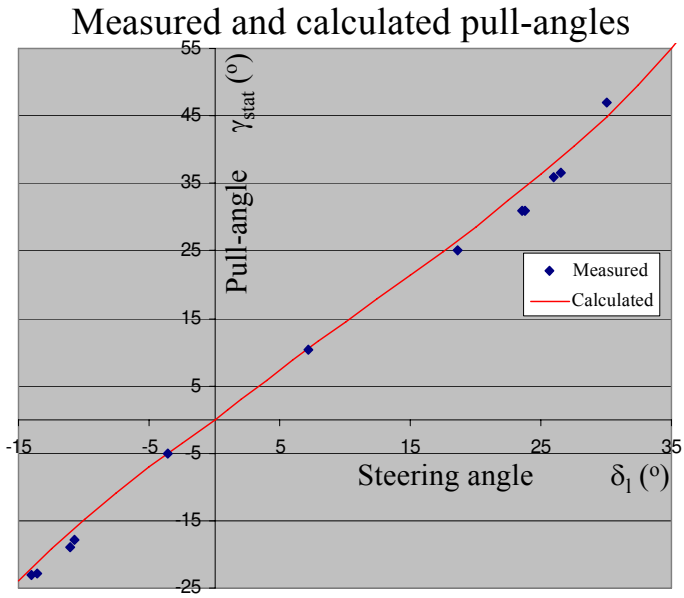


Figure 9
Comparison of measured and simulated pulling angle

Figure 9 shows the comparison of the calculated and the measured steady-state pulling angles. The continuous line shows the function of equation 4, and the squares are the date gained from the field measuring. This result validates the determination of the steady-state pull-angle determination.

From the recorded measurement data, the steering and pull angle data have been taken (Figure 10 upper part). I have created a computer model of the vehicle train and set the parameters in accordance with the measured vehicle. Using the recorded speed and steering data, I have reproduced the simulation with the model. The pull angle produced by the simulation then was compared with the measured one (Figure 10 lower part), after which after I draw the following conclusions. The model used for determining the transient state of the pulling angle gives satisfactory results; therefore, it can be used as the parameter of vehicle stability determination.

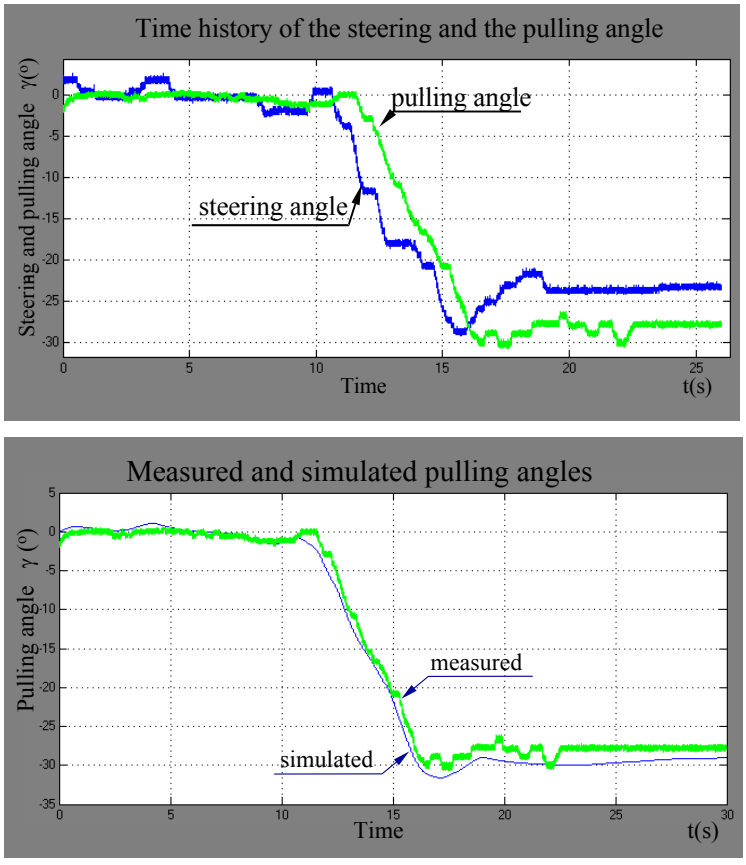


Figure 10
Comparison of measured and simulated pulling angle

Conclusions, summary

Stability determinations were developed using a self-developed model park of agricultural transport vehicles. Among the stability determination methods, the pulling angle based determination proved to be the most accurate, and the easiest to measure its determination parameter. The pulling angle determination model was validated by a series of field measurments. The expected pulling angle can be compared with the measured one, and from the deviation the stability of the vehicle train can be determined. The advantage of this method is that the pull angle deviation can be caused both by the pulling vehicle and the trailer; in this way abnormal behavior of either of them can be determined.

Acknowledgments

The author would like to express his appreciation to the Forestry of of Kisalföldi Erdőgazdaság Rt. for providing the vehicles, area and the crew for the measuring, and prof. Pérer Szendrő and Lajos Laib for the support, as well as the leaders of Óbuda University for providing the working conditions.

References

[1] Towing Definitions on the ProPride Inc. home page. http://www.propridehitch.com/trailer_sway/towing_definitions.html downloaded at 02.01.2010

[2] Mitschke, M: Dynamic der Kraftfahrzeuge. 2. Auflage, Band C, Berlin Springer 1990

[3] Ferenc Farkas, Sándor Halász: Embedded Fuzzy Controller for Industrial Applications, Acta Polytechnica Hungarica, Vol. 3, No. 2, 2006, pp. 41-63

[4] Horváth Béla (Editor): SR-8 kihordó gépesítési információ Panax Kft. Nyomdaüzeme, 1996, p:30

The Numerical Modeling of Transient Regimes of Diesel Generator Sets

Tiberiu Tudorache, Cristian Roman

Electrical Engineering Faculty, University Politehnica of Bucharest
313 Splaiul Independentei, Sect. 6, Bucharest, Romania
e-mail: tudorach@amotion.pub.ro

Abstract: This paper deals with the numerical modeling of a diesel generator set used as a main energy source in isolated areas and as a back-up energy source in the case of renewable energy systems. The numerical models are developed using a Matlab/Simulink software package and they prove to be a powerful tool for the computer aided design of complex hybrid power systems. Several operation regimes of the equipment are studied. The numerical study is completed with experimental measurements on a Kipor type diesel-electric generator set.

Keywords: Diesel generator set; numerical models; experimental analysis

1 Introduction

The increasing demand for energy, the continuous reduction in existent resources of fossil fuels and the growing concern regarding environmental pollution have compelled mankind to explore new production technologies for electrical energy using clean renewable sources such as wind energy, solar energy, etc.

Among the electric power technologies using renewable sources, those based on the conversion of solar and wind energy are clean, silent and reliable, with low maintenance costs and small ecological impact. Sunlight and the kinetic energy of the wind are free, practically inexhaustible, and involve no polluting residues or greenhouse gases emission. Along with these advantages, however, electric power production systems using as primary sources exclusively solar and wind energy pose technical problems due to uncontrollable wind speed fluctuations and to the day – night and summer – winter alternations. As a consequence, in autonomous regimes, the power supply continuity of a local grid should be backed-up by other reliable and non-fluctuant sources of primary energy, such as diesel generator sets. Such systems, designed for the decentralized production of electric power using combined sources of primary energy, are called hybrid systems.

By combining several renewable energy sources (wind/PV/diesel), a hybrid system may ensure increased reliability in the electric power supply to consumers, at optimum quality parameters, in the condition of a minimum required electric power during an imposed time period.

The increased interest in using diesel generator sets as the main energy source in isolated areas or as an emergency source in the case of renewable-based power systems can be observed by the great number of papers and studies carried out in this area. The research conducted in this domain refers to aspects such as: island operations of diesel generator sets [1], simulations of diesel/pv/wind hybrid power systems [2]-[3], numerical modelings using special regulation techniques such as neural networks and fuzzy logic [4]-[6], etc. Various aspects about these regulation techniques can be found also in [7]-[8].

The topic proposed in this paper refers to the development of numerical models for the simulation of the operation regimes of diesel generators integrated as a back-up energy source in hybrid power systems as back-up energy source. The numerical analysis is completed by experimental investigations on a Kipor type diesel generator.

2 Diesel Generator Sets

Diesel generator sets convert fuel energy (diesel or bio-diesel) into mechanical energy by means of an internal combustion engine, and then into electric energy by means of an electric machine working as generator.

2.1 The Main Parts and Characteristics of Diesel Generator Sets

The main parts of a diesel generator are: the internal combustion engine, usually air- or water-cooled; the electric generator usually of synchronous type; the mechanical coupling; the support chassis; the battery for generator start-up; the fuel tank; the starter motor; the command panel, etc.

The main characteristics of a diesel generator set are: rated power, rated voltage, rated frequency, number of phases, etc.

The diesel generator sets are usually designed to run at 3000 rpm or 1500 rpm at a frequency of 50 Hz. The primary movers are internal combustion engines equipped with mechanical regulators to keep the imposed speed, integrated in the injection pump and adjusted to obtain an output frequency of about 52 Hz without load and 50 Hz for rated load.

The proper operation of a diesel generator set is determined to a great extent by two main components, the *speed regulator* and the *voltage regulator*. The performance of these components are vital for the operation and utilization of diesel generator sets, their purpose being to precisely maintain the imposed parameters of electric power (voltage and frequency).

The *speed regulator* is designed to keep constant the internal combustion engine speed by changing the quantity of fuel consumed by the motor. The direct result of this speed regulation is a stable frequency of voltage at the generator terminals (since frequency is proportional to the generator speed). A constant frequency requires good precision and a short response time from the speed regulator. The speed regulator starts regulating when various electric loads are connected or disconnected at the generator terminals.

There are a lot of speed regulation systems, starting from the simpler spring-based ones up to complex hydraulic and electronic ones able to regulate dynamically the fuel admission valve to keep the speed constant in a given range with response times at load changes smaller than 1-3 seconds.

The main role of the *voltage regulator* is to control the voltage at the generator terminals and to keep it constant by limiting as fast as possible the voltage peaks and over voltages that occur due to load variations. The quantity that the voltage regulator acts on is the excitation current that changes the voltage amplitude at the generator terminals.

2.2 Utilizations of Diesel Generator Sets

Diesel generator sets are equipment that work either in *isolated regime,* in areas where there is no possibility to connect to the grid (e.g. the electric energy supply of households, chalets, holiday houses, isolated objectives, equipment in industrial sites, electric installations for outdoor entertainment events, military equipment, telecommunications, etc.), or as *emergency* regime, as a reserve electric power source, in the event of electric power blackouts. In *emergency regime* the diesel generator sets usually supply only vital consumers, like fire pumps, elevators, safety lighting installations, banks, hospitals, government buildings, offices, supermarkets and large restaurants, hotels, malls, stadiums, airports, fuel stations, private houses, and industrial sites where specific processes do not allow for blackouts, become incontrollable or generate important losses without electric power, etc.

Usually, in parallel with diesel generator sets, UPS systems are used, with a buffer role, able to ensure for short periods the continuity of power supply for vital consumers, until the diesel generator sets are started-up. The minimum combined time necessary for the detection of a grid voltage drop, the start-up of internal combustion engine, reaching the stabilized regime of the generator (frequency and voltage) and the load connection is typically at least 10-15 seconds.

In the case of power systems based on renewable energies, given the fluctuant character of unconventional energy sources (e.g. wind or solar energy), diesel generator sets takes on particular importance, their role being to ensure the continuity of electric power for the local grid during periods when the renewable sources of energy become unavailable or insufficient.

3 The Numerical Modeling of a Diesel Generator Set

The analysis of the complex aspects of a diesel generator set requires the development of reliable numerical models that allow for simulation in different operation regimes, specifically in conditions as close as possible to the reality of the assembly internal combustion engine – the synchronous generator.

The synchronous generator represents a key component of a diesel generator set. It converts the mechanical power produced by the primary mover into electrical power. The numerical models that can be used in the study of the synchronous generator can be classified into circuit models and field models, with the ones most used in electric drive systems being the circuit models.

From the circuit models category of the synchronous machine, the most used is the orthogonal model (d-q) that allows for the study of different operation regimes of the machine, specifically the study of the assembly machine – the regulation system.

The orthogonal model (d-q) of the synchronous machine is based on the decomposition of the three-phase machine in an equivalent bi-phase one, Fig. 1. In this way the phasors system specific to the three-phase machine is decomposed along the longitudinal axis *d* and transversal axis *q* respectively.

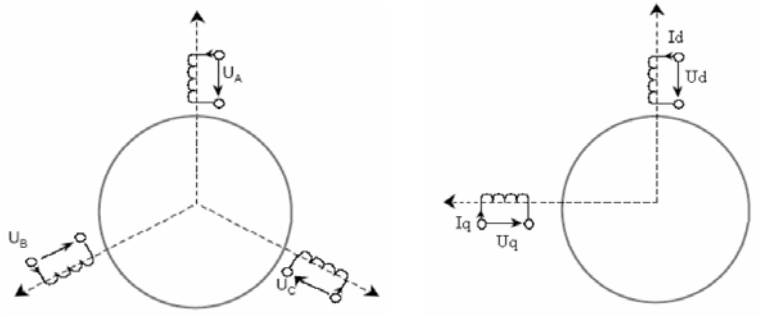


Figure 1
Transformation of a three-phase machine in an equivalent bi-phase one

The voltage equations of the synchronous generator written in a coordinate system $d_\theta - q_\theta$ situated at the electric angle θ with respect to the fixed coordinate system $d - q$ is as follows:

$$\underline{u}_{s\theta} = R_s \underline{i}_{s\theta} + \frac{d\underline{\Psi}_{s\theta}}{dt} + \underline{j}\omega\underline{\Psi}_{s\theta}, \tag{1}$$

$$\underline{u}_{r\theta} = R_r \underline{i}_{r\theta} + \frac{d\underline{\Psi}_{r\theta}}{dt}, \tag{2}$$

where $\underline{u}_{s\theta}$ and $\underline{u}_{r\theta}$ are the stator and rotor voltage phasors, R_s and R_r are the stator and rotor resistances, $\underline{i}_{s\theta}$ and $\underline{i}_{r\theta}$ are the stator and rotor current phasors, $\underline{\Psi}_{s\theta}$ and $\underline{\Psi}_{r\theta}$ are the stator and rotor magnetic flux phasors and ω is the pulsation [9].

In order to study the operation of a diesel generator set either in *isolated regime* or in *emergency regime,* dedicated numerical models were developed using the Matlab/Simulink software package.

3.1 Generator Set Operating in Isolated Regime

The numerical modeling scheme used for the simulation of a diesel generator set in *isolated regime* is presented in Fig. 2, where we can distinguish: the *generator set*, the electric *load*, the normally open *switch* I1, and the *measurement* blocks.

The generator set block is composed of the diesel internal combustion engine model and the synchronous generator model, and of the voltage and speed regulation systems, Fig. 3. The main data of the simulated generator set are as follows: rated power 4.5 kVA, rated voltage 400 V, rated frequency 50 Hz.

The development of numerical models for studying the operation of the diesel generator set in continuous regime permitted us to estimate the answer of the generator in case of load variations.

In order to exemplify the effectiveness of the numerical models, we simulated the transient regime of the equipment in the case of a sudden connection of a resistive load. The initial state of the generator set is of stand-by type with no-load.

These simulation results emphasize also the capacity of voltage and speed regulation systems to keep the imposed values of generator speed and voltage.

The mechanical power (per unit) initially developed by the combustion engine is very small because the synchronous generator coupled with the engine is running without load, Fig. 4a). After suddenly connecting the resistive load, the mechanical power developed by the engine increases rapidly and after approximately 1.5 seconds, a new operation stabilized regime is reached.

During the transient regime, due to the sudden coupling of the load, the machine speed (per unit) decreases abruptly, but for a very short period of time (~0.25 seconds), after which it stabilizes at the imposed value as a result of the action of the speed regulation system, Fig. 4b).

The voltage (per unit) at the generator terminals also demonstrates a significant decrease for a short time, after which it stabilizes in less than 0.5 sec. due to the action of the voltage regulation system, Fig. 4c).

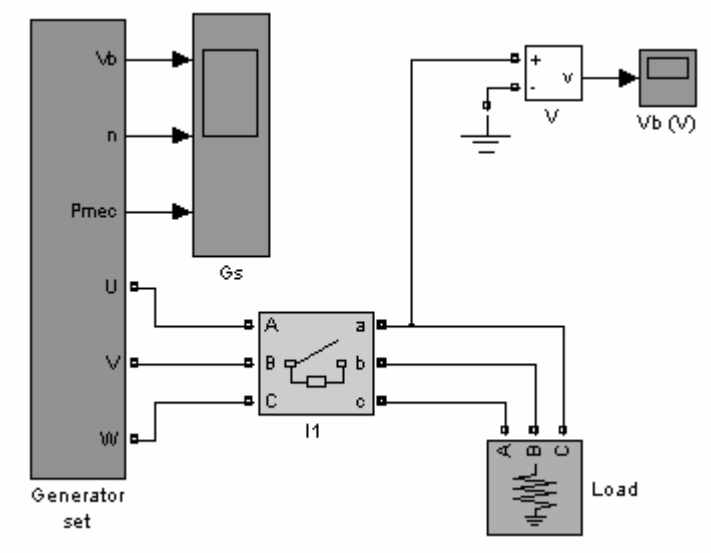
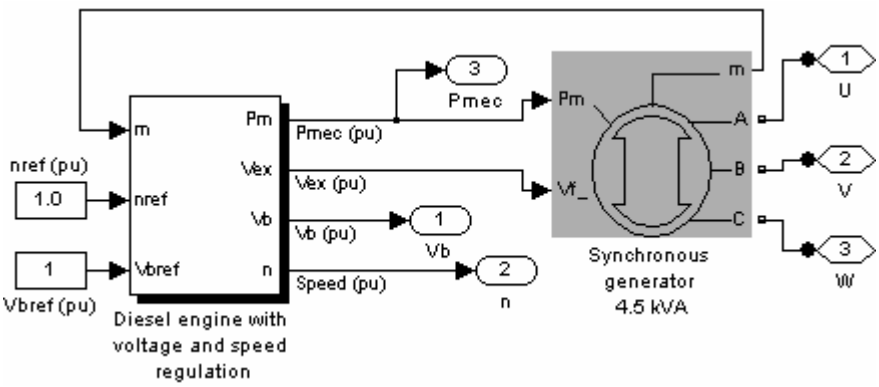
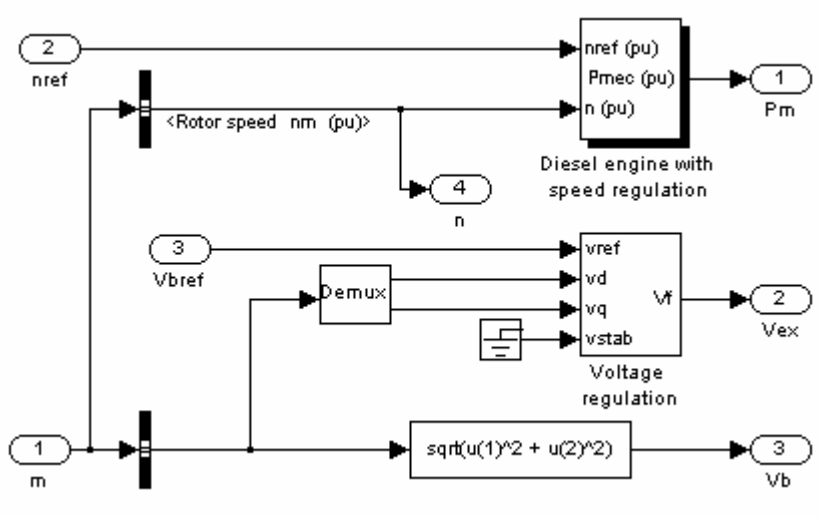


Figure 2
Simulink model of a generator set operating on a local grid

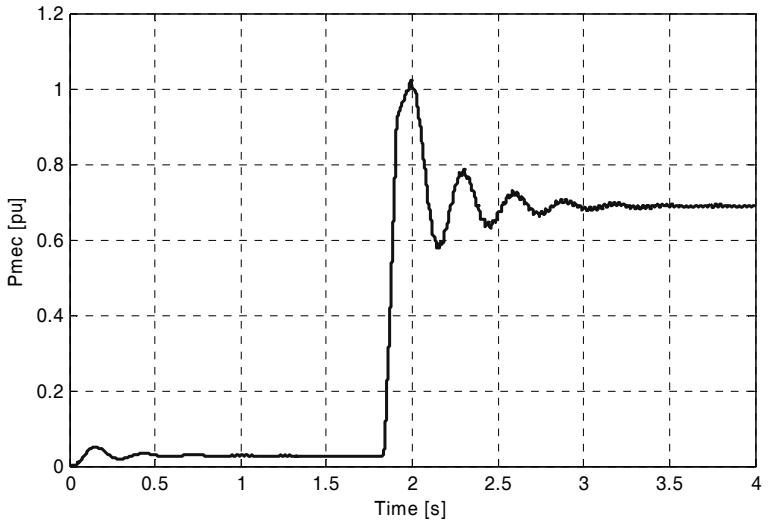


a)

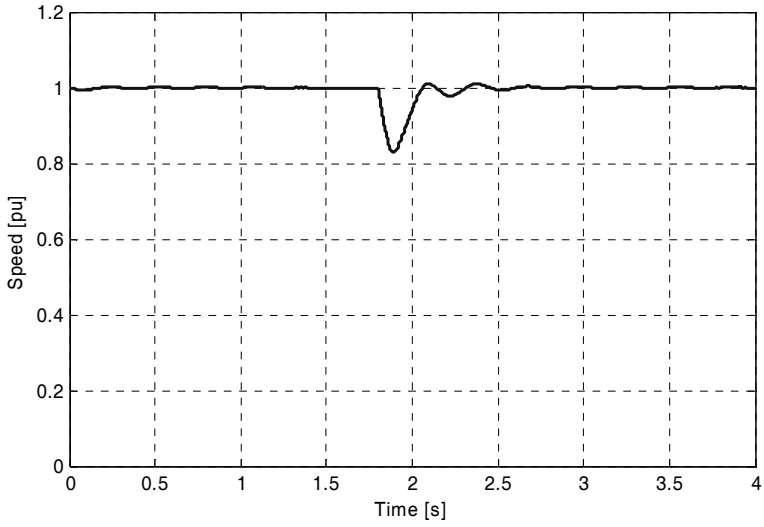


b)

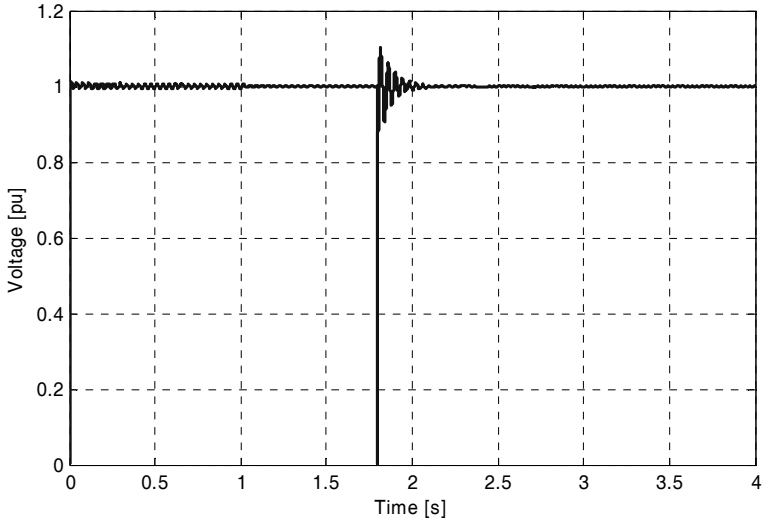
Figure 3
Simulink model of a generator set with voltage and speed regulation systems;
a) global scheme of the generator set; b) Diesel engine and regulation systems



a)



b)



c)

Figure 4
Time variation of generator set quantities during the transient regime after applying a resistive load; a) mechanical power; b) rotor speed; c) output voltage

3.2 Generator Set Operating in Emergency Regime

In the case of important equipment or objectives, electric power consumers are usually grouped into vital consumers and non-vital ones. An electric power security solution for the vital consumers is to back-up their supply by means of a diesel generator set. The installation of the diesel generator set should be done so that during a blackout resulting from a grid fault, it is possible to keep connected only the vital consumers. In the case of wind and/or solar renewable power systems, a black out could occur in the event of insufficient solar or wind power.

Such a situation is presented in Fig. 5, where we can identify the following blocks: the Grid (three-phase, 400 V, 50 Hz, which can be represented by a renewable power system), the Generator set (4.5 kVA), the Three-phase short-circuit block, vital consumers (2 kW resistive), and non-vital consumers (10 kW resistive).

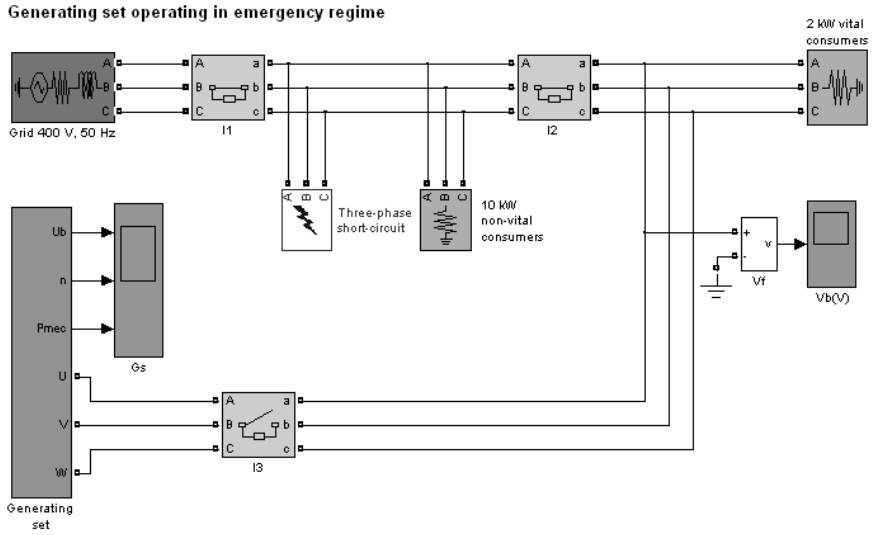


Figure 5
Generator set operating as emergency power source

The two groups of resistive type consumers have the rated powers 10 kW (non-vital consumers) and 2 kW (vital consumers), and they are supplied from the grid. When a three-phase short-circuit occurs at the grid level, the normally closed breakers I1 and I2 cut the energy supply for the non-vital consumers (10 kW), and the normally open breaker I3 connects the vital consumers (2 kW) to the diesel generator set.

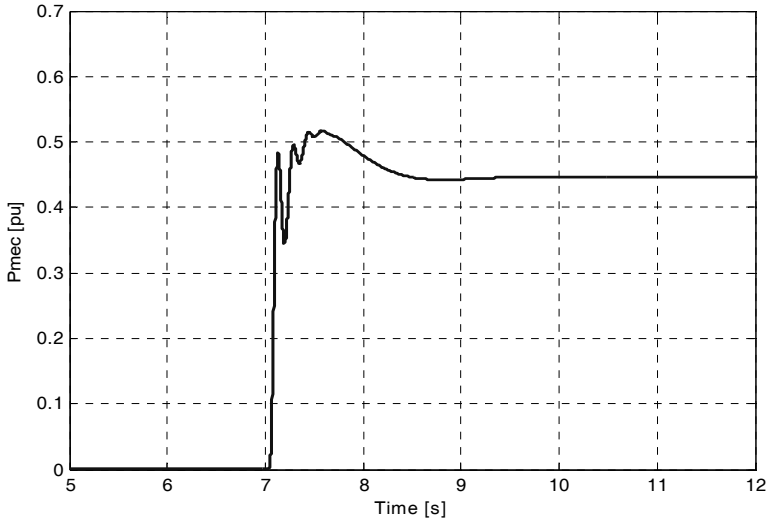
The three-phase short-circuit occurs in our case after 4 seconds from the simulation start. At time instant 4.01 seconds, the I1 and I2 breakers open, and at time instant 7.01 seconds, the I3 breaker closes in order to supply the vital

consumers. Thus the vital consumers remain unsupplied for 3 seconds, the period necessary for the diesel generator set to start-up. If the vital consumers cannot tolerate short periods of time without electricity, then the diesel generator should work in stand-by regime, being ready at any time to be connected to the load. Moreover, if the vital consumers are sensitive to voltage and frequency variations, which can occur during the commutation from the grid to the diesel generator set, a UPS system should be provided as a buffer until the extinction of the transient regime.

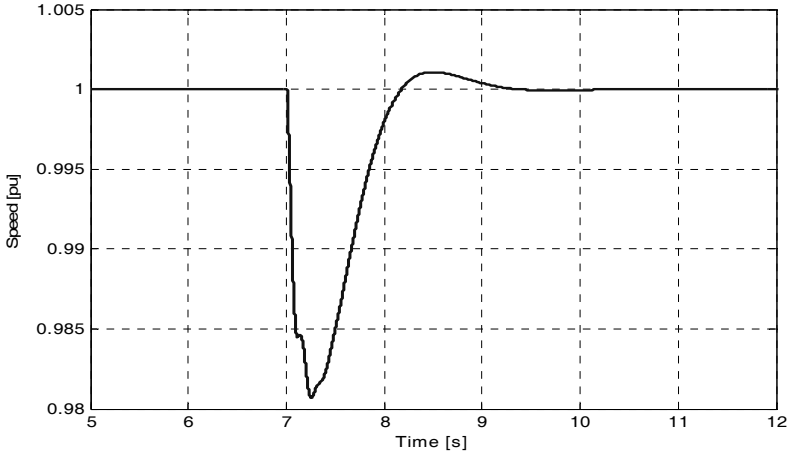
The system answer in the case of emergency is presented in Fig. 6, where we can identify the time evolutions of mechanical power, rotor speed and output voltage (per unit).

When the short-circuit occurs (after 4 seconds from the simulation start), the mechanical power produced by the generator set increases from a small value (no-load regime) and stabilizes at the value imposed by the regulation systems, Fig. 6a).

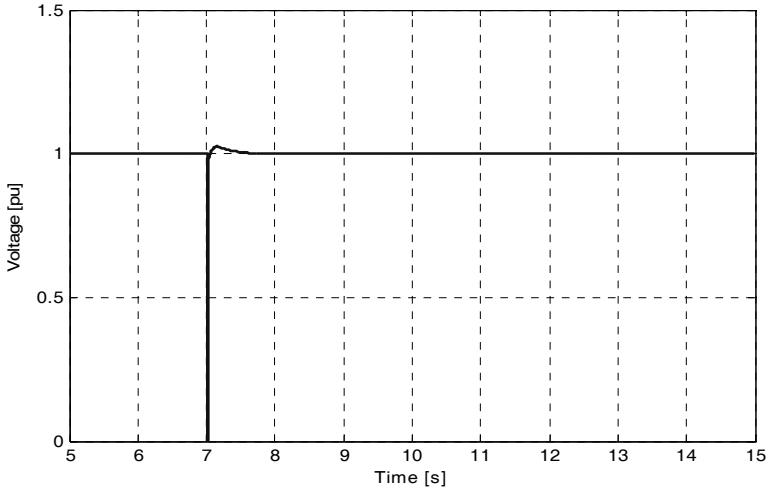
The output voltage (Fig. 6c) demonstrates an significant decrease, but for a short period of time, after which it comes back very quickly at the rated value. Such a voltage drop can be prevented by means of a properly sized UPS system.



a)



b)



c)

Figure 6
Time variation of generator set quantities when operating in emergency regime;
a) mechanical power; b) rotor speed; c) output voltage

4 Experimental Results

The goal of the experimental analysis of the paper was to register the transient phenomena after coupling a load to the diesel generator set initially operating in stand-by regime.

The diesel generator set subject to the experimental tests is a Kipor type, rated power 4.5 kVA, rated voltage 230 V, rated frequency 50 Hz, equipped with automatic start-up.

As presented in Fig. 7, the experimental set-up consists of: a diesel generator set, a laptop with a data acquisition card, conditioning circuits and transducers, load resistance, an ampere-meter, a voltmeter and an on/off breaker.

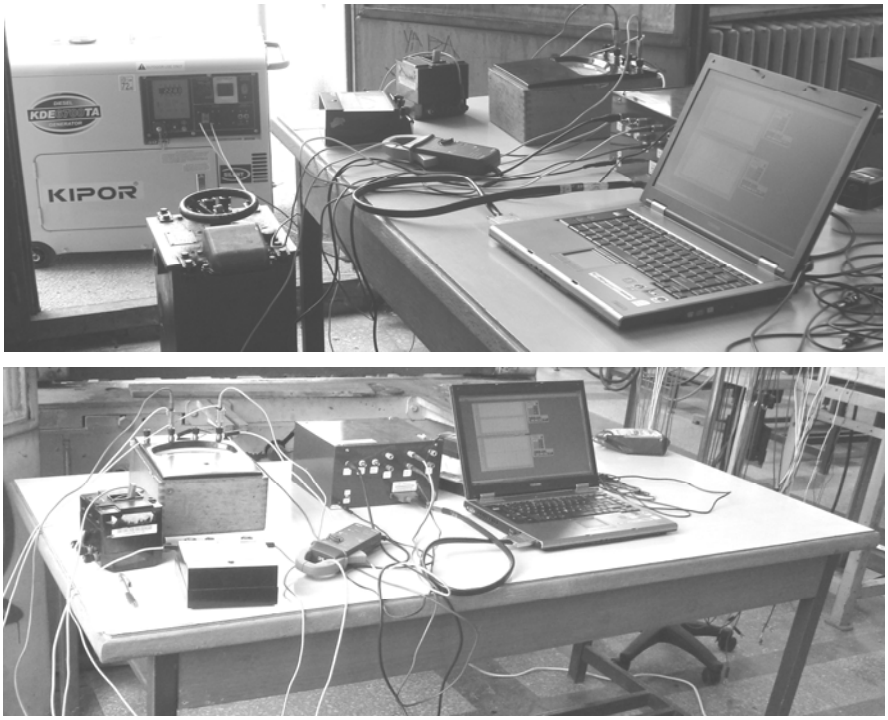


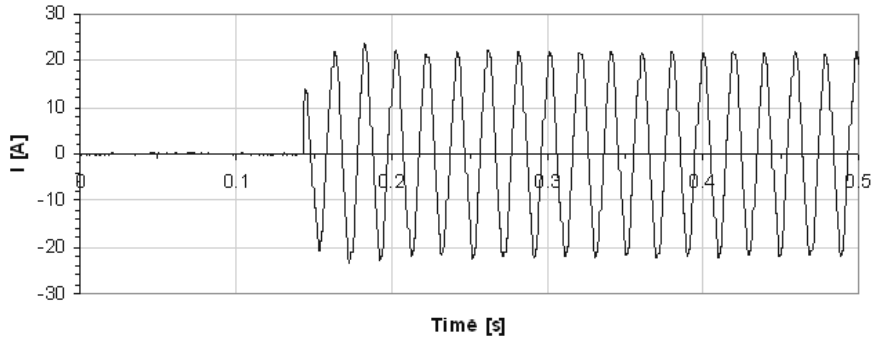
Figure 7
Experimental set-up used for investigations

Using a data acquisition card, we stored the time evolution of the output current and voltage produced by the generator set after applying a resistive load at the generator terminals.

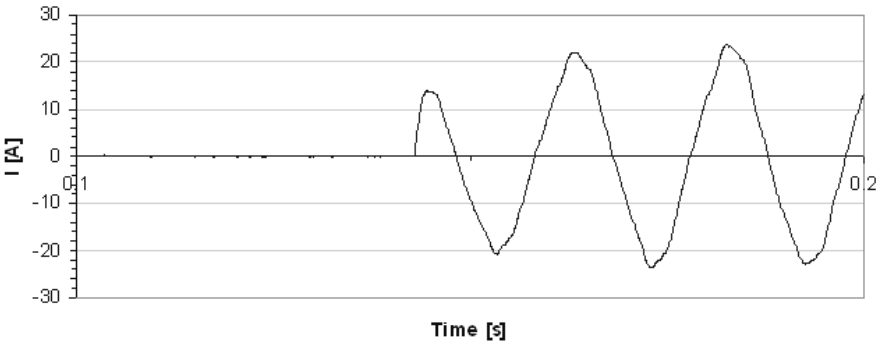
At the beginning, the diesel generator set was operating in no-load regime, the load current being null, the generated voltage 230 V and the frequency 51.6 Hz.

When operating on a local grid, the diesel generator set used for experiments is designed in such a way that starting up the machine under load condition is not recommended. Thus, after the start-up of the diesel generator set, we closed the load circuit by acting on the on/off breaker, the rms value of the absorbed current reaching 15.2 A. The power delivered to the resistive load is 3.35 kW. We can notice in Fig. 8 that the approximate duration of the transient regime is 0.13 s.

By studying the experimental results illustrated in Fig. 9, we can observe that the frequency of the voltage at the generator terminals presents a decreasing trend from 51.6 Hz in no-load operation to 51.02 Hz under load operation of 66% from the rated load of the generator (a specific trend for this type of equipment).

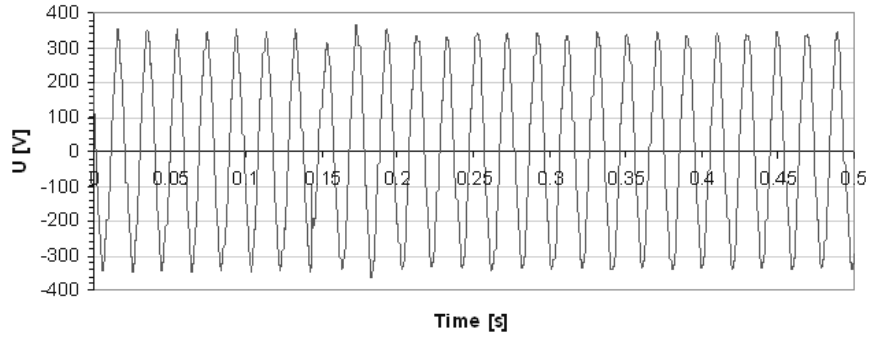


a)

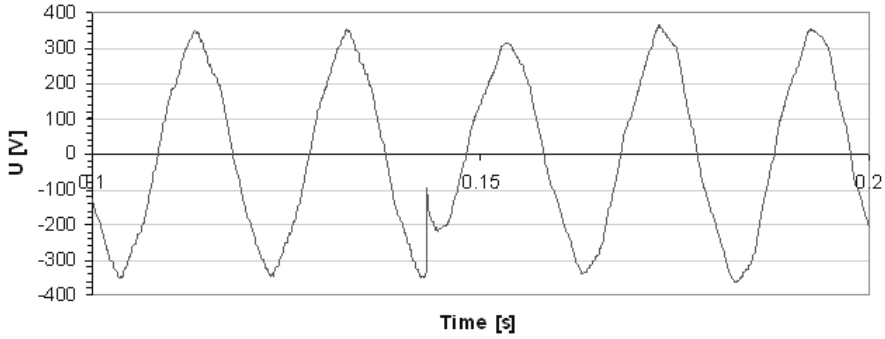


b)

Figure 8
Time variation of load current after applying a resistive load; a) global view; b) detail



a)



b)

Figure 9
Time variation of voltage at generator terminals after applying a resistive load;
a) global view; b) detail

We can observe in Fig. 9 a voltage drop at the generator terminals of about 70%, but it happens for a short period of time, after which the voltage regulator brings the voltage back to the imposed value.

Conclusions

This paper deals with aspects specific to the operation of diesel generator sets working both in isolated regime and as a back-up electric energy source for vital consumers who are usually supplied from the grid or from hybrid power systems based on renewable sources.

The study of diesel generator sets was treated both via numerical modeling and experiments in laboratory.

The numerical models used for the study of the dynamic regimes of diesel generator sets were developed using a Matlab/Simulink software package. Via

numerical simulations, we determined the time variations of the main quantities useful in the study and operation of diesel generator sets (mechanical power, voltage, speed). We should observe also the usefulness, efficiency and flexibility of developed numerical models in the study of the behavior of power systems for different topologies, charge levels, load types, etc.

The experimental measurements aimed to determine the dynamic response of the diesel generator set after a sudden load connection. The diesel generator set that was experimentally tested was a Kipor 6700 one, with a rated power of 4.5 kVA.

Acknowledgement

This work was supported by the Romanian Ministry of Education and Research, in the frame of the research grant PNII, 21-075/2007.

References

[1] R. J. Best, D. J. Morrow, D. J. McGowan, P. A. Crossley: Synchronous Islanded Operation of a Diesel Generator, IEEE Transactions on Power Systems, Vol. 22, pp. 2170-2176, 2007

[2] K. Karasavvas: Modular Simulation of a Hybrid Power System with Diesel, Photovoltaic Inverter and Wind Turbine Generation, Journal of Engineering Science and Technology Review, Vol. 1, pp. 38-40, 2008

[3] T. Tudorache, A. Morega: Optimum Design of Wind/PV/Diesel/Batteries Hybrid Systems, MPS 2008

[4] W. Shi, J. Yang, T. Tang: RBF NN-based Marine Diesel Engine Generator Modeling, Proc. of American Control Conference, 2005

[5] H. Wei, R. Min, T. Yingqi: A Fuzzy Control System of Diesel Generator Speed, APPEEC 2009 Conference

[6] W. Shi: Multi-Neural Networks Control of Marine Diesel Engine Generator Set, Proc. of 4th International Conference on Fuzzy Systems and Knowledge Discovery 2007

[7] M. L. Tomescu, S. Preitl, R.-E. Precup, J. K. Tar: Stability Analysis Method for Fuzzy Control Systems Dedicated Controlling Nonlinear Processes, Acta Polytechnica Hungarica, Vol. 4, No. 3, pp. 127-141, 2007

[8] J. Žilková, J. Timko, P. Girovský: Nonlinear System Control Using Neural Networks, Acta Polytechnica Hungarica, Vol. 3, No. 4, pp. 85-94, 2006

[9] C. Ghiță: Electrical Machines (in Romanian), Matrix-Rom Publishing House, Bucharest, 2005

Proposed Methods to Increase the Output Efficiency of a Photovoltaic (PV) System

Fatima Zohra Zerhouni, M'hamed Houari Zerhouni, Mansour Zegrar, M. Tarik Benmessaoud, Amine Boudghene Stambouli, Abdelhamid Midoun

Electronics Department
University of Sciences and Technology Mohamed Boudiaf (U.S.T.M.B.O)
BP 1505, Oran El M'Naouer, Oran, Algérie
E-mail: zerhouni_fz@yahoo.fr, aboudghenes@yahoo.com

Abstract: Lately, the use of solar energy has seen considerable development. Transformation in electric energy is one of its applications which attracts considerable interest, owing to the fact that it makes it possible to solve a major problem in isolated cities that lack electrical supply networks. For solar energy use, the current drawback remains its high cost. This problem can be resolved through different improvements in terms of power production. For that, different axis of research can be explored [1-3].

Over the past few years, solar cells arrays (SCA) have been connected to various loads in a direct coupled method. A PV module can produce the power at a point, called an operating point, anywhere on the current- voltage (I-V) curve. The coordinates of the Operating point are the operating voltage and current. There is a unique point near the knee of the I-V curve, called a maximum power point (MPP), at which the module operates with the maximum efficiency and produces the maximum output power. The point of maximum power is the desired operating point for a PV array to obtain maximum efficiency. A PV array is usually oversized to compensate for a low power yield during winter months. This mismatching between a PV module and a load requires further over-sizing of the PV array and thus increases the overall system cost [3-5]. To mitigate this problem, different methods have been developed [6-12]. A maximum power point tracker (MPPT) can be used to maintain the PV module's operating point at the MPP. The system in a direct coupled method cannot always operate at maximum power point (MPP) of the solar array when the load, irradiance or temperature changes. A PV-load coupling system should be able to maximize the energy output of the PV generator, which should operate always at its maximum power point (MPP) in order to achieve maximum global efficiency. Some authors have supported the direct coupling between the PV generator and the load [2-3]. Several methods and algorithms to track the maximum power point have been developed [2-12].

We focused our effort on improving the matching between PV sources and loads through the development of new performing methods. In this paper, a rapid approach for peak power tracking is proposed. The system PV generator - load is optimized, when the working point in direct coupling is quite near the MPP of the PV generator, so that the global efficiency of the system is acceptable. Another method is proposed. It is based on reconfiguring on-line the SCA by changing the connections between different modules in

order to minimize the losses due to load and operation conditions. Two strategies for improvements are chosen for testing in this study, whatever the loads and operating conditions.
The experimental system used a microcontroller. It is robust when undergoing environment changes and load variations. The performance of the methods is verified through simulations and experiments.

Keywords: photovoltaic; MPPT; configuration; optimum power; switching

1 Introduction

The conversion of solar energy into electric energy is performed by means of photovoltaic (PV) generators. Photovoltaic offer the highest versatility among renewable energy technologies. Electricity produced from photovoltaic (PV) systems has a far smaller impact on the environment than traditional methods of electrical generation. The most attractive features of solar panels are the nonexistence of movable parts, the very slow degradation of the sealed solar cells and the extreme simplicity of its use and maintenance. Another advantage is the modularity. All desired generator sizes can be realized, from the mill watt range to the megawatt range. Solar energy is a pollution-free source of abundant power. During their operation, PV cells need no fuel, give off no atmospheric or water pollutants and require no cooling water. The use of PV systems is not constrained by material or land shortages and the sun is a virtually endless energy source.

2 Modeling of A PV Generator

The basic component of a SCA (solar cells array) is the photovoltaic cell. A photovoltaic (PV) array under uniform irradiance Es exhibits a current-voltage characteristic with a unique maximum power point (MPP) where the array produces maximum output power, which changes as a consequence of the variation of the irradiance level and of the SCA temperature T[3-12].

The simple equivalent circuit is sufficient for most applications. Likewise, the PV generator current and, consequently, the power vary with the cells' operation temperature T and irradiance Es. A photovoltaic array can be represented by an equivalent circuit composed of a current generator, a sensitive diode D to the light, a series resistor Rs and a shunt resistance Rsh, as shown in Figure 1.

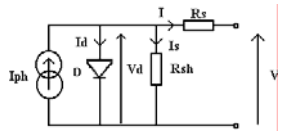


Figure 1
Equivalent circuit of a PV cell

By inspection of Figure 1, Kirchhoff's first law provides the current–voltage(I-V) characteristics. One can write the equation [3]:

$$I = I_{ph} - I_d - I_s \tag{1}$$

$$I_d = I_o.\left[\exp(\frac{V_d}{A.U_T}) - 1\right] \tag{2}$$

Where

I_{ph} is the photocurrent source equal to the short-circuit current I_{cc}

I_d is the generated current,

I_o is the saturation current of a solar array,

q is the electron charge,

K_B is the Boltzmann's constant,

and A is the ideality factor for a p–n junction.

$$I = I_{ph} - I_o.\left[\exp\frac{V_d}{A.U_T} - 1\right] - \frac{V_d}{R_{sh}} \tag{3}$$

Or $$V_d = V + R_s.I \tag{4}$$

then,

$$I = I_{ph} - I_o\left[\exp(\frac{V + R_s.I}{A.U_T}) - 1\right] - \frac{V + R_s.I}{R_{sh}} \tag{5}$$

By considering Rsh very large, one can approximate (5) by:

$$I = I_{cc} - I_o\left[\exp(\frac{V + R_s.I}{A.U_T}) - 1\right] \tag{6}$$

$$V = -R_s.I + A.U_T.Ln\left[\frac{I_{cc} - I + I_o}{I_o}\right] \tag{7}$$

The equation (6) of the cell current *I* depends on the cell voltage V with the saturation current *I*o and the diode factor A.

The expression (6) can represent the operation of a solar cell appropriately, but it presents some inconveniences in practical use. The expression possesses an implicit character: the current appears on both sides of the equation, forcing its solution through iterative methods or approximate solutions. The Newton Raphston iterative method is used.

Photovoltaic systems are designed around the photovoltaic cell. Since a typical photovoltaic cell produces less than 3 watts at approximately 0.5 volt dc, cells must be connected in series or/and in parallel configurations to produce enough power for high-power applications. In our case, a photovoltaic module is constituted of 36 solar polycrystalline cells in series (Kyocera LA 361 K51) mounted 35° south. Figure 2 illustrates the operating characteristic curves of the solar array under given irradiance and temperature. In Figure 2, V and I are the output voltage and the output current of the SCA, respectively. P is the power. A photovoltaic array under uniform irradiance Es and fixed temperature T exhibits a current-voltage characteristic with a unique point, called the maximum power point (MPP), where the array produces maximum output power Popt (Figure 2).

a) I-V characteristic b) P-V characteristic

Figure 2

Electrical characteristic of the SCA at Es=100%, T=25°C

$$P = V.I = -R_s.I^2 + A.U_T.I.Ln\left[\frac{I_{cc} - I + I_o}{I_o}\right] \tag{8}$$

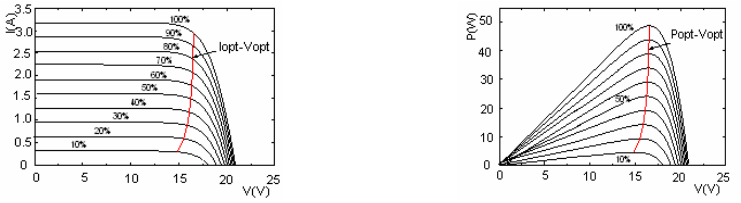
The curve in Figure 2a consists of different regions: one is the current source region, and the other is the voltage source region. In the voltage source region, the internal impedance of the solar array is low on the right side of the power curve, and in the current source region, the internal impedance of the solar array is high on the left side of the power curve. The maximum power point of the solar array is located at the knee of the power curve (zone 3). According to the maximum power transfer theory, the power delivered to the load is maximum when the source internal impedance matches the load impedance in a direct coupling. The peak Popt in Figure 2 is provided by solving the following equation [3]:

$$\frac{dP}{dI} = -2.R_s.I + A.U_T.Ln\left[\frac{I_{cc} - I + I_o}{I_o}\right] - \frac{A.U_T.I}{I_{cc} - I + I_o} = 0 \tag{9}$$

The current Iopt is the solution of the following equation:

$$I_{cc} = I_{opt} + I_o.\left[\exp[\frac{(2.Iopt.Rs)}{A.U_T} + \frac{I_{opt}}{I_{cc} - I_{opt} + I_o}] - 1\right] \tag{10}$$

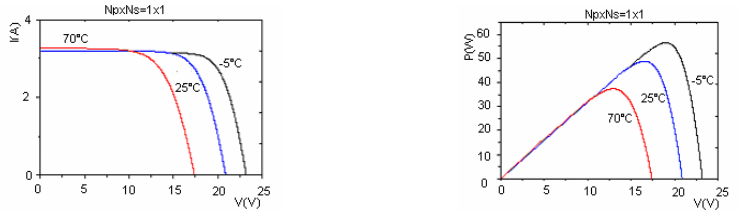
Iopt is replaced in equation (7) to obtain Vopt. The product Vopt. Iopt result in the optimal power Popt. There is a unique Popt on each P–V characteristic curve. Vopt and Iopt varie according to the sun irradiation ant temperature. Figure 3 shows the characteristic curves at different irradiances Es. Figure 4 shows the characteristic curves at different temperatures T.



a) I-V characteristics b) P-V characteristics

Figure 3

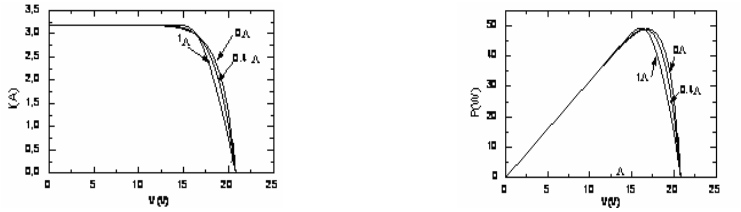
I-V and P-V characteristics of a solar module under varied solar irradiance



a) I-V characteristics b) P-V characteristics

Figure 4

I-V and P-V characteristics of a solar module under varied temperatures



a) I-V characteristics b) P-V characteristics

Figure 5

Influence of the series resistance Rs on the I-V and P-V characteristics of a PV module

a) I-V characteristics b) P-V characteristics

Figure 6
Influence of the series resistance Rsh on the characteristics of a PV module

The series resistance and the shunt resistance of the panel have a large impact on the slope of the I-V curve, as seen in Figures 5 and 6 respectively.

3 The Maximum Power Point Tracker

3.1 Introduction

Maximum Power Point Tracker (MPPT) is often used to increase the energy conversion efficiency for a photovoltaic energy source. The maximum power is transferred to the load when the impedance source matches the load one. To accomplish this objective, a switching converter is placed between the PV source and the load. With an MPPT control, it is possible to reach the output panel characteristics around the optimal power point. The output power of photovoltaic (PV) panels varies with atmospheric conditions (solar irradiance level and temperature) as well as their optimum voltage (Vopt) and current (Iopt). It is crucial to operate the PV energy conversion systems near the maximum power point (MPP) to increase the power yield of the PV system. This problem has attracted the interest of several authors [3-12]. Maximum power point tracking (MPPT) algorithms are usually implemented in the power electronic interface between the PV panel and a load. Maximum power point matching of a PV array with the load maximises the energy transfer by operating the load as closely as possible to the MPP line of the PV output, whatever the loads and working conditions. In our case, a Maximum power point tracker is a DC to DC convertor.

3.2 Optimum Power Point Detector

The current-voltage curve in a PV generator changes with the irradiance and the temperature conditions. This situation implies the need for a specific dynamic searching algorithm for the maximum and safe power point of the coupling PV-load. So, a system is inserted between the SCA and the load as shown in Figure 7.

This system is called MPPT (Maximum Power Point Tracker). In order to make the most of the available solar energy for any load and working conditions, the MPP can be tracked by using the MPPT [3-12]. In Figure 7, the input voltage and current are respectively (V_1, I_1). The output voltage and current are respectively (V_2, I_2).

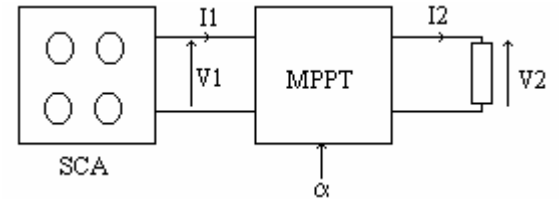


Figure 7
Connexion SCA – load via MPPT

α is the static converter duty cycle [3]. It is a control parameter. α varies. α allows an optimal impedance adaptation:

$$p_2 = p_1 \tag{11}$$

P_2 is the output power;

P_1 is the input power;

α_{opt} is determined by:

$$V_1 = Vopt \tag{12}$$

$$I_1 = Iopt \tag{13}$$

Vopt and Iopt are the optimal voltage and the optimal current respectively.

In general, α can be determined according to the load [3]. So the load parameters are required, but unfortunately these parameters are not constant. The study is then for a specific load and cannot be applicable if the load changes. Also, the optimal voltage Vopt and the optimal current Iopt depend on the climatic conditions, on the particular moment of the day and season. These variations are problems. Therefore, it is important to design a flexible system which can be adapted to large types of loads and under different operating conditions in order to reduce the mismatch losses in a directly coupled load to a SCA.

Another method is the fixed reference voltage control method. It is the simplest control method. This method regulates the solar array terminal voltage and matches it to a fixed reference voltage, which is the MPP of the SCA. Indeed, the V_{opt} variation according to V_{oc} (V_{oc}: the SCA open circuit voltage) is linear whatever the working conditions [3]. The approximation is:

$$V_{opt} = \frac{75.V_{oc}}{100} \tag{14}$$

α_{opt} is calculated according to:

$$\alpha_{opt} = \frac{V_2}{V_{opt}} \tag{15}$$

To measure V_{oc}, a pilot cell in open circuit is used. It is at the same SCA's working conditions.

$$V_{oc} = N_s.V_{ocs} \tag{16}$$

V_{oc}: SCA open circuit voltage;

V_{ocs}: the cell pilot open circuit voltage;

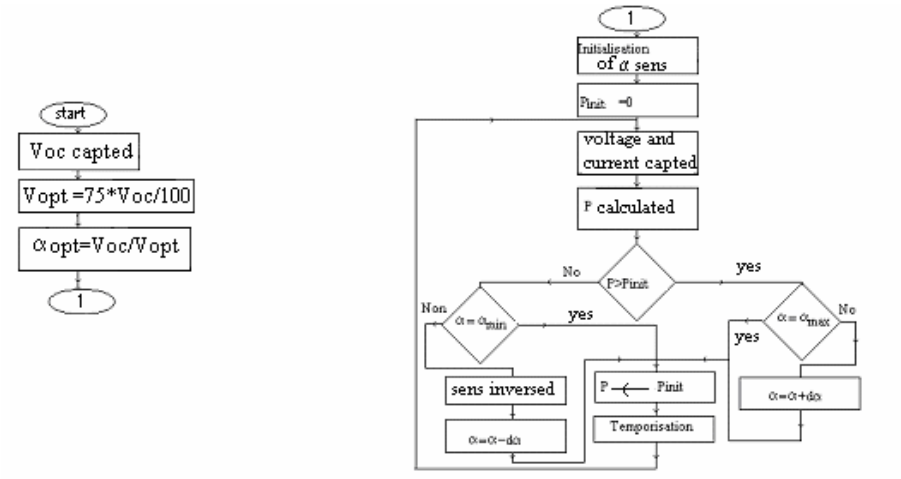
N_s: the series solar cell's number.

In practice, the Voc value will be different from the real because the photovoltaic cells are not all identical. To overcome this problem, one measures real SCA V_{oc} while cutting off the load's supply for a short moment, without disturbing the system's operation. There is also another control method, independent of any knowledge of the SCA or the load characteristics, to find the optimum power point by comparing powers [3].

3.3 Adopted Method

We combined the methods practically. In the first time, in real time, the system is in open circuit. Model Vopt is calculated according to the equation (14). The voltage operation is measured. Then, the cyclic ratio is calculated, according to the relation (15). The principle is to adjust the actual operating point of the PV array so that the actual power P approaches the optimum value Popt as closely as possible. This cyclic ratio brings the operation point around the optimum power point [3]. We then track the optimum power point tracking per powers comparison as shown in flowchart (Figure 8). According to the power's variation, the cyclic ratio is readjusted. At the end, the system will oscillate between three points which encircle the optimum power point.

Thus, in this zone, the optimum power point is then determined. The MPPT can always track the MPP. At this level, if the load and/or the irradiance change abruptly, the operating point is deviated from that optimal. The cyclic ratio is changed automatically, in a gradual way, until it reachs the MPP as shown in the flowchart in Figure 8b. However, to reduce the convergence time towards the optimum power point, a method is proposed. A relation between the new cyclic ratio to that preceding the change is established.



a) Jump to the MPP b) power's comparison

Figure 8
MPP research flow chart

In Figure 9, one supposes that after having determined the MPP, the operating point is A'. A sudden increase in the load will bring back the operating point to the point B'. The cyclic ratio to adopt in this case is calculated according to:

$$\alpha'_{opt} = \frac{\alpha_{opt}.V'.I_{opt}}{V_{opt}.I'} \tag{17}$$

with:

α_{opt}: Cyclic ratio around the MPP before the abrupt change;

I_{opt}: SCA current around the MPP before the change occurred;

V_{opt}: SCA voltage around the MPP before the abrupt change;

V': SCA voltage after the abrupt change;

I': SCA current after the abrupt change;

α'_{opt}: New cyclic ratio.

The point B' is brought back around the MPP at the point C (equation 17). Then, the MPP tracking by comparing powers is adopted (Figure 8b)).

Another case is studied. The operating point is A'. For an abrupt reduction in load, the new operating point is point A (Figure 9). The current in A and the current in A' are closed and closed to the SCA short circuit current. The cyclic ratio adopted to bring back the operating point A around the MPP (noted B) will be a particular case of the equation (14), according to:

$$\alpha'_{opt} = \frac{\alpha_{opt}.V'}{V_{opt}} \tag{18}$$

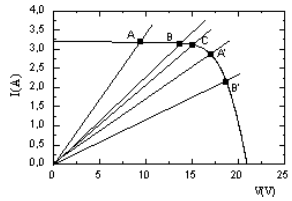


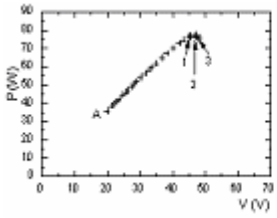
Figure 9
MPP research

3.4 Practical Results

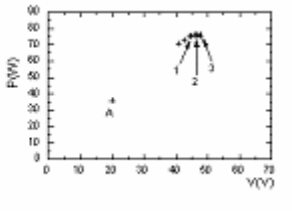
The practical results consider a SCA of three modules in series. A DC buck converter has been realized experimentally and used. The experimental results are given below.

Figure 10 illustrates the practical results of the power P against the voltage V, for an irradiance Es=56% and a temperature T=27 °C. The MPP research characteristic in a gradual way is illustrated by Figure 10a. The operating point is initially at the point A (V_A=20.06 V, I_A=1.76 A, P_A=35.41 W). According to the algorithm established (Figure 8b) [3], the program compares the powers and readjusts gradually α to progress towards the MPP (P_{opt}=76.92 W, V_{opt}=46.68 V). The system operation will be stabilized at the end between the points 1, 2, and 3 which encircle the MPP.

Figure 10b illustrates the practical result for the same working conditions but by adopting the jump method towards the MPP according to the equations (17) and (18). Next, the method to move toward the MPP is by steps as shown in Figure 10b. It is seen that intermediate steps are avoided and convergence towards the MPP is fast. At the end, the system is operating between points 1, 2 and 3.



a) MPP gradual tracking



b) Jump towards the MPP

Figure 10
Experimental P-V characteristics, Es=56%, T=27 °C

Another case is also verified practically. During the SCA operation, a sudden load change occurs. It generates the operating point's displacement towards a power lower than the optimum one. The operation point must thus be brought back towards the optimum power zone. The developed program calculates in real time the cyclic ratio to approach the MPP quickly. The result is shown in Figure 11. This Figure shows an example of the SCA current evolution.

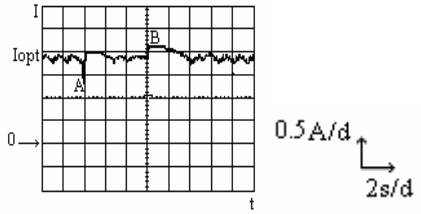


Figure 11
Experimental system response after abrupt load changes

The operating current is around I_{opt}, and an abrupt increase in the load (at "A") occurs. The current falls. The system calculates and adopts in real time α. So, an operation at I_{opt} occurred. Suddenly, at 'B', a load reduction occurs. The operating current increases. The system determines in real time α. The system's operation is stabilized gradually around the MPP at the I_{opt} current and an optimum operation is ensured. Our new MPPT control adjusts continually the static converter duty cycle. Transitory effects are immediately detected and the new MPP rapidly recovered. This method has been applied for a resistive load and for a DC motor. The MPP research has been successfully verified. The MPP is then rapidly recovered.

4 The Reconfiguration Method

4.1 Introduction

Another method presented in this paper is based on reconfiguring On-line the SCA by changing the connections between different modules in order to minimize losses. The number of modules required to build an SCA depends on the power required and also on the types of modules used and the operating conditions (solar radiation, temperature, etc.).

4.2 Basics for the Proposed Method

In Figure 12 the SCA characteristics for two different configurations (N_pxN_s=2x3 and 1x6 with Ns: modules number in series and Np branches number in parallel)

as well as that of two different loads Za and Zb are plotted. In this case, the irradiance and temperature are kept constant. As can be seen the power corresponding to point A_1 is higher than that of point A_2, which means that the first configuration (2x3) provides better load matching for this load than the second configuration (1x6). Whereas for load Zb, the power on point B_1 is higher than that of point B_2, which means that in this case the second configuration is better. It is then necessary to select on line and in real time the SCA appropriate configuration for a given load and under given working conditions in order to reduce the mismatch losses [3] [13-14].

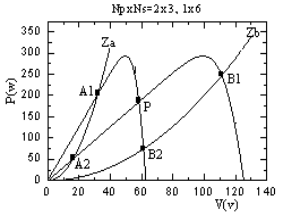


Figure 12
The P-V SCA and loads characteristics

Figure 12 shows that the two P-V SCA characteristics intersect on a point noted 'P'. At this operating point, either configuration could be used. Point 'P' is taken as a reference. For loads where their operating points are at the point 'P' left, configuration 1 (noted series configuration) should be used. And for all loads with operating points situated at the point 'P' right, configuration 2 (noted parallel configuration) is to be used. Therefore it is necessary to find a way to determine on line the point's 'P' position: the voltage Vcom and the current Icom corresponding to this point. Then, the load voltage and current is compared to Vcom and Icom. The SCA switches then onto the appropriate configuration.

4.3 Our Adopted Method

The principle of the proposed method consists of determining On-Line and in Real-Time for a fixed number of photovoltaic modules which configuration is the best for a given load under given working conditions and the switching the SCA into that configuration. Unfortunately, the reference point P is not fixed but varies with weather conditions (solar radiation, temperature, etc.). Therefore, it is necessary to determine the positions of point P as these conditions change; in other words, to find a way to determine the current (Icom) and the voltage (Vcom) corresponding to point P as these conditions vary. A simulation program has been written to determine the variation of the commutation point as the solar radiation and/or the temperature change. Figures 13 and 14 show the simulation results for an SCA made of 6 modules, allowed to commutate between two configurations (1X6 and 3X2) [3].

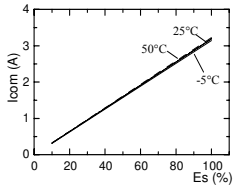


Figure 13
Variation of Icom versus solar radiation and temperatures

From Figure 13, it is clear that the commutation current varies linearly with the solar radiation and is hardly affected by the temperature. The simulation program has been run for other different configurations and the results led to the same conclusion. Consequently the current Icom could be written as follows.

$$Icom=Ki.Es \tag{19}$$

where Es is the solar radiation and Ki is a constant which depends on the number and the type of modules used.

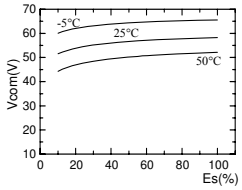


Figure 14
Variation of Vcom versus solar radiation and temperatures

The variation of Vcom as a function of the solar radiation for different temperatures has been determined and is illustrated in Figure 15. As can be seen, the characteristic is non linear at a low solar radiation level, so a simple linear procedure has been applied. As a consequence, for a given temperature the commutation voltage Vcom could be written as follows:

$$Vcom=ai.Es+bi \tag{20}$$

where ai and bi are constants which are derived from the linear procedure at 25°C. The Figure also shows that the commutation voltage Vcom is significantly affected by the temperature T. As T increases Vcom diminishes linearly. To implement the reconfiguration method using the commutation voltage Vcom as reference, equation (20) should be corrected in order to take into account the linear variation of Vcom as a function of the temperature.

After modification equation (20) becomes

$$Vcom=ai.Es+bi+Ai.(T-25) \tag{21}$$

where ai and bi are as defined previously and Ai is the slope of the curve Vcom=f(T) at 25°C.

4.4 Practical Results

The control system based on equations (19) and (21) has been implemented (Figure 15) and an example of the practical results is given in Figure 16 [3].

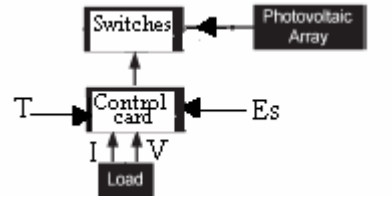


Figure 15
Our experimental system realized

A variable load has been used in order to cover different operating conditions. As can be seen, the SCA is switched into the appropriate configuration (Figure 16).

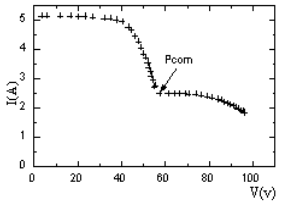


Figure 16
Practical characteristic I-V of the GP using the reconfiguration method at Es=78%, T=27 °C

Conclusion

The purpose of the work is to optimize the system's operation. The main reason for the described methods is to supply in an efficient manner a stand alone system using renewable energy sources such as photovoltaic ones. This study presents a simple but efficient method to improve a photovoltaic system's operation. The aim of the first method is to reach the MPP rapidly, in real time and automatically. It is used to maximize the photovoltaic array output power, irrespective of the temperature and irradiation conditions and of the load electrical characteristics. Thus, a new concept to track the MPP rapidly has been developed. The second method is an optimum switching of photovoltaic (PV) connected to a load, in real time. By rearranging the series and parallel connections between different modules, the matching between the load and SCA is improved. This enables the system to react to changes in loads and environmental conditions such as temperature and irradiance. The quality of load matching in a PV power system is improved by these proposed methods because the power available to the load is greater, whatever the changes in the load and/or in the climatic conditions. The experimental results show that the use of the two proposed method's control increases the PV output power. Our methods are in real time. Then, these methods take physical variations and aging of SCA and other effects such as shading into account.

References

[1] Tiberiu Tudorache, Mihail Popescu: FEM Optimal Design of Wind Energy-based Heater, Acta Polytechnica Hungarica, Vol. 6, No. 2, 2009, pp. 55-70

[2] Appelbaum J: The Operation of Loads Powered by Separate Sources or by a Common Source of Solar Cells: IEEE Transactions on Energy Conversion; Vol. 4, No. 3, September 1989, pp. 351-357

[3] Zerhouni F. Z: Développement et optimisation d'un générateur énergétique hybride propre à base de PV-PAC, Doctoral thesis, university of sciences and technology Mohamed Boudiaf USTOMB, Algeria, 2009

[4] V. Salas, E. Olias, A. Barrado, A. Lazaro: Review of the Maximum Power Point Tracking Algorithms for Stand Alone Photovoltaic Systems, Solar Energy Materials & Solar Cells, Vol. 90, 2006, pp. 1555-1578

[5] Daiki Tokushima, Masato Uchida, Satoshi Kanbei, Hiroki Ishikawa, Haruo Naitoh: A New MPPT Control for Photovoltaic Panels by Instantaneous Maximum Power Point Tracking, Electrical Engineering in Japan, Vol. 157, No. 3, November 2006, pp. 73-800

[6] Chen-Chi Chu, Chieh-Li Chen: Robust Maximum Power Point Tracking Method for Photovoltaic Cells: A Sliding Mode Control Approach, Solar Energy, Vol. 83, 2009, pp. 1370-1378

[7] D. Brunelli, D. Dondi, A. Bertacchini, L. Larcher, P. Pavan, L. Benin: Photovoltaic Scavenging Systems: Modeling and Optimization Microelectronics Journal, 2009, pp. 1337-1344

[8] Weidong Xiao, Student Member, Magnus G. J. Lind, William G. Dunford, and Antoine Capel: Real-Time Identification of Optimal Operating Points in Photovoltaic Power Systems, IEEE Transactions on Industrial Electronics, Vol. 53, No. 4, August 2006, pp. 1017-1026

[9] Enrique M. a, E. Dura´n a, Sidrach-de-Cardona bM., Andu´ jar J. M: Theoretical Assessment of the Maximum Power Point Tracking Efficiency of Photovoltaic Facilities with Different Converter Topologies, Solar Energy, Vol. 81, 2007, pp. 31-38

[10] Liu Liqun, Wang Zhixin: A Variable Voltage MPPT Control Method for Photovoltaic Generation System, wseas transactions on circuits and systems, Vol. 8, No. 4, April 2009, pp. 335-349

[11] Arias, J, Linera, F. F, Martin-Ramos, J, Pernia, A. M, Cambronero, J: A Modular PV Regulator Based on Microcontroller with Maximum Power Point Tracking, Industry Applications, 39th IAS Annual Meeting, conference Record of IEEE, 2, 2004, pp. 1178-1184, Industry Application Society Annual Meeting IAS'2004, Seattle,USA

[12] Weidong Xiao, Nathan Ozog, William G. Dunford: Topology Study of Photovoltaic Interface for Maximum Power Point Tracking, IEEE

Transactions on Industrial Electronics, Vol. 54, No. 3, June 2007, pp. 1696-1704

[13] Salameh, Z. M., Liang, C: Optimum Switching Points for Array Reconfiguration Controller, Conference Record of the Twenty First IEEE, Vol. 2, 1990, pp. 971-976

[14] Salameh Z. M, Mulpur A. K., Dagher F: Two Stage Electrical Array Reconfiguration Controller, Solar Energy, Vol. 44, 1990, pp. 51-55

A Coalgebra as an Intrusion Detection System

Daniel Mihályi, Valerie Novitzká

Department of Computers and Informatics
Faculty of Electrical Engineering and Informatics
Technical University of Košice
Letná 9, 042 00 Košice, Slovakia
valerie.novitzka@tuke.sk, daniel.mihalyi@tuke.sk

Abstract: In this paper we construct a coalgebra for an intrusion detection system to describe the behaviour of a packet stream together with selected actions in the case of intrusions. We start with an extension of the notion of the many-typed signature to the generalised signature and we construct the category of packets as a basic structure of our approach. A defined endofunctor captures the expected behaviour of the packet stream. The constructed coalgebra enables the description of the behaviour of the packet stream together with the reaction to intrusions.

Keywords: Coalgebra; Category theory; Intrusion detection system

1 Introduction

The main purpose of our research [5], [6], [7] is the construction of behavioural categorical models based on coalgebras for large program systems. There are only quite simple examples of using coalgebras in actual programs. In this contribution we show how it is possible to use our results for nontrivial systems from the area of real applications in informatics. We chose an Intrusion Detection System (IDS) to show how its behaviour can be modelled in categorical manner by a coalgebra.

The main purpose of an IDS is to disclose potential unwanted network activities. Many contemporary tendencies and trends are mostly pointed towards signature-based methods for attack-recognition. The idea of this method rests on the comparison of actually observed network traffic and the collection of known attack descriptions [4]. In our approach, we present another abstract means of notion *signature*. The well known notion of the universal algebra, a *many-typed signature* we extend to a *generalised signature*. Because there we deal with complex packet structures, we need to describe them in more complex mathematical structures. We use families of sets to describe heterogeneous informatic structures, e. g. records, and enclose them into a category.

In our approach, we formulate an IDS in the theory of coalgebras of semipolynomial endofunctors [3] over generalised signatures which are depicted in an abstract frame of category theory [1]. Our approach we formulate in the following steps:

1 first of all, we define a generalised signature containing the structure of treating packets and its chosen properties;

2 in the next step we construct the category of packets;

3 then we determine the semipolynomial endofunctor over this category;

4 afterward we characterize symptoms of network attacks and intrusions;

5 finally we excogitate the coalgebra of a semipolynomial endofunctor over a category of packets by means of which we describe the behaviour of infinite packet streams.

2 Generalised Signature

First of all we have to construct a *generalised signature* as an extension of the algebraic signature as a pair

$$\sum_p = (\bar{T}, F) \tag{1}$$

consisting of a finite collection of *Church's type names* $\bar{T}$ and a finite collection of *operation specifications* on Church's types denoted by *F*. This set includes the structure specification of treating packets e.g. *version*, *ttl*, *protocol,* etc. and structural properties of packets like *dsize*, *itype*, *content* etc. In operation specifications we distinguish three families:

1 *constructor-operation specifications* denoted by $f: \sigma_1 \lozenge \sigma_2 \rightarrow \tau;$

2 *deconstructor-operation specifications* denoted by $f: \sigma \rightarrow \tau_1 \lozenge \tau_2;$

3 *derived operation specifications* denoted by $f: \sigma \rightarrow \tau$ and $f: \sigma_1 \lozenge \sigma_2 \rightarrow \tau_1 \lozenge \tau_2.$

where $\sigma, \tau \in \bar{T}$ are arbitrary types from $\bar{T}$. The symbol $\lozenge$ is a placeholder for the type operation of product, coproduct and function. Then the specification of the signature Σ_p is treating a packet which we denote by *p*.

Table 1
IDS Signature

BEGIN Signature
Σ_p
Begin types

$\bar{T}$ ={*actions, protocols, ips, port, message, natip, nat0, char*}
End types
Begin opns
$\mathcal{F}$ = {*alert, log, drop, activate:* → *actions,*
255, icmp, tcp: → *protocols,*
ttl: → *nat0,*
port: → *nat0,*
mac addr: hex x *hex* x *hex* x *hex* x *hex* x *hex* → *mac,*
ip addr: natip x *natip* x *natip* x *natip* → *ips,*
ver : → *nat,*
message : → *char,*
dsize : → *nat0,*
content : → *char,*
itype : → *nat0}*
End opns
END Signature

3 Category *Packet*

In the second step we need to construct the category $Packet$ (Figure 1) of packets, where objects are treated packets denoted by $p_1, p_2, \ldots$ as non trivial heterogeneous structures – records, and morphisms *next:* $p_i \rightarrow p_{i+1}$ express homomorphous transition into the next packet of a given stream.

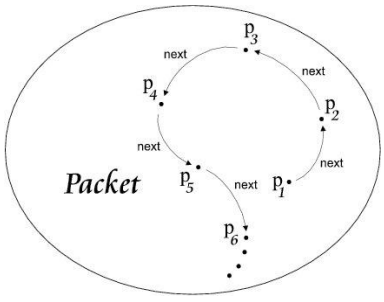


Figure 1
The category of packets

For any object p holds the universal mapping property mentioned in [1] in the following way: for any object $p \in Packet_{Obj}$ and projection morphisms $f:p \rightarrow ver$, $g:p \rightarrow ttl$, $h:p \rightarrow protocol$, $i:p \rightarrow s\_addr$, $i:p \rightarrow d\_addr$ there exists one (multiple) morphism

$$\langle f, g, h, i, j \rangle : p \rightarrow Nat \times Nat0 \times Protocols \times IPs \times IPs \tag{2}$$

depicted in the Figure 2 by dashed arrow.

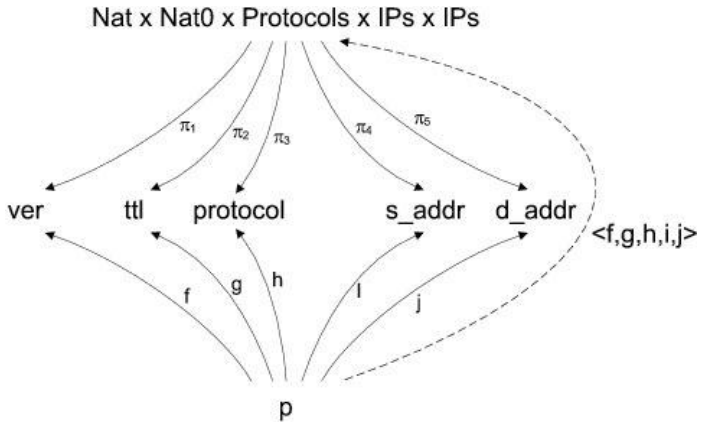


Figure 2
The universal projection property on structure p

3.1 Stream Automata

With respect to problems related to intrusion detections, we start from the theory of *stream automata* published in [2]. The authors represent trivial models of dynamical systems behaviour on infinite streams consisting of set elements. For instance, we can define an automata as a triple

$$SA = (Q, hd : Q \rightarrow P, tl : Q \rightarrow Q) \tag{3}$$

where Q is a set of (internal) states, $hd: Q \rightarrow P$, resp. $tl: Q \rightarrow Q$ are *head* resp. *tail* functions of a given stream.

If we consider trivial packets we also get a "trivial system" that can be described by display and one button. Then we can enunciate the principle: display packet when the button is pressed.

3.2 Coalgebra without Detection

In our approach, for a given trivial stream of packets without intrusive detection, we introduce the appropriate coalgebra *(ρ_p <hd,tl>)* in the following way. Infinite stream of packets we denote by ρ_p as state space of the coalgebraic structure

$$\langle hd, tl \rangle : \rho_p \to p \times \rho_p \tag{4}$$

We specify stream coalgebraic operations head resp. tail as *hd:* $\rho_p \to p$, resp. *tl:* $\rho_p \to \rho_p$ where ρ_p represents morphism compositions in the category $Packet$

$$p_1 \xrightarrow{next} p_2 \xrightarrow{next} \dots \tag{5}$$

We can formulate dynamics (behaviour) of infinite stream ρ_p as a sequence

$$\left(hd(\rho_p), hd(tl(\rho_p)), hd(tl^2(\rho_p)), \dots\right) \tag{6}$$

where $p_1 = hd(\rho_p)$, $p_2 = hd(tl(\rho_p))$,…

4 Semipolynomial Endofunctor

Next, we construct a semipolynomial functor over objects and morphisms of the category $Packet$ as

$$T : Packet \to Packet \tag{7}$$

defined in the following way

$$T(p) = X \times p \tag{8}$$

and

$$T(next(p)) = X \times next(p) \tag{9}$$

where X denotes observed values of a given packet. Then, the transition coalgebraic structure has the following form

$$\langle hd, tl \rangle : \rho_p \to T(\rho_p) \tag{10}$$

This structure gives us some observations of the network behaviour from outside based on observable values.

5 The Coalgebra

5.1 The Coalgebra with Detection

Now we extend the coalgebra introduced in 4.2 to the coalgebra with detection of unwanted network intrusions.

For the demonstration example, we show in Table 2 three selected specifications *A,B,C* of usual network intrusions by [8], whereas their real intendment is in parenthesis. We can consider the values listed below in the form of equalities as the symptoms of a potential network remote attack.

If from captured packet are observed some known symptoms mentioned above, then the coalgebra (system) responds by making one of the following preferred reactions, such as

• *alert,* which generates appropriate attention on the screen,

• *log,* for intrusion protocolling,

• *drop,* which ignores the intrusive fact by throwing away the incriminated packet and activation.

Table 2
Specifications of network intrusions

A	*B*	*C*
(ICMP Ping NMAP)	*(TCP Portscan)*	*(DOS Cisco attempt)*
IP Protocol == icmp	*MAC Addr == MACDAD*	*Port == 80*
dsize == 0	*IP Protocol == 255*	*dsize == 1*
itype == 8	*IP TTL == 0*	*content == "|13|"*

Now we need to extend the definition of the semipolynomial functor to include the detection of the known intrusions. We can formalize the activity of the whole system by mapping

$$atack(p) \mapsto (p, next(p), intrusion\_type(p)) \tag{11}$$

where *intrusion_type(p)* is a function of the form

$$intrusion\_type(p) : I \to actions \tag{12}$$

where *I* is a particular type of intrusion.

5.2 The Coalgebra as an IDS

As the last step, we construct coalgebra as intrusion detection system

$$\left(\rho_p, \left\langle hd, tl, intrusion\_type \right\rangle\right) \tag{13}$$

which is explicitly characterised by the following operations

- Immediate observation of treating packet *hd:* $\rho_p \rightarrow p$
- State modification *tl:* $\rho_p \rightarrow \rho_p$ and
- Generation of appropriate action (*intrusion_type(p):* $\rho_p \rightarrow p \circledR actions^I$) in the form

$$\left\langle hd, tl, intrusion\_type \right\rangle : \rho_p \rightarrow p \times \rho_p \times p \circledR actions^I \tag{14}$$

where $p \circledR actions^I$ expresses the generation of the appropriate reaction *actions* according to the given intrusion type $I=A+B+C$ in the appurtenant field of packet p, i.e. coincidence was captured between an intrusive pattern of network traffic and symptoms from Table 2.

5.2.1 Example

Behaviour of the system described by the coalgebra (13) can be modelled "step by step" by the following sequence

$$\begin{aligned} (p_1, p_2, \ldots) \quad &\mapsto \quad (p_1, (p_2, p_3, p_4,), A \mapsto alert) \quad \mapsto \\ &\mapsto \quad (p_1, p_2, (p_3, p_4,), \varepsilon) \quad \mapsto \\ &\mapsto \quad (p_1, p_2, p_3, (p_4,), C \mapsto alert) \quad \mapsto \\ &\mapsto \quad \ldots \end{aligned}$$

In the event that one of intrusions A,B or C is detected, some of the predefined actions from signature Σ_p are performed.

The example shows the situation where on any pattern of network traffic are treated packets p_1, p_2, p_3, p_4. On the packet p_1 was captured intrusion "*ICMP Ping NMAP*" by the specification A from Table 2, and on the packet p_2 was captured intrusion "*DOS Cisco attempt*" by the specification C from the same table.

5.3 The Final Coalgebra

Finally we turn our attention to constructing the final coalgebra. Let $\mathcal{T}coalg$ be the category of coalgebras over the semipolynomial endofunctor T where objects are

coalgebras on infinite data structures and morphisms are structure preserved homomorphisms between coalgebras. Its final object is the final coalgebra

$$\left(\rho_w, \left\langle observer, nextstat, i_t \right\rangle\right) \tag{15}$$

over the semipolynomial endofunctor T where *observer* is the generalized operation for performing an immediate observation on a data element of infinite data structure, *nextstat* is the next state operation and i_t is generator of the appropriate action.

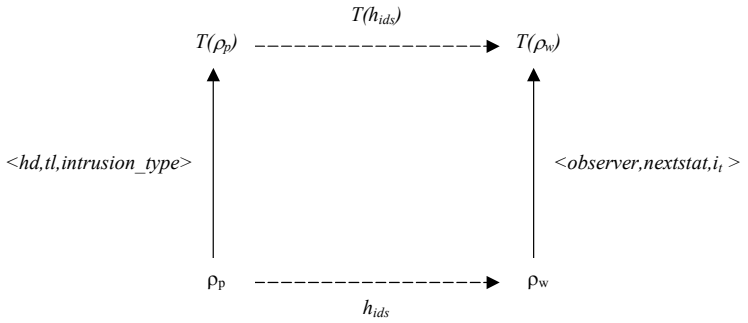


Figure 3
Homomorphism of the final coalgebra

For every operation *hd,tl* or *intrusion_type* of the intrusion detection coalgebra $(\rho_p, \langle hd,tl,intrusion\_type\rangle)$ in the packet state space ρ_p of the category *Packet* there exists a unique morphism (behavioural relation) in the category of coalgebras $Tkoalg$

$$\left\langle hd, tl, intrusion\_type \right\rangle \rightarrow \left\langle observer, nextstat, i_t \right\rangle \tag{16}$$

Where the diagram at Figure 3 commutes.

We call the homomorphism h_{ids}: $\rho_p \rightarrow \rho_w$ *infinite stream packet behavior* of a given computer network. This behaviour is realized stepwise by repeated evaluation of the coalgebraic structure. From these facts we see that the mapping h_{ids} captures stepwise particular packet observations by means of operation *hd*, which originate from the increased application of operation *tl*.

Conclusions

In this paper we have shown how coalgebras can be used for the modelling of real program systems. Our contribution contains the step by step construction of a coalgebra for an IDS. The constructed coalgebra describes the behaviour of infinite stream of packets with the detection of possible intrusions. This model covers also actions executed in the case of intrusions.

Our results demonstrate that coalgebras can be useful for a wide spectrum of large program systems. Of course, this paper deals only with one area of program systems, but in following research we will concern ourselves with the modelling of other systems, e.g. database systems or distributed systems by coalgebras.

In the future we would like to extend coalgebraic models with resource-oriented modal logics for proving bisimilarities on states produced by a system.

Acknowledgement

This work was supported by VEGA Grant 1/0175/08: Behavioural Categorical Models for Complex Program Systems.

References

[1] Barr, M., Wells, C.: Category Theory for Computing Science. Prentice Hall International (UK) Ltd., 66 Wood Lane End, Hertfordshire, UK, 1990

[2] Hasuo, I.: Modal Logics for Coalgebras-A Survey. Tech. rep., Institute of Technology, Tokyo, 2003

[3] Jacobs, B.: Introduction to Coalgebra. Towards Mathematics of States and Observations (draft), 2006

[4] Kazachkin, D. S., Gamayunov, D. Y.: Network Traffic Analysis Optimization for Signature-based Intrusion Detection Systems. Computational systems lab of Moscow State University's Faculty of Computational Math and Cybernetics, 2008

[5] Novitzká Valerie, Jenčík Marián, Mihályi Daniel, Slodičák Viliam, Ľaľová Martina: Behaviour of Program Systems in Terms of Categories, Computer Science and Technology Research Survey, Košice, Elfa, 2009, pp. 31-36, ISBN 978-80-8086-131-5

[6] Novitzká, V., Mihályi, D., Verbová, A. Coalgebras as Models of System's Behaviour. In AEI 2008, International Conference on Applied Electrical Engineering and Informatics '2008 (Athens, Greece, 2008), DCI FEI Technical University, Košice, pp. 31-36

[7] Slodičák Viliam, Mihályi Daniel: Coalgebras for Program Behavior in Toposes and Comonads, Proceedings of the Tenth International Conference on Informatics - Informatics 2009, Košice, Herľany, November 23-25, 2009, Košice, elfa, s.r.o., 2009, 10, pp. 125-135, ISBN 978-80-8086-126-1

[8] The Snort Team, s.-t. Snortusers manual. The Snort Project, 2008

Swarm Behavior of the Electromagnetics Community as regards Using Swarm Intelligence in their Research Studies

Asim Egemen Yilmaz

Department of Electronics Engineering, Faculty of Engineering
Ankara University
06100 Tandogan, Ankara, Turkey
e-mail: aeyilmaz@eng.ankara.edu.tr

Abstract: Recently, swarm intelligence and its applications have gained much popularity among researchers of various disciplines. The main aim of this study is to investigate the situation for the electromagnetic theory and microwave technology practitioners and try to point to an interesting analogy.

Keywords: Swarm intelligence; ant colony optimization; particle swarm optimization; electromagnetic theory; microwave theory

1 Introduction

The term "swarm intelligence", since its introduction by Beni and Wang in 1989 in the context of cellular robotic systems [1], has been a major multidisciplinary attraction center for researchers dealing especially with complex inverse (e.g. design and synthesis) problems. Typically, swarm intelligence systems consist of a population with members having some characteristic behaviors and interacting locally with each other within their environment. In these systems, the members individually behave freely to a certain extent and interact with each other. Even though there is no dictating centralized mechanism, these interactions yield a global behavior which is more organized and directive than that of a stand-alone individual.

Today, two main algorithms come to mind when the phrase "swarm intelligence" is mentioned. These are the Ant Colony Optimization (ACO) (considered initially in 1992 by Dorigo in his Ph.D. thesis [2], and later formalized by Dorigo et al. in [3]), and the Particle Swarm Optimization (PSO) (developed by Kennedy and Eberhart in 1995 [4]) methods. Both algorithms were developed by researchers observing the behaviors of animals living as swarms/colonies and being inspired

by them; and in more than a decade, they have proved to be successful in solving various complex problems due to their intelligent and systematic metaheuristic approaches. Originally, ACO was designed for combinatorial optimization problems, whereas PSO was designed for continuous ones. However, by now, successful versions of continuous ACO (e.g. [5-7]), and discrete PSO (e.g. [8-10]) have been developed.

ACO depends on the following principles: Initially, ants have random movements, but upon finding food they lay down pheromone trails returning home. Other ants have a tendency to follow these pheromones instead of keeping their random behavior. Over time, all pheromone trails start to evaporate and reduce their attractiveness. However, since pheromones over shorter paths are traced faster, and new pheromones are laid over the same path, the arate at which new pheromes are layed out surpasses the evaporation rate. Due to this positive feedback mechanism, the popularity of shorter paths (i.e. pheromone density) increases in an accelerated manner. This is the key of the success of the ACO for the solution of problems such as the traveling salesman problem.

On the other hand, PSO depends on the following main principles: The behavior of each individual in the swarm has three components, among which there is a balance. The first is random behavior (i.e. the tendency for searching and exploration), the second is social behavior (i.e. the tendency to observe and follow the other swarm members), and the final one is cognitive behavior (i.e. the tendency to revisit places of good memories). With these mechanisms, a swarm systematically searches the space, and the members are attracted rapidly to the best solution when found. This is the key of the success of PSO for the solution of multidimensional continuous optimization problems.

2 About the Analysis in this Study

It is a well known fact that swarm intelligence has gained much popularity and has diffused to a wide spectrum of research areas over the past two decades. Moreover, the increase in popularity and diffusion is still accelarating. A rough quantitative measure can be given as follows: As of July 2007, the number of citations of the 1995 dated original PSO proceeding (i.e. [4]) was about 1200 in Scopus, whereas as of January 2009, this number was almost 3000. Namely, the number of citations in the between July 2007 and January 2009 (a period of 1.5 years) is much more (almost 1.5 times) than the ones in the first 12 years.

In order to understand the rate of increase in the popularity of swarm intelligence among the electromagnetics community, an analysis of the publication archives was performed in this study. In fact, the results of a similar analysis were recently reported by Poli in a review article [11] (A more detailed version of Poli's study is

also available [12]). But unlike ours, the main aim of Poli's general-purpose study was to visualize the spread spectrum application areas of PSO, to identify each of these application areas, to construct an up-to-date PSO bibliography, and to perform taxonomy among the relevant publications. It is noteworthy that only the publications in IEEE Xplore database were included in Poli's analysis.

In this work, our main aim is to identify the increasing popularity of swarm intelligence (not only PSO, but also ACO) specifically among the electromagnetics community. Databases other than IEEE Xplore have also been included in the study. For this analysis, in addition to the major periodicals of electromagnetic theory and microwave technology, major general-purpose electrical/electronics engineering periodicals and the publications devoted to optimization and intelligence research have also been investigated. The periodicals included in this analysis are listed in Table 1. For a journal to be included in the analysis, the abstracting/indexing of the journal (i.e. the fact that the journal is being indexed in Thomson Scientific's Science Citation Index, Science Citation Index Expanded, etc.) were not considered, since there are many newly incepted journals which are not in some major indexes yet. The only restrictions for inclusion were that each journal would have ISSN numbers and would publish articles with original contributions after a peer-review process.

Since the archives of the major international symposium/conference proceedings are not as completely accessible and extensively searchable as the journal archives, the analysis was limited to journal and magazine publications. As another general rule, review articles were not accounted; only research articles were been considered.

By the time this analysis was performed, the results regarding the year 2009 were still immature; hence, the articles already published in 2009, or the ones with DOI but scheduled for publication in the upcoming issues in 2009 (i.e. articles in press) were not included in the analysis. For example, queries for Elsevier's *Expert Systems with Applications* returned one article which was still in press; hence, the relevant periodical is identified with a "×" (i.e. No Match) in Table 1.

For electromagnetic theory and microwave technology periodicals (e.g. *IEEE Transactions on Antennas and Propagation*, Taylor & Francis' *Electromagnetics*, etc.), any of the terms "swarm", "colony", "PSO", "ACO" (i.e. with the Boolean expression OR) was searched inside the Title, Abstract, and Keyword fields. The results of the queries were examined one by one in order to eliminate any occurrences of these terms for other purposes. For example, the query for *IEEE Transactions on Magnetics* returned some papers including the word "colony" of the "colony matrix" term, or the word "swarm" inside the "swarm points" term; such occurrences were identified and such papers were not accounted.

Table 1

List of investigated journals/magazines and information whether they have published papers/articles matching with the search criteria (*in alphabetical order with respect to the publisher*)

Publishing Institute / Company	**Related Journal / Magazine Name**	**Match**
ACES (American Computational Electromagnetics Society)	ACES Journal	√
AGU (American Geophysical Union)	Radio Science	×
Brno University of Technology	Radioengineering	√
CRL Publishing	Engineering Intelligent Systems	×
Electromagnetics Academy	PIER (Progress in Electromagnetics Research)	√
	PIERS Online	√
	PIER Letters	×
	PIER B	√
	PIER C	×
	PIER M	√
Electromagnetics Academy / Brill Publishing	Journal of Electromagnetic Waves and Applications (JEMWA)	√
Elsevier Science	AEÜ – International Journal of Electronics and Communications	√
	Applied Soft Computing	×
	Computers & Electrical Engineering	×
	Expert Systems with Applications	×
Emerald Group	The International Journal for Computational Mathematics and Electrical & Electronic Engineering (COMPEL)	×
ETRI (Korean Electronics and Telecommunications Research Institute)	ETRI Journal	×
Hindawi Publishing Corporation	International Journal of Antennas and Propagation	√
	International Journal of Microwave Science and Technology	×
	Journal of Artificial Evolution and Applications	√
IEEE (Institute of Electrical and Electronics Engineers)	Antennas and Propagation Magazine	√
	Antennas and Wireless Propagation Letters	√
	Microwave Magazine	×
	Microwave and Wireless Components Letters	×
	Transactions on Antennas and Propagation	√
	Transactions on Evolutionary Computation	×
	Transactions on Magnetics	√
	Transactions on Microwave Theory & Techniques	√
IEICE (Institute of Electronics, Information and Communication Engineers)	Electronics Express	×
	Transactions on Communications	×
	Transactions on Fundamentals of Electronics, Communications and Computer Sciences	×
IET (Institution of Engineering and Technology)	Electronics Letters	√
	Microwave, Antennas & Propagation	√
Inderscience Publishers	International Journal of Computational Science and Engineering (IJCSE)	×
	International Journal of Artificial Intelligence and Soft Computing (IJAISC)	×
IOS Press	International Journal of Applied Electromagnetics and Mechanics	√
John Wiley & Sons	Microwave and Optical Technology Letters	√
	International Journal of RF and Microwave Computer-Aided Engineering	√
	Expert Systems – The Journal of Knowledge Engineering	×
MIT Press	Evolutionary Computation	×
Research India Publications	International Journal of Computational Intelligence Research (IJCIR)	×

Table 1 (cont'd)

Schiele & Schön	Frequenz – Journal of RF Engineering and Telecommunications	√
Springer Verlag	Electrical Engineering (Archiv für Electrotechnik)	√
	Journal of Computational Electronics	×
	Radioelectronics and Communications Systems	×
	Swarm Intelligence	×
Taylor & Francis	Electromagnetics	√
	International Journal of Electronics	√
TÜBİTAK (The Scientific and Technological Research Council of Turkey)	Turkish Journal of Electrical Engineering and Computer Sciences (ELEKTRİK)	×
U.R.S.I. (Union Radio-Scientifique Internationale) Germany / Copernicus GmbH	Advances in Radio Science	√

For the general-purpose electrical/electronics engineering periodicals (e.g. Elsevier's *AEÜ – International Journal of Electronics and Communications*, IET's *Electronics Letters*, etc.) the same queries were performed. But in this instance, the query results were additionally examined in order to determine whether the applications in the studies were related to electromagnetic theory and microwave technology (by searching any of the terms "antenna", "propagation", "microwave", "electromagnetic", "wavelength" (i.e. with the Boolean expression OR) inside the Full Texts (or in Any Field if the feature is supported by the publisher's search engine)). The ones not related to electromagnetic theory or microwave theory were eliminated.

For the optimization and intelligence research periodicals (e.g. *IEEE Transactions on Evolutionary Computation*, Wiley's *Expert Systems*, etc.), in addition to the standard query, any of the terms "antenna", "propagation", "microwave", "electromagnetic", "wavelength" (i.e. with the Boolean expression OR) were also searched inside the Full Texts (or in Any Field if the feature was supported by the publisher's search engine). Similarly, the results of the queries were one by one examined in order to eliminate low relevancies.

Tables 2 and 3 list the annual distribution of the published articles in non-IEEE and IEEE periodical publications, respectively. It should be noted that the years before 1999 have also been included in the analysis; but since there are no publications in those years, they have been omitted in the tables for space considerations. In Figure 1, the total distributions can be seen as a bar graph.

Table 2

Number of electromagnetic theory related papers/articles using swarm intelligence (periodicals not published by IEEE) - (*order of appearance as in Table 1*)

Journal / Magazine Name	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**Journal Total**
ACES Journal				1				1		**2**
Radioengineering			1			2	2	1	2	**8**
PIER (Progress in Electromagnetics Research)								3	6	**9**
PIERS Online									1	**1**
PIER B									2	**2**
PIER M									1	**1**
Journal of Electromagnetic Waves and Applications (JEMWA)			1		1	1	2	3	2	**10**
AEÜ – International Journal of Electronics and Communications									1	**1**
International Journal of Antennas and Propagation									1	**1**
Journal of Artificial Evolution and Applications									1	**1**
Electronics Letters						1		1	3	**5**
Microwave, Antennas & Propagation								1		**1**
International Journal of Applied Electromagnetics and Mechanics								2		**2**
Microwave and Optical Technology Letters				1		2	6	7	4	**20**
International Journal of RF and Microwave Computer-Aided Engineering			1				1	2	2	**6**
Frequenz – Journal of RF Engineering and Telecommunications									1	**1**
Electrical Engineering (Archiv für Electrotechnik)									2	**2**
Electromagnetics							2	1	2	**5**
International Journal of Electronics					1					**1**
Advances in Radio Science							1			**1**
ANNUAL TOTAL	**0**	**0**	**3**	**2**	**2**	**6**	**14**	**22**	**31**	**80**

Table 3
Number of electromagnetic theory related papers/articles using swarm intelligence (periodicals published by IEEE) - (*order of appearance as in Table 1*)

Journal / Magazine Name	2000	2001	2002	2003	2004	2005	2006	2007	2008	Journal Total
Antennas and Propagation Magazine							1	1		**2**
Antennas and Wireless Propagation Letters						1	6	2	1	**10**
Transactions on Antennas and Propagation					2	5	2	11	7	**27**
Transactions on Magnetics			2		1	3	4	3	8	**21**
Transactions on Microwave Theory & Techniques						1			1	**2**
ANNUAL TOTAL	**0**	**0**	**2**	**0**	**3**	**10**	**13**	**17**	**17**	**62**

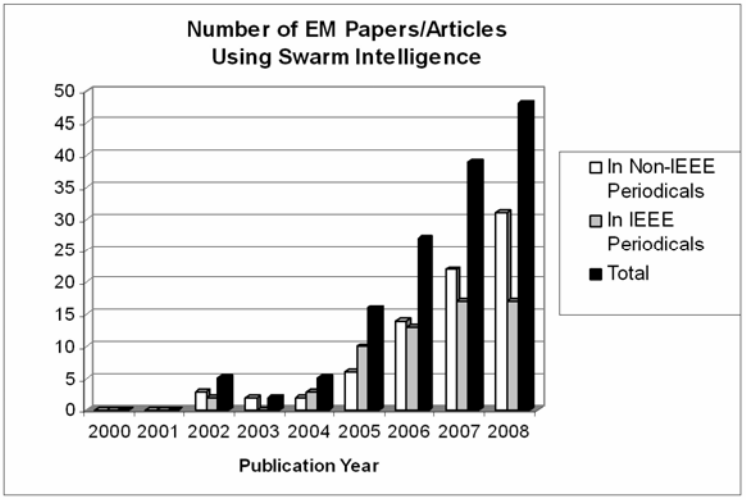


Figure 1
Annual increase in number of electromagnetic theory related journal papers/articles using swarm intelligence

3 The Analogy

Roughly speaking, the behaviors of all researchers resemble swarms and colonies, and the following analogies can be claimed.

From the ant colony perspective, a researcher lays out virtual pheromones to the "optimal publication" paths, where this optimality depends on his/her own personal criteria. In other words, most of the time, such criteria determine the address of submission for new manuscripts. These criteria might be the aim and scope, turn-around time, abstracting/indexing, the impact factor of the journal, interest and timing matches with any special issues, the tendency of the editorial/advisory board to the subject, etc. Optimal publication paths are attractive not only for that researcher but also for his fellows or any other researcher with similar areas of interest.

On the other hand, from the particle swarm optimization perspective, each researcher in the academic world has the tendency to explore new research areas, the application of novel methods to existing problems, etc. (i.e. random search and exploration behavior). At the same time, each researcher has the tendency to observe and follow the new trends, approaches, methods and applications in his/her own research area (i.e. social behavior) by means of tracking up-to-date literature. Moreover, each researcher has the tendency to revisit and extend his previous critically acclaimed and highly cited studies in order to create new publications (i.e. cognitive behavior).

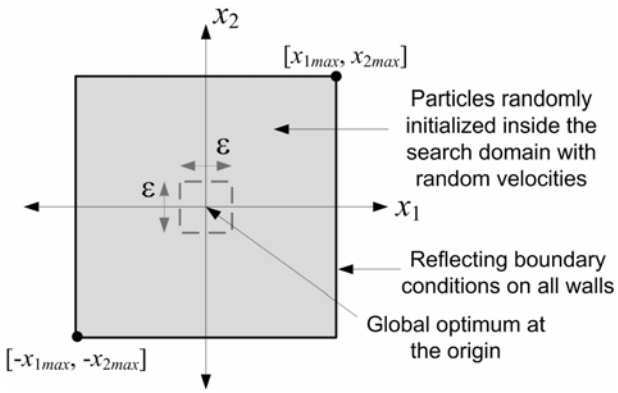


Figure 2
The two-dimensional search domain for the optimization problems

All the above facts are sufficient for a qualitative (and more or less subjective) analogy between the swarm/colony members and scholars. For a quantitative figure of the merit the the analogy, the following analysis was performed. Two-dimensional benchmark problems were solved by means of the particle swarm optimization, for which the search space is seen in Figure 2. For position and

velocity updates in each iteration, the formulation of Shi and Eberhart [13] was used:

$$v_n = c_1 v_n + c_2 u_1 (pbest - x_n) + c_3 u_2 (gbest - x_n) \quad (1)$$

$$x_n = x_n + \Delta t v_n \quad (2)$$

where v_n is the velocity component of the particle in the nth dimension, and x_n is its coordinate at the nth dimension. Certainly, these two operations are repeated for both dimensions. In these equations; the so-called inertial weight c_1, is a measure indicating the tendency to preserve the velocity along the previous course. Although not existing in the original PSO paper [4], inertial weight was introduced later by Shi and Eberhart [13] in order to improve the performance of the method; moreover, they showed that the ideal choice for the inertial weight is to decrease it linearly from 0.95 to 0.4 [14]. c_2 and c_3 are measures indicating the tendencies to converge to the *pbest* and *gbest*, which are the personal and global best positions, respectively. In early PSO researches, c_2 and c_3 were usually chosen to be 2.0; whereas for recent works 1.494 seems to be a more preferred value. u_1 and u_2 are random numbers between 0.0 and 1.0. The step size in time Δt is chosen to be unity for simplicity. The particles are kept inside the search space by means of the reflecting boundary conditions defined by Xu and Rahmat-Samii [15]. It is also common practice to put limitations to the velocities of the particles, since there is a probability that overspeedy particles jump over the global optimum. In this study, $|v_{n,max}|$ is set to 0.005.

As a first problem, global optimum for the two-dimensional sphere function (i.e.

$$f(\mathbf{x}) = \sum_{i=1}^{n} x_i^2 \quad (3)$$

where n=2) was searched; for which the number of particles inside ε vicinity of the global optimum at each iteration is investigated. It is expected that once the global optimum is found by any member, the others will be quickly attracted to that point. In order to have a more robust experiment, 1000 independent PSO executions were performed. The number of particles was taken to be 100; and the number of iterations was set equal to 100. The search space was chosen to be a square [-5000.0, 5000.0] × [-5000.0, 5000.0], where ε is chosen to be 0.1. For all independent executions, the global optimum was found by the swarm accurately (i.e. all the parameter settings mentioned in the previous paragraph and this paragraph were appropriate). The variation of the number of particles attracted to the global optimum (averaged over all executions) vs. the iteration is seen in Figure 3.

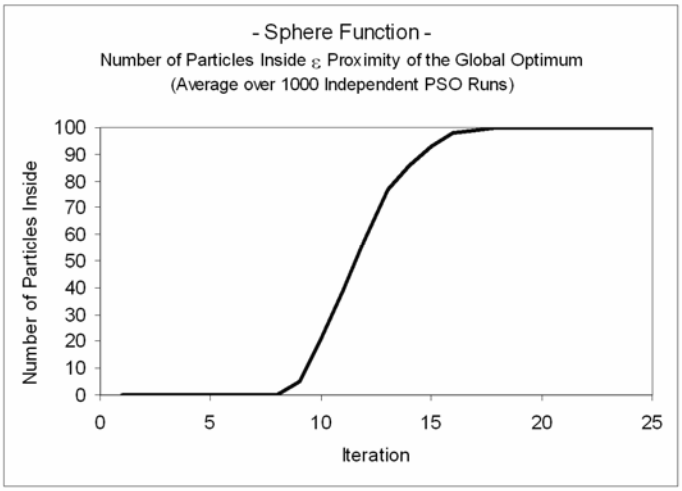


Figure 3
Number of particles around the global optimum vs. iteration for the two-dimensional sphere function

As a second problem, global optimum for the two-dimensional Rastrigin function (i.e.

$$f(\mathbf{x}) = 10n + \sum_{i=1}^{n} (x_i^2 - 10\cos(2\pi x_i)) \tag{4}$$

where n=2) was searched. Again, 1000 independent PSO executions were performed. This time, the number of particles is 100; and the number of iterations is 500. The search domain is a square [-5.12, 5.12] × [-5.12, 5.12], where ε is chosen to be 0.1. For all independent executions, the global optimum was found by the swarm accurately (i.e. all the parameter settings were appropriate). The variation of the number of particles attracted to the global optimum (averaged over all executions) vs. the iteration is seen in Figure 4.

Finally, the bar graph seen in Figure 1 is redrawn as an *XY*-line chart in Figure 5, where the years 1995-1999 are also added. As we see in Figure 3, for the sphere function, the number of particles around the global optimum is monotonically increasing. Moreover, the second derivative of the curve for the initial ramp-up phase is positive. With the parameter setup, 20-25 iterations became sufficient for the whole swarm to be attracted around the global optimum. This is most probably because of the fact that there are no local minima for the sphere function.

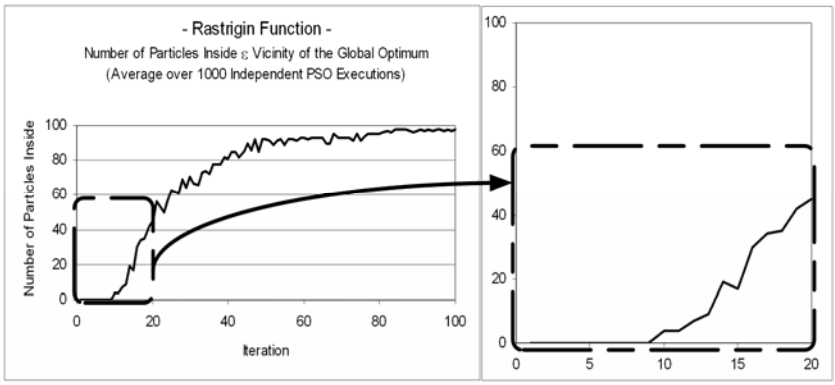


Figure 4
Number of particles around the global optimum vs. iteration for the two-dimensional Rastrigin function

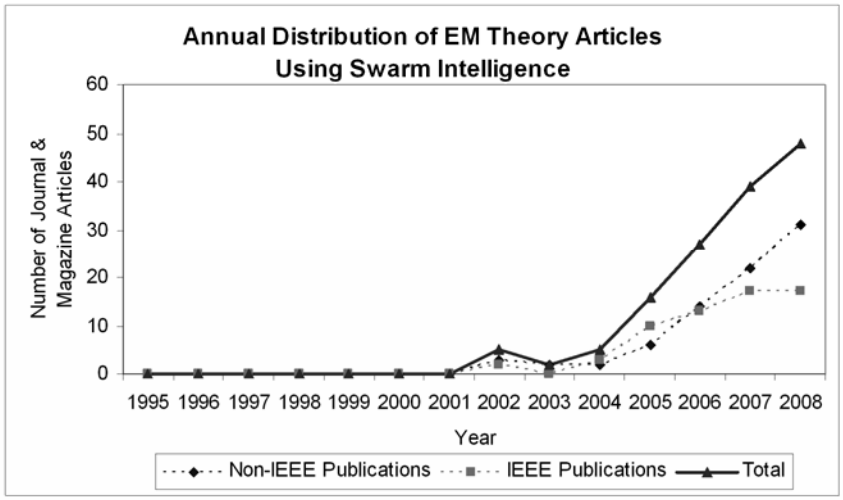


Figure 5
The *XY*-line chart version of the graph seen in Figure 1

For the Rastrigin function, on the other hand (Figure 4), even though there is an overall increase in the curve, it can be seen that there are fluctuations at some points. In addition, although there is an accelerated ramp-up phase, it is not as smooth as in the sphere function case. Moreover, it takes more time for the whole swarm to be attracted to the global optimum. This is most probably because of the fact that there are multiple local optima for the Rastrigin function; and sometimes some particles are erroneously but temporarily attracted backwards by one of the local optima.

Comparing the characteristics of Figures 3, 4, and 5, we can see that for all cases it takes some time for any swarm member to encounter the optimum; and an attraction center is created afterwards. More specifically, the detail of Figure 4 is very similar to Figure 5 in character. This means that the behavior of the electromagnetics community resembles the behavior of the particles which search the global optimum of a multimodal function. This makes sense, since there are researchers (particles) which are sometimes temporarily attracted to other interesting research areas and topics (optima); and this causes fluctuations in the number of swarm intelligence related published articles.

Finally, a rough guess by looking at Figure 3, 4 and 5 can be made as follows: the number of such publications might be continue increasing over the next 5-10 years.

Conclusions

In this study, the behavior of the electromagnetics community as regards using swarm intelligence in their research studies is analyzed. It is observed that the attraction of the subject is similar to the attraction of a global optimum of a multimodal function; and the movements of the researchers are similar to those of swarm members. Although the "swarm intelligence" topic is semi-humorously chosen in this study, "the fact of being an attraction center" can be generalized to any promising new method or technique. Moreover, this type of behavior is not unique for the electromagnetics community; it is also applicable for other societies of various disciplines.

References

[1] G. Beni, J. Wang: Swarm Intelligence in Cellular Robotic Systems, presented at NATO Advanced Workshop on Robots and Biological Systems, Tuscany, Italy, 1989

[2] M. Dorigo: Optimization, Learning and Natural Algorithms (in Italian), PhD Thesis, Politecnico di Milano, Italy, 1992

[3] M. Dorigo, G. DiCaro, L. M. Gambardella: Ant Algorithms for Discrete Optimization, Artificial Life, Vol. 5, No. 2, pp. 137-172, 1999

[4] J. Kennedy, R. Eberhart: Particle Swarm Optimization, in Proceedings of the IEEE International Conference on Neural Networks, 1995, pp. 1942-1948

[5] G. Bilchev, I. C. Parmee: The Ant Colony Metaphor for Searching Continuous Design Spaces, in Proceedings of the AISB Workshop on Evolutionary Computation, Vol. 993 of LNCS, T. C. Fogarty, ed., Berlin: Springer-Verlag, 1995, pp. 25-39

[6] N. Monmarché, G. Venturini, M. Slimane: On How Pachycondyla Apicalis Ants Suggest a New Search Algorithm, Future Generation Computer Systems, Vol. 16, pp. 937-946, 2000

[7] J. Dréo, P. Siarry: A New Ant Colony Algorithm Using the Heterarchical Concept Aimed at Optimization of Multiminima Continuous Functions, in Proceedings of the Third International Workshop on Ant Algorithms (ANTS'2002) Vol. 2463 of LNCS, M. Dorigo, G. Di Caro and M. Sampels, ed.s, Berlin: Springer-Verlag, 2002, pp. 216-221

[8] J. Kennedy, R. C. Eberhart: A Discrete Binary Version of the Particle Swarm Algorithm, in Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics, 1997, pp. 4104-4109

[9] M. Clerc: The Swarm and the Queen: Towards a Deterministic and Adaptive Particle Swarm Optimization, in Proceedings of the IEEE Congress on Evolutionary Computation, 1999, pp. 1951-1957

[10] C. K. Mohan, B. Al-kazemi: Discrete Particle Swarm Optimization, presented at Workshop on Particle Swarm Optimization, Purdue School of Engineering and Technology, Indianapolis:IN, 2001

[11] R. Poli: Analysis of the Publications on the Applications of Particle Swarm Optimisation, Journal of Artificial Evolution and Applications, Vol. 2008, Article ID 685175, doi:10.1155/2008/685175

[12] R. Poli: An Analysis of Publications on Particle Swarm Optimisation Applications, Univ. Essex Comp. Sci. & Elect. Eng. Tech. Rep. CSM-469, [online], May 2007 (rev. Nov. 2007), Available online: http://www.essex.ac.uk/dces/research/publications/technicalreports/2007/tr-csm469-revised.pdf

[13] Y. Shi, R. Eberhart: A Modified Particle Swarm Optimizer, in Proceedings of the IEEE International Conference on Evolutionary Computation, 1998, pp. 69-73

[14] Y. Shi, R.C. Eberhart: Empirical Study of Particle Swarm Optimization, in Proceedings of the IEEE Congress on Evolutionary Computation, pp. 1945-1950, 1999

[15] S. Xu, Y. Rahmat-Samii: Boundary Conditions in Particle Swarm Optimization Revisited, IEEE Transactions on Antennas and Propagation, Vol. 55, No. 3, pp. 760-765, 2007

Skewness and Kurtosis in Function of Selection of Network Traffic Distribution

Petar Čisar

Telekom Srbija, Subotica, Serbia, petarc@telekom.rs

Sanja Maravić Čisar

Subotica Tech – College of Applied Sciences, Subotica, Serbia, sanjam@vts.su.ac.rs

Abstract: The available literature is not completely certain what type(s) of probability distribution best models network traffic. Thus, for example, the uniform, Poisson, lognormal, Pareto and Rayleigh distributions were used in different applications. Statistical analysis presented in this paper aims to show how skewness and kurtosis of network traffic samples in a certain time interval may be criterions for selection of appropriate distribution type. The creation of histogram and probability distribution of network traffic samples is also discussed and demonstrated on a real case.

Keywords: skewness; kurtosis; network traffic; histogram; probability distribution

1 Introduction

Skewness characterizes the degree of asymmetry of a given distribution around its mean. If the distribution of the data are symmetric then skewness will be close to 0. Positive skewness indicates a distribution with an asymmetric tail extending toward more positive values. Negative skewness indicates a distribution with an asymmetric tail extending toward more negative values.

A measure of the standard error of skewness (SES) can roughly be estimated, according to [9], as the square root of 6/N, where N represents the number of samples. If the skewness is more than twice this amount, then it indicates that the distribution of the data is non-symmetric and it can be assumed that the distribution is significantly skewed. If the skewness is within the expected range of chance fluctuations in that statistic (i.e. ± SES), that would indicate a distribution with no significant skewness problem.

Kurtosis characterizes the relative peakedness or flatness of a distribution compared with the normal distribution. For normally distributed data the kurtosis is 0. Positive kurtosis indicates a relatively peaked distribution. Negative kurtosis indicates a relatively flat distribution. As with skewness, if the value of kurtosis is too big or too small, there is concern about the normality of the distribution. In this case, a rough formula for the standard error for kurtosis (SEK) is the square root of 24/N. Since the value of kurtosis falls within two standard errors (i.e. ± SEK), the data may be considered to meet the criteria for normality by this measure. These measures of skewness and kurtosis are one method of examining the distribution of the data. However, they are not definitive in concluding normality. What should also be examined is a graph (histogram) of the data; and further, one should consider performing other tests for normality such as the Shapiro-Wilk or the Kolmogorov-Smirnov test.

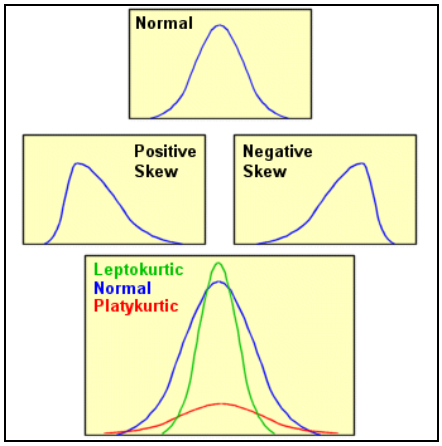


Figure 1
Illustration of skewness and kurtosis

Skewness is a network characteristic that can be successfully implemented in statistical anomaly detection algorithms, as is shown in [12]. When trying to distort the distribution parameters (mean, median, etc.), the attacker puts more weight into one of the tails of the probability density function, which leads to an asymmetric, skewed distribution. It is well-known that the standard characterizing parameters of a distribution are the mean (or median), the standard deviation, the kurtosis, and the skewness. It is reasonable to assume that there will be some empirical distribution of the skewness in the case when there is no attack is available, since most of the time, the system is not normally attacked. The statistical intrusion detection algorithm can compare the skewness of the sample to the expected value of the skewness in order to decide if an attack is taking place or not.

The authors have dealt with the topic of statistical intrusion detection in publications [1]-[8].

2 Different Types of Probability Distribution

Using the software package "Matlab", sequences of 40 random numbers with various types of probability distributions are generated (Figure 2 – function "Export"), as the way to simulate network traffic.

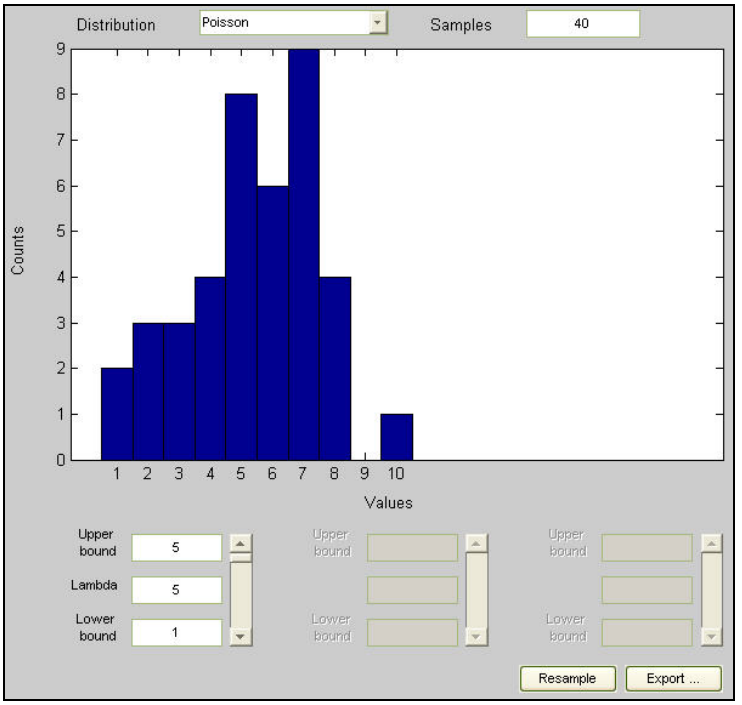
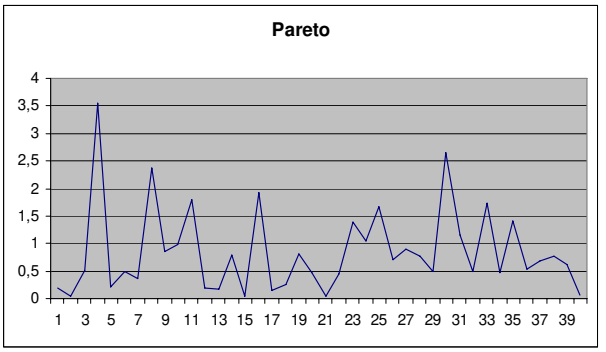
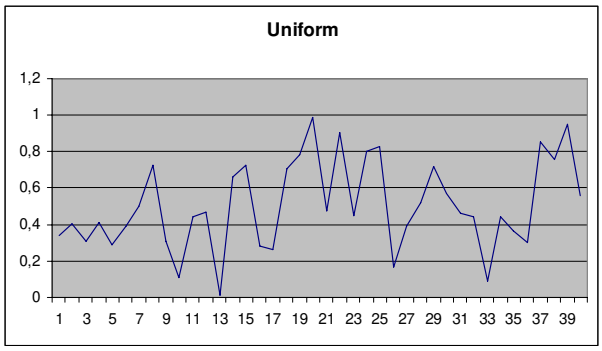
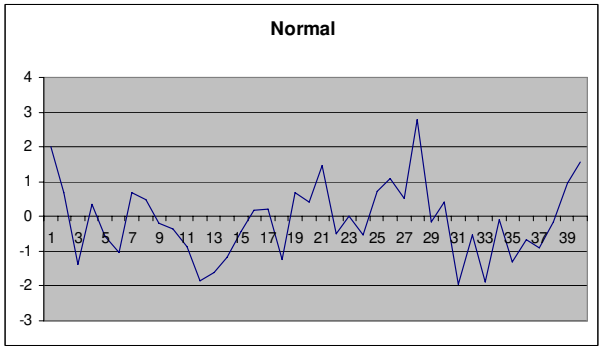


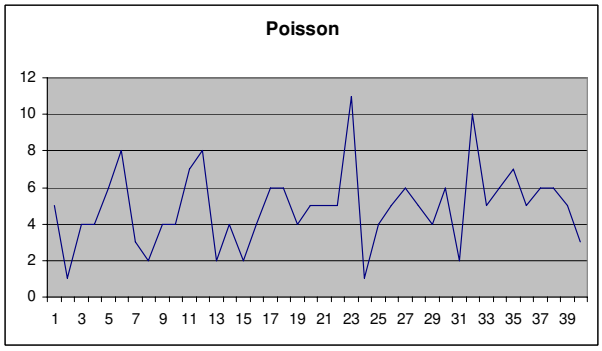
Figure 2
Random number generation (Poisson distribution)

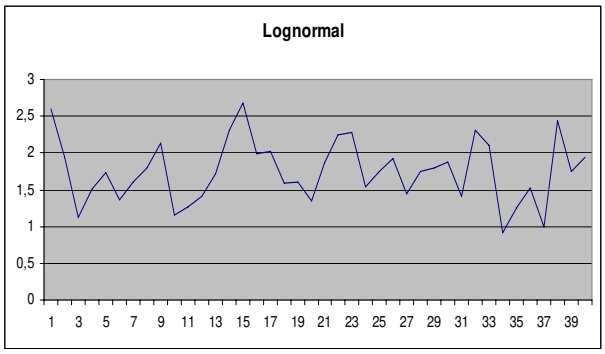
The following different types of distribution are examined: Pareto, normal, Poisson, lognormal, Rayleigh, chi-square and Weibull. Their graphical representation is shown in Figure 3.

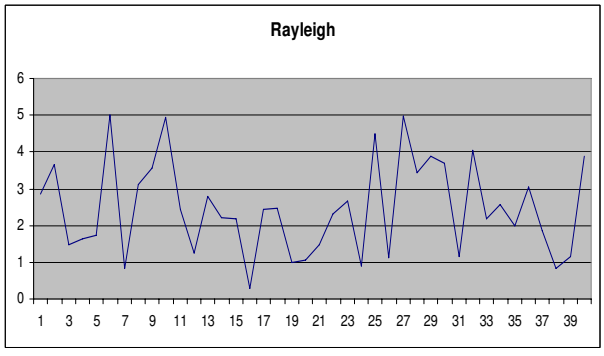


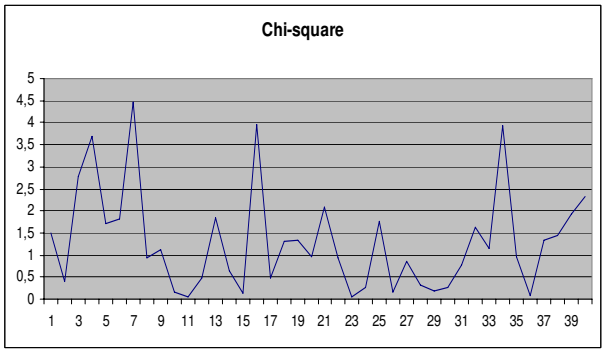

Uniform
1,2
1
0,8
0,6
0,4
0,2
0
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39


Normal
4
3
2
1
0
-1
-2
-3
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39


Poisson
12
10
8
6
4
2
0
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39


Lognormal
3
2,5
2
1,5
1
0,5
0
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39


Rayleigh
6
5
4
3
2
1
0
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39


Chi-square
5
4,5
4
3,5
3
2,5
2
1,5
1
0,5
0
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39

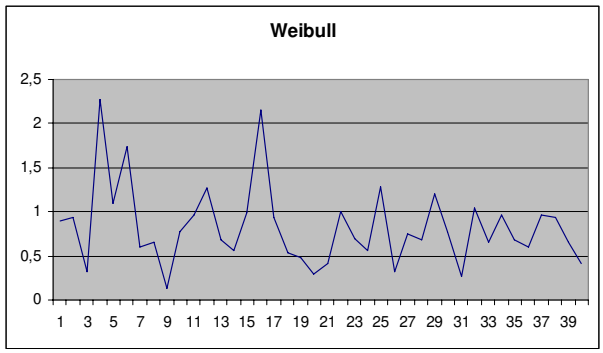


Figure 3
Examined types of distribution

For the purpose of the statistical comparison of generated curves, the authentic traffic samples y_t (local maxima) of a real ISP are also analyzed and graphically represented.

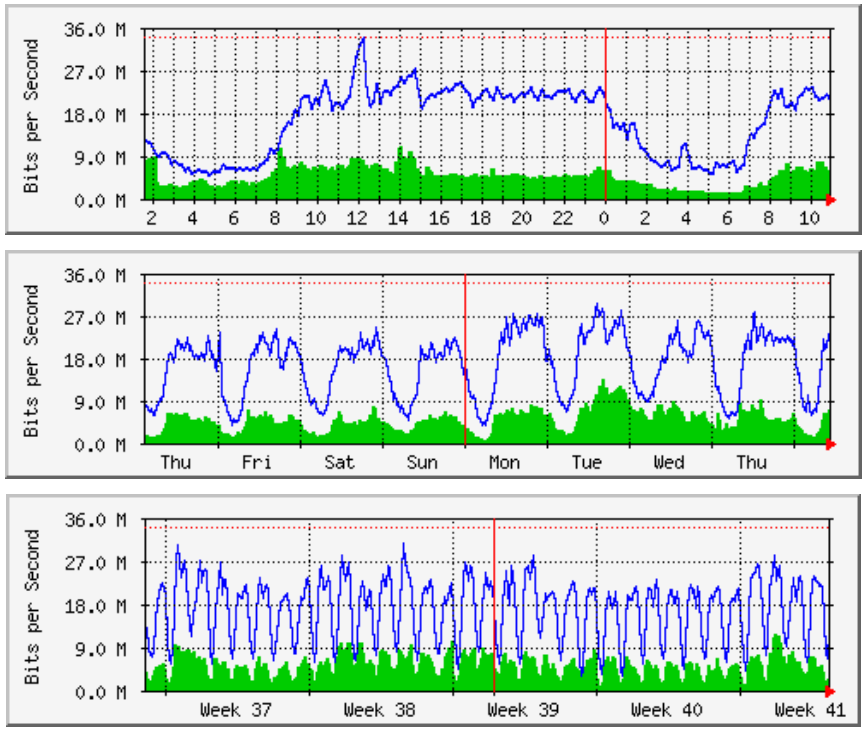


Figure 4
Network traffic curves

Table 1
Samples of network traffic

Sample	y_t (daily)	y_t (weekly)	y_t (monthly)
1	12	21	23
2	10.5	22.5	30
3	8.5	23	27
4	10.5	20	27
5	18	20.5	25
6	22	23.5	27
7	25.5	24	22
8	20	21	24
9	33.9	23	23
10	25	25	20
11	24	25.5	24.5
12	26.5	24.5	26.5
13	27.5	22	28
14	23	25.5	27
15	25	27	23
16	24	28	22.5
17	23	27	26.5
18	23	28	31
19	22	25.5	22.5
20	23	30	22.5
21	23	29	27
22	23	26.5	25
23	23	29	26
24	16	26.5	28
25	16	27.5	21
26	9	26	24
27	11.5	25	22
28	8.5	24	22
29	8.5	23.5	22
30	14	22	23
31	23	22.5	27
32	23	24	29
33	20	24	25
34	23	25	25
35	23	23	22

The numerical values of all samples are presented in the following table.

Table 2

Samples for various distributions

	Pareto	Uniform	Normal	Poisson	Lognormal	Rayleigh	Chi-square	Weibull	ISP (daily)
1	0.187	0.3418	1.9969	5	2.6035	2.8425	1.5095	0.8927	12
2	0.0449	0.4018	0.697	1	1.9416	3.6687	0.4015	0.9414	10.5
3	0.5183	0.3077	-1.3664	4	1.1205	1.489	2.7817	0.3218	8.5
4	3.5492	0.4116	0.363	4	1.5011	1.6232	3.6918	2.2777	10.5
5	0.2081	0.2859	-0.567	6	1.7366	1.7224	1.7237	1.1012	18
6	0.4941	0.3941	-1.0442	8	1.3615	5.012	1.8079	1.7357	22
7	0.3545	0.503	0.6971	3	1.6055	0.8327	4.4586	0.6054	25.5
8	2.3838	0.722	0.484	2	1.7944	3.1239	0.9332	0.6562	20
9	0.8559	0.3062	-0.1938	4	2.1348	3.5528	1.1275	0.131	33.9
10	0.9793	0.1122	-0.3781	4	1.1604	4.9411	0.1487	0.7701	25
11	1.7948	0.4433	-0.8864	7	1.2742	2.43	0.044	0.9571	24
12	0.1825	0.4668	-1.8402	8	1.4038	1.2626	0.4828	1.271	26.5
13	0.176	0.0147	-1.6282	2	1.7206	2.8012	1.8466	0.6853	27.5
14	0.7949	0.6641	-1.1738	4	2.3076	2.2195	0.6304	0.5569	23
15	0.0444	0.7241	-0.4154	2	2.6753	2.1782	0.1261	0.9892	25
16	1.9163	0.2816	0.1751	4	1.9844	0.294	3.9641	2.1489	24
17	0.1393	0.2618	0.2294	6	2.0198	2.4474	0.4758	0.9316	23
18	0.2621	0.7085	-1.2409	6	1.5909	2.4579	1.315	0.5318	23
19	0.8116	0.7839	0.7	4	1.6081	1.0003	1.332	0.4805	22
20	0.477	0.9862	0.4269	5	1.3498	1.0748	0.9661	0.2888	23
21	0.0495	0.4733	1.4548	5	1.8651	1.4749	2.0914	0.4108	23
22	0.4463	0.9028	-0.5102	5	2.2465	2.3204	0.947	1.0002	23
23	1.397	0.4511	-0.0067	11	2.2796	2.6605	0.044	0.6905	23
24	1.0421	0.8045	-0.5255	1	1.538	0.9073	0.26	0.5594	16
25	1.6721	0.8289	0.7177	4	1.7407	4.4974	1.771	1.2808	16
26	0.712	0.1663	1.0884	5	1.9303	1.1374	0.154	0.3163	9
27	0.8934	0.3939	0.5006	6	1.4374	4.9586	0.8429	0.7507	11.5
28	0.7689	0.5208	2.7718	5	1.7461	3.4222	0.3112	0.6776	8.5
29	0.4928	0.7181	-0.1603	4	1.8019	3.891	0.188	1.2044	8.5
30	2.6427	0.5692	0.4295	6	1.8782	3.6857	0.2644	0.7746	14
31	1.1575	0.4608	-1.9668	2	1.4134	1.1608	0.7791	0.2662	23
32	0.4969	0.4453	-0.546	10	2.3082	4.0349	1.6375	1.0455	23
33	1.7429	0.0877	-1.8884	5	2.1038	2.1679	1.1526	0.6499	20
34	0.4764	0.4435	-0.108	6	0.9098	2.5793	3.9352	0.9679	23
35	1.4026	0.3663	-1.3161	7	1.2547	1.9957	0.9714	0.6829	23
36	0.5321	0.3025	-0.6726	5	1.5198	3.0429	0.0734	0.5983	
37	0.6811	0.8518	-0.9024	6	0.9969	1.8625	1.3493	0.9609	
38	0.7662	0.7595	-0.1548	6	2.4398	0.8267	1.4533	0.9396	
39	0.6136	0.9498	0.9472	5	1.7478	1.1558	1.9199	0.6502	
40	0.0594	0.5579	1.5504	3	1.9379	3.8697	2.3274	0.421	

As emphasized in Chapter 1, if the skewness and kurtosis are within the expected ranges of chance fluctuations in that statistic (i.e. ± SES and ± SEK), this implies that the distribution has no significant skewness problem. The different distributions shown in the table above (N = 40) result in SEK = 1.55 and SES = 0.77. Also, for the ISP (N = 35) the values of SEK = 1.66 and SES = 0.83 were obtained. By calculating the kurtosis and skewness for each distribution, the values presented in the following table were obtained.

Table 3
Kurtosis and skewness for analyzed distributions

	Pareto	Uniform	Normal	Poisson	Lognormal	Rayleigh	Chi-square	Weibull	ISP
kurtosis	2.82	-0.64	0.11	1.09	-0.38	-0.69	1.08	2.96	-0.53
skewness	1.6	0.16	0.38	0.58	0.187	0.41	1.23	1.47	-0.44

In the table above, the values outside the calculated limits are marked in darker shading, indicating distribution types that have greater skewness and kurtosis than is allowed. In this sense, these distributions are not appropriate in applications that describe network traffic.

3 The Variations of Skewness and Kurtosis

Daily, weekly and monthly network traffic curves of the Internet user were analyzed in order to determine the extent of variations of skewness and kurtosis. Based on these curves, the appropriate descriptive statistics are calculated (Table 4).

Table 4
Descriptive statistics of network samples

	daily	weekly	monthly
Mean	19,75428571	24,68571429	24,85714286
Standard Error	1,090581087	0,430499863	0,455393185
Median	23	24,5	25
Mode	23	24	27
Standard Deviation	6,451964719	2,546871536	2,694142417
Sample Variance	41,62784874	6,486554622	7,258403361
Kurtosis	-0,526430851	-0,602769614	-0,614090489
Skewness	-0,441718333	0,172823583	0,343951523
Range	25,4	10	11
Minimum	8,5	20	20
Maximum	33,9	30	31
Sum	691,4	864	870
Count	35	35	35
Confidence Level (99.0%)	2,975535291	1,1745734	1,2424922

By analyzing the results obtained in the table above, it can be concluded that the values for skewness and kurtosis remain within allowable limits for all the time periods. Only in the case of skewness is a change of sign detected.

4 Histogram and Probability Distribution

According to [10], the purpose of a histogram is to graphically summarize the distribution of a univariate data set. The histogram graphically shows the following:

1 center of the data

2 spread (i.e. the scale) of the data

3 skewness of the data

4 presence of outliers

5 presence of multiple modes in the data

These features provide strong indications of the proper distributional model for the data.

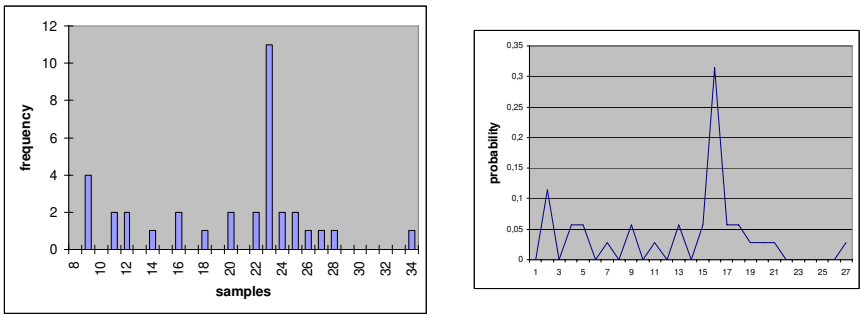
The most common form of the histogram is obtained by splitting the range of the data into equal-sized bins (called classes). Then for each bin, the number of points from the data set that fall into each bin is counted. That is

- vertical axis - frequency (i.e. counts for each bin)
- horizontal axis - response variable

The histogram of the frequency distribution can be converted to a probability distribution by dividing the tally in each group by the total number of data points to achieve the relative frequency [11]. In accordance with the aforementioned, the following graphical presentation related to network samples of ISP is obtained.

Based on Figure 5, it can be concluded that the most frequent samples approximately correspond to the mean of network traffic flow.

histogram **probability distribution**



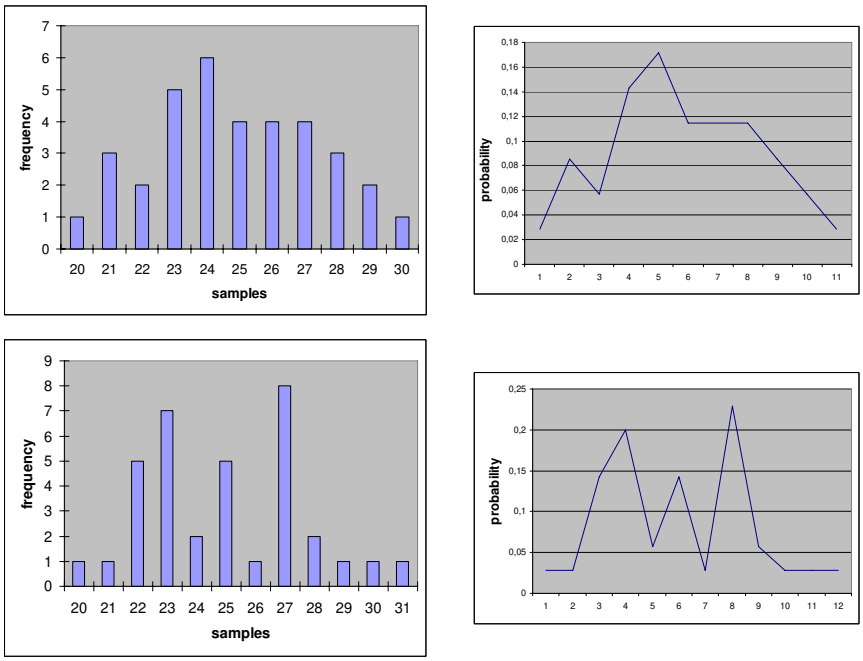


Figure 5
Histogram and probability distribution of network samples

Conclusions

Through analysis of the appropriate values, it can be concluded that the Pareto and Weibull distributions provide kurtosis and skewness that are significantly outside the permitted range. Looking at the skewness, it is necessary to note that the chi-square distribution gives values also higher than the allowed. Among the analyzed types of distribution, based on kurtosis and skewness as criteria for describing network traffic, the appropriate distributions are the uniform, the normal, the Poisson, the lognormal and the Rayleigh. It is also shown that the most frequent network sample in the histogram is approximately equal to the mean of the traffic flow.

In the next phase of research, the skewness and kurtosis of network traffic will be examined in terms of the reliability of network anomaly detection. The expectation is to try to find their application in the field of information systems security.

References

[1] Sorensen, S.: Competitive Overview of Statistical Anomaly Detection, White Paper, Juniper Networks, 2004

[2] Gong, F.: Deciphering Detection Techniques: Part II Anomaly-based Intrusion Detection, White Paper, McAfee Security, 2003

[3] SANS Intrusion Detection FAQ: Can You Explain Traffic Analysis and Anomaly Detection? Available at http://www.sans.org/resources/idfaq/ anomaly_detection.php

[4] Dulanović, N., Hinić, D., Simić, D.: An Intrusion Prevention System as a Proactive Security Mechanism in Network Infrastructure, YUJOR - Yugoslav Journal of Operations Research, Vol. 18, No. 1, pp. 109-122, 2008

[5] Lazarevic, A., Kumar, V., Srivastava, J.: Managing Cyber Threats: Issues, Approaches and Challenges, Chapter: A survey of Intrusion Detection techniques, Kluwer Academic Publishers, Boston, 2005

[6] Lončarić, S.: Osnove slučajnih procesa, Sveučilište u Zagrebu, Fakultet elektrotehnike i računarstva, Zavod za elektroničke sustave i obradbu informacija, Available at http://spus.zesoi.fer.hr

[7] Siris, V., Papagalou, F.: Application of Anomaly Detection Algorithms for Detecting SYN Flooding Attacks. Available at http://www.ist-scampi.org /publications/papers/siris-globecom2004.pdf

[8] CAIDA, the Cooperative Association for Internet Data Analysis: Inferring Internet Denial-of-Service Activity, University of California, San Diego, 2001

[9] Tabachnick, B. G., Fidell, L. S.: Using Multivariate Statistics (3rd ed.), New York: Harper Collins, 1996

[10] National Institute of Standards and Technology, Engineering Statistics Handbook, http://www.itl.nist.gov/div898/handbook/pmc/pmc.htm

[11] NetMBA-Business Knowledge Center, Available at http://www.netmba.com/ statistics/histogram/

[12] Buttyan, L., Schaffer, P., Vajda, I.: Resilient Aggregation with Attack Detection in Sensor Networks, Second IEEE International Workshop on Sensor Networks and Systems for Pervasive Computing (PerSeNS), IEEE Computer Society Press, Pisa, Italy, 2006

The Impacts of Natural Disasters on Power Systems: Anatomy of the Marmara Earthquake Blackout

Bülent Oral, Ferdun Dönmez

Marmara University, Istanbul, Turkey
boral@marmara.edu.tr, ferdun.donmez@holidayinnbursa.com

Abstract: An earthquake is an instantly occurring and unpredictable natural event. The potential and effects of earthquakes and other natural disasters on power systems are system faults. The faults are not only limited to the physical damage of power systems, but power quality disturbances may also take place. The Marmara Earthquake, which occurred in Turkey on Aug 17, 1999, caused death and catastrophe. After the earthquake, the Turkish Power System collapsed. This is the largest power blackout in Turkey in last twenty years. In this study, the impact of the Marmara earthquake on the Turkish power system are described and the Marmara Earthquake Blackout is examined in detail as regards the qualitative behavior of the power system.

Keywords: Power System Blackouts; Power System Disturbances; Overvoltages in Power Systems

1 Introduction

Disasters are sudden, uncontrollable, and mostly unexpected events. According to the source, the classification of disasters has two broad categories: as natural disasters and man-made disasters. A lot of disastrous events may be classified under the broad category of natural disasters, including earthquakes, hurricanes, tornados, avalanches, volcanic eruptions, landslides, floods, etc. On the other hand, terrorism, war, nuclear power plant accidents, and airplane crashes are examples of man-made disasters [1]. The impacts of natural disasters are often greatly prolonged and exacerbated by disruptions to critical infrastructure systems. Critical infrastructure includes the electric power system, water, and transportation [2].

The potential effects of earthquake and other natural disasters on the power system are system faults [3]. The faults are not only limited to physical damage of power systems, but power quality disturbances may also take place. They may cause

severe power outages. The definition of a power outage, measured by its duration, appears to be similar among countries. A customer interruption in the U.K. is defined as a power cut lasting more than 3 minutes, while it is 1 minute in Sweden and less than 5 minutes in the U.S. There is no official definition with regards to the size of blackouts. A large blackout could refer to incidents that affect over a million people in various locations [4].

Large blackouts are initiated by a single event that gradually leads to cascading outages and eventually to the collapse of the entire system [5]. They are more complex than smaller ones as they often involve cascading events in which the primary failure triggers a sequence of secondary failures that lead to a blackout in a large area of the grid [4]. Therefore, a cascading failure is the main mechanism of large blackouts [6].

Recent cascading failures in several power systems in the world require urgent attention. The US-Canada region failure during the August 2003 blackout affected 8 states in the U.S. and 2 provinces in Canada and left 50 million people in the dark. The Italian blackout of Sept 28, 2003 affected 57 million people. Around 19,000 MW of electricity load was lost over a 277,000 square kilometer area. The Scandinavian blackout of Sept 23, 2003 affected approximately 5 million people, cutting off around 3,000 MW of generating capacity in Sweden, and 1,850 MW in Denmark [4]. The US/Canada blackout, the Scandinavian blackout and the Italian blackout have shown that the technical issues related to power system security cannot be completely overcome. It is necessary to investigate the failure, to analyze the cause leading to the blackouts and to identify potential blackouts [7].

Over the last twenty years, there have been many significant blackouts for various reasons such as natural disasters, supply shortages, the restructuring of the electricity industry, etc. Apart from the news reported in the media, sources of information on blackouts are limited. However, anatomies of the blackouts have been studied by researchers in literature [2], [5], [7-12].

An earthquake with a moment magnitude of 7.4 on the Richter scale occurred on August 17, 1999 at 03:02 a.m. and affected the northwest of Turkey. The earthquake is called the Marmara Earthquake. After the earthquake, the Turkish Power System collapsed.

Although this was the largest power blackout in Turkey in last twenty years, the impacts of the earthquake on the power system and the causes of the large blackout have not been investigated in literature. The main goal of this paper is to investigate the Marmara Earthquake Blackout. The blackout is studied for its impact on the Turkish Power System in the context of the location and importance of the earthquake region and the theory of power system disturbances.

2 The Earthquake Region

On August 17, 1999 an earthquake with an epicenter near Izmit became the most terrifying disaster in recent Turkish history. The impact of the earthquake on the population and the economy was mainly felt in seven cities in the Marmara Region (Kocaeli, Sakarya, Yalova, İstanbul, Bolu, Bursa, and Eskisehir). The death toll was 18,373, another 48,901 people injured. Reportedly 93,000 housing units and 15,000 small business units collapsed or were badly damaged [13].

The Marmara Region, where the major impact of the earthquake occurred, is very important to the Turkish economy, both in terms of production and consumption capacities. This area accounts for 23% of the total population of Turkey. The seven cities, Kocaeli, Sakarya, Yalova, Istanbul, Bolu, Bursa, and Eskisehir, represent 33.4% of the Turkish GDP; further these cities produce 46.7% of total industry value added [14]. The Marmara region, mainly Kocaeli, Sakarya, and Yalova, is the center of the Turkish oil, textile, automobile, petrochemical, and tire industries. With an average income level per person that far exceeds the national average, the region also plays a very important role in terms of domestic consumption demand. Turkey main heavy industry is located in the Marmara Region. Figure 1 shows the location of the earthquake region on the map.

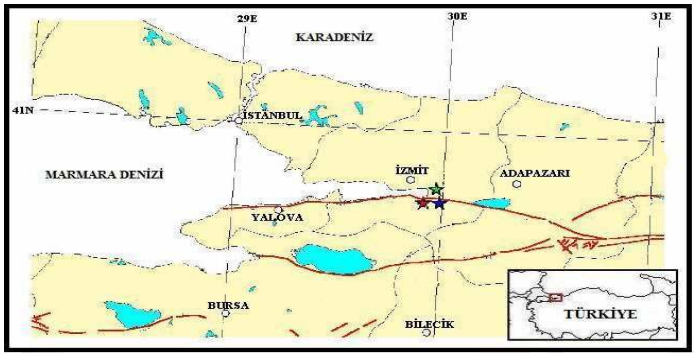


Figure 1
Location of the Earthquake Region

The region plays an important role regarding electricity consumption. At the end of 1999, the region used 32.2% of Turkey's electricity, which was 28 Billion kWh per anum [16]. In Table 1, the key indicators of the region are listed; the population, GDP and electricity consumption, and their percentages as a share of the total in Turkey.

Table 1
The Earthquake Region Indicators and Rations in Turkey

City	*Population*	*GDP ($)*	*%*	*Consumption (MWh)*	*%*
Bolu	554,473	184,628	1.0	730,288	0.8
Bursa	1,991,811	6,767,332	3.7	4,544,886	5.2
Eskisehir	662,599	2,275,641	1.2	849,913	1.0
Istanbul	9,382,894	40,506,151	21.8	16,509,300	18.8
Kocaeli	1,200,953	7,818,017	4.2	4,278,546	4.9
Sakarya	736,223	2,028,927	1.1	744,253	0.8
Yalova	166,617	715,127	0.4	595,927	0.7
Total	**14,695,570**	**60,295,823**	**33.4**	**28,253,113**	**32.2**

3 The Power System Characteristics

At the time of the Marmara Earthquake, the operational structure of the Turkish Power System was as follows. In 1999, TEAS (the Turkish Electric Generation Transmission CO.) was responsible for the generation and transmission of electricity throughout the country. Voltage levels of "National Grid Transmission System" were 154 kV and 380 kV. The length of the Turkish Power System transmission lines totaled 41,880 km, and the transformer capacities equaled 53,766 MVA, while 13,700 km and 18,470 MVA belonged to 380 kV [15]. An illustration of National Grid Transmission System for Turkey at 380 kV is displayed in Figure 2.

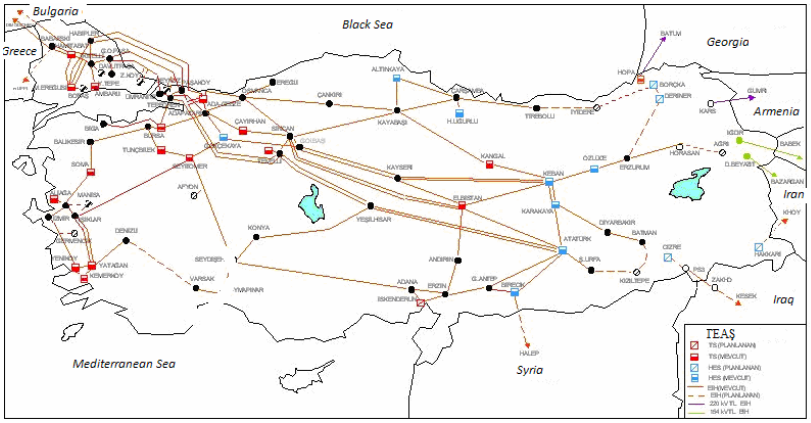


Figure 2
View of 380 kV Transmission System

The system was observed and managed by five regional operation divisions of National Load Dispatch Department. These were;

- Northeast Anatolia (NEA)
- Northwest Anatolia (NWA)
- West Anatolia (WA)
- Mid Anatolia (MA)
- Southeast Anatolia (SEA)

Each regional operation division was responsible in its region for load distribution and the operation of the interconnected system components. The operation divisions sent monthly reports to the National Load Dispatch Department. The connections among the regions are shown in Figure 3 [16].

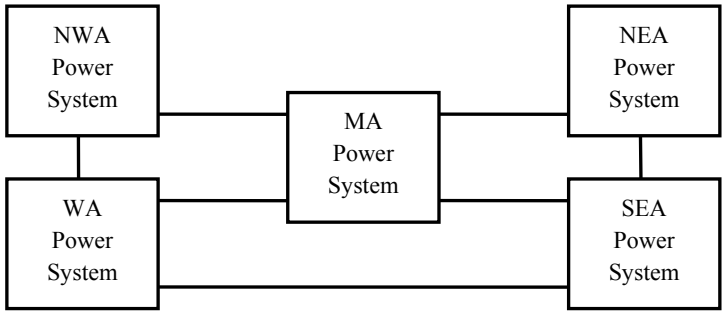


Figure 3
Connections among the Regions

In 1999, the installed power in Turkey was 26.1 GW. 60% was thermal, and 40% was from hydroelectric sources. 24.3% of the installed capacity came from lignite-fueled power plants. The most important of these were Afşin Elbistan (1,360 MW), Soma (1034 MW), Seyitömer (600 MW), Yatağan (630 MW), Yeniköy (420 MW) and Kemerköy (630 MW). Natural gas-fueled power plants, which accounted for the second major share in Turkey, produced 23.5% of installed capacity. The Power plants of Ambarlı (1,350 MW), Hamitabat (1,200 MW) and Bursa (1,432 MW) were natural gas-fueled power plants of the largest capacity [15]. All of them are in the Marmara region.

Figure 4 shows the monthly instantaneous peak loads for 1999. The instantaneous peak load of Turkish Power System was 18,938 MW at 17:30 on 8 December. Figure 4 also demonstrates the load distribution over the regions of the Marmara Earthquake, which was in month of August [15-17]. It was observed that the instantaneous peak load was higher in the earthquake region NWA than in the other regions.

Months	MW
January	17,137
February	17,392
March	17,671
April	16,320
May	15,758
June	16,042
July	17,088
August	**17,063**
October	16,063
September	16,818
November	18,170
December	18,938

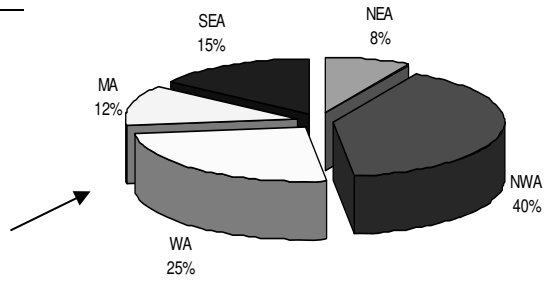


Figure 4

Monthly Instantaneous Peak Loads and load distribution over regions in August, 1999

Table 2 shows the hourly load values on the third Wednesdays of each month in 1999 [17]. In August, the minimum hourly load was 10,300 MW. Except in January, the minimum hourly load range was between 10,000 and 12,600 MW. However, the maximum load range was between 14,000 and 17,600 MW. As a result, the daily load fluctuation in 1999 was between 4,000 and 6,000 MW.

Table 2

The Hourly Load Values on the Third Wednesdays of Each Month in 1999

Date	The Hourly Load (MW)	
	Minimum	Maximum
20 January	8,300	12,200
17 Februray	11,300	16,800
17 March	11,600	17,600
21 April	10,400	15,400
19 May	10,000	14,000
16 June	10,400	15,300
21 July	11,200	16,200
18 August	**10,300**	**16,000**
15 September	10,500	15,000
20 October	10,250	16,400
17 November	11,750	17,750
15 December	12,600	17,750

4 The Earthquake Impact on the Power System

The Marmara Earthquake occurred on August 17 at local time 03:02 am. After the first shock, the Turkish Power System was badly affected and the system experience a blackout, except for some regions, WA and a region isolated from national grid, which were supplied by international connections (Bulgaria, Georgia and Iran) [16]. Therefore, except for WA, the thermal and hydroelectric power plants were out of service.

The time period of the minimum load drawn on the Turkish power system is usually between 02:00-07:00 am. This is a characteristic of Turkish power system. The Marmara Earthquake occurred during this period. Some of the daily load curves of August 1999 are presented in Figure 5 [17].

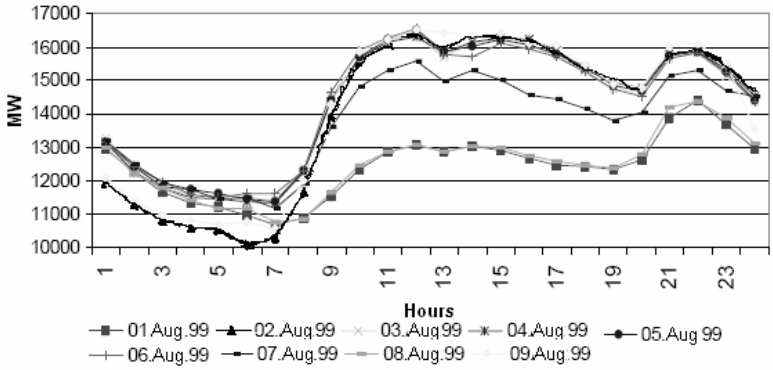


Figure 5
Daily Load Curves belonging to August 1-9, 1999

The generation-consumption balance of the regions on the day before the earthquake is given in table 3 [16]. NWA was the highest in both generation and consumption. In additional, WA was the region of second-highest consumption and was balanced in generation–consumption.

Table 3
Generation-Consumption of The Regions at the Time of Minimum Load on 16 August 1999 (MWh)

Region	**Generation**	**Consumption**
NWA	3,978	3,588
WA	2,234	2,574
SEA	2,622	1,420
MA	441	1,069
NEA	385	566
Total	**9,682**	**9,217**

At the time of the earthquake, the Turkish power system was supplied heavily by thermal plants. Figure 6 shows the power flow between the regions [18]. In general, the power flow of the Turkish power system is from the east regions, which have important hydroelectric power plants, to the west regions, where there are extensive industrial facilities. However, at the time of the earthquake, the power flow was not in that direction. The Marmara Earthquake occurred in the summer of 1999. The summer was dry, the water levels of the hydroelectric power plants were low, and the water was used for irrigation. At the same time, take or pay agreements had to be applied for natural gas power plants; therefore, the natural gas power plants were on service at the earthquake time.

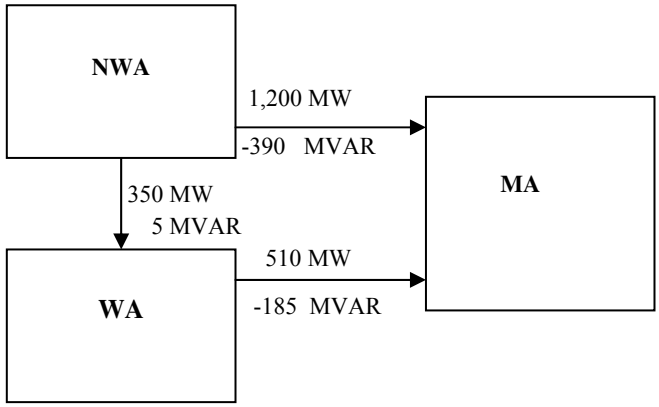


Figure 6
The Power Flow between the Regions at the Minimum Time Range on 17 August 1999

Consequently, the Marmara Earthquake occurred,

- at the time range of minimum load on the Turkish Power System;
- in Northwest Anatolia, which was the most developed region regarding the generation-consumption of electricity, industry and population density;
- during a situation when the power flow was to the east from the west.

When the earthquake occurred, the load of the NWA suddenly dropped. The region drew more than 35% of the total Turkey load. Excluding the isolated regions, the total drawn load of MA, NEA and SEA was roughly 4.0 GW. The load of WA was roughly 3.0 GW and 30% of the total drawn load of Turkey. Figure 7 shows the change to the system load during the earthquake [18]. Therefore, the load of the Turkish power system was suddenly rejected. As a result of disturbances to the power system, cascading failures occurred in the system.

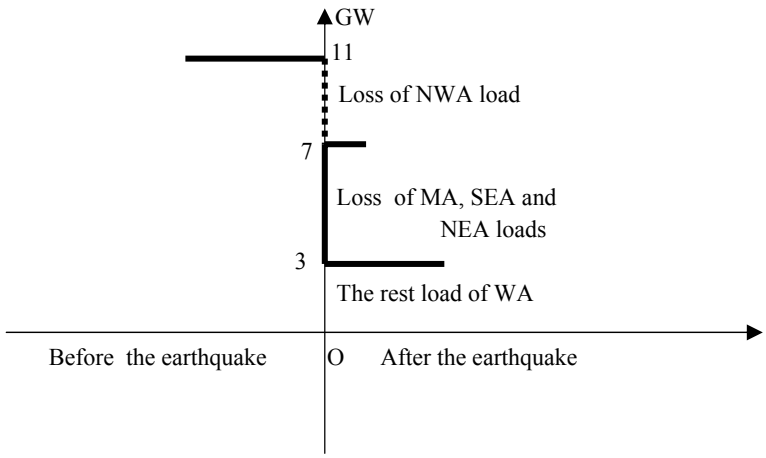


Figure 7
Sudden load reduction of Turkish Power System in the Marmara Earthquake

During the earthquake, the NWA power demand was covered by Hamitabat, Ambarli, and Unimar ve Bursa natural gas power plants. At the same time, the region exported power to MA. The power flow was roughly 1.2 GW [18]. The power connections were provided between NWA and MA by power transmission lines of 380 kV.

380 kV Osmanca and Adapazarı transformer substations, which are located between NWA and MA regions, and power transmission lines of NWA region were affected by the earthquake. Therefore, transformer substations of NWA were switched off, and the region1s load was suddenly lost.

The effect of a sudden loss of demand can develop the following possible results on the power system:

- System frequency rise,
- System voltage rise,
- Transmission overload,
- Transient instability,
- System oscillations.

These actions affect power system for particular instants. These time intervals can be from 1/10 seconds to seconds for system frequency and voltage; however, they can be from seconds to minutes for transmission overload and system oscillations [3].

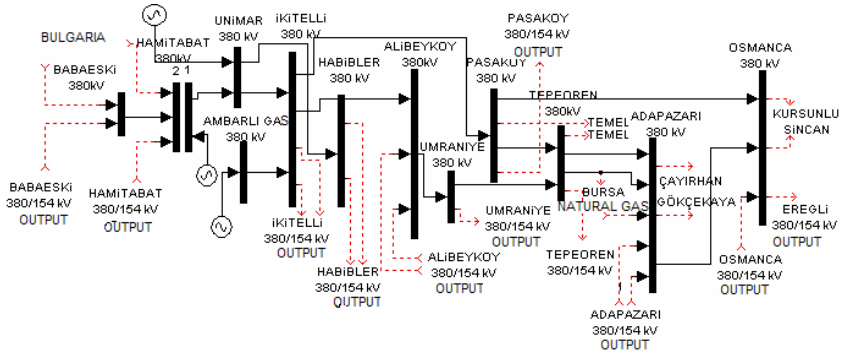


Figure 8
Power System of Northwest Anatolian Region (1999)

At the time of the earthquake, the NWA transmission and distribution lines were out of service because of pylon breaking and short circuits in the power lines. The substations in the severely affected earthquake regions were opened by buchholz relays. In addition, medium voltage and low voltage substations suffered damage. Consequently, immediately after the earthquake, the region load was rejected. At the same time, connection substations between NWA and MA suffered damage, when power was flowing from NWA to WA.

The rate of frequency change after the load rejection depends on the inertia constant and generation – load imbalance [19]. In other words, the power system frequency is directly related to the rotational speed of the generators supplying the system. There are slight variations in frequency as the dynamic balance between load generation changes. The size of the frequency shift and its duration depend on the load changes. Frequency variations that go beyond the accepted limits for normal steady-stead operation of the power system can be caused by faults such as a large block of load being disconnected, or a large source of generation going off-line on the bulk power transmission system [20]. At the instance of the load rejection, the breaker opens and the current rapidly goes to zero. The voltage drop over the generator internal impendence becomes zero, as well, causing a sudden step change in generator output voltage amplitude and phase. When running as a generator connected to the grid, the electrical frequency of the generator is synchronized to the grid frequency [19].

After the blackout, the power system restoration started quickly and it was recovered between 03:30 and 08:30 am. Normal operating conditions of the power system were provided with the exception of the earthquake region. During the system restoration process, an aftershock earthquake occurred in Duzce at 08:54 am. Because of this earthquake, the 380 kV lines between MA and NWA went out of service. Therefore, the Turkish Power System was interrupted again, with the exceptions of the regions powered by international connections and the Black Sea region. At 14:00 pm of the same day, the power system recovered and put into

service at normal operating conditions except for the region seriously affected by the earthquake [18].

Conclusions

Earthquakes are natural, sudden, unpredictable, uncontrollable and short-lasting natural events with destructive effects. The earthquakes not only damage the power system, but also make disturbances on power system. An earthquake occurred on August 17, 1999 in Turkey, and the Turkish Power System was affected by this earthquake and the power system collapsed.

- The blackout affected the biggest industrial and electricity consuming region of Turkey.
- The earthquake occurred at the time range of the minimum load on the Turkish Power System while the power flow was to east from west and the power system was operated predominantly by thermal power plants.
- When the earthquake occurred, the region load, which was more than 35% of the total load in Turkey, was suddenly rejected.

Because of the effect of this sudden loss of demand and cascading failures, disturbances on the power system occurred. As a result of these disturbances, instant frequency and voltage increased and a large blackout happened on the Turkish Power System. After twelve hours, the Turkish Power System recovered to its normal operating conditions.

Acknowledgement

The authors would like to thank Marmara University Scientific Research Projects Commission for its support in the form of a grant (FEN-YLS-290107-0041).

References

[1] Quarantelli E. L.: Different Types of Disasters and Planning Implications, Disaster Research Center, No. 169, 1991

[2] Chang S. E., McDaniels T. L., Mikawoz J., Peterson K.: Infrastructure Failure Interdependencies in Extreme Events: Power Outage Consequences in The 1998 Ice Storm, Springer Natural Hazards, Vol. 41, 2007, pp. 337-358

[3] Knight U. G.: Power Systems in Emergencies From Contingency Planning to Crisis Management, John Wiley& Sons Ltd., 2001

[4] Yu W., Pollitt M. G.: Does Liberalisation Cause More Electricity Blackouts? Evidence From a Global Study of Newspaper Reports, Cambridge EPRG Working Paper, No. 0902, January 2009

[5] Pourbeik P., Kundur P. S., Taylor C. W.: The Anatomy of A Power Grid Blackout Root-Causes and Dynamics of Recent Major Blackouts, IEEE Power and Energy Magazine, Vol. 4, No. 5, Sept.-Oct. 2006, pp. 22-29

[6] Nedic D. P., Dobson I., Kirschen D. S., Carreras B. C., Lynch V. E.: Criticality in A Cascading Failure Blackout Model, Elsevier Electrical Power and Energy Systems, V: 28, 2006, pp. 627-633

[7] Chen X., Deng C., Chen Y., Li C.: Blackout Prevention: Anatomy of the Blackout in Europe, in Proceedings of IEEE Power Engineering Conference, IPEC 2007, Singapore, 3-6 Dec., 2007, pp. 928-932

[8] Liu G. Y., Wang Y. J., Liu C. W.: Scenario Simulation of An Electric Power System and Application to Its Seismic Mitigation, in Proceedings of 4th International Conference on Earthquake Engineering, Taipei, Taiwan, 12-13 Oct., 2006, pp. 1-10

[9] Hines P., Apt J., Talukdar S.: Large Blackouts in North America: Historical Trends and Policy Implications, Elsevier Energy Policy, Vol. 37, 2009, pp. 5249-5259

[10] Bialek J. W.: "Recent Blackouts in U.S. and Continental Europe: is Liberalization to Blame?, Cambridge Working paper, No. 34, Department of Applied Economics, 2004

[11] Eleschova Z., Belan A.: Blackout in Power System, in Proceedings of 9th International Scientific Conference on Electric Power Engineering, Switzerland, 13-15 May, 2008, pp. 313-316, Brno

[12] Dizdarevic N., Majstrovic M., Coko S. C., Mandic N., Benovic J.: Causes, Analyses and Countermeasures with Respect to Blackout in Croatia on January 12, 2003, in Proceedings of the CRIS International Workshop on Power System Blackouts, Lund, Sweden, 3-5 May, 2004, pp. 1-21

[13] Ewing B. T., Kruse J. B., Ozdemir Ö.: Disaster Losses in The Developing World: Evidence from The August 1999 Earthquake in Turkey, Turkish Economic Association, Discussion Paper, No. 19, 2004

[14] DIE (State Institute of Statistics): 1999 Gross Domestic Product Results According to Provincials, 2000 [in Turkish]

[15] TEAS (Turkish Electric Generation Transmission CO): Electric Generation Transmission Statistics in 1999, APK-377, Ankara, 2000 [in Turkish]

[16] TEAS (Turkish Electric Generation Transmission CO): 1999 Operating Activities Report, Load Balancing Department, Ankara, 2000 [in Turkish]

[17] DPT (State Planning Organization): VIII. Development Plan Energy Special Commission Report, Ankara, 2001 [in Turkish]

[18] TEAS (Turkish Electric Generation Transmission CO): August 1999, Activities Reports, Load Balancing of Regions, Ankara, 1999 [in Turkish]

[19] Terzija V. V., Akke M.: Synchronous and Asynchronous Generators Frequency and Harmonics Behaviour after a Sudden Load Rejection, IEEE Transactions on Power Systems, Vol. 18, No. 2, 2003, pp. 730-736

[20] Dugan R. C., McGranaghan M. F., Santosa S., Beaty H. W.: Electrical Power System Quality, McGraw-Hill, Second Edition, New York, 2003

Liquefaction Resistance of Chlef River Silty Sand: Effect of Low Plastic Fines and other Parameters

Mostefa Belkhatir, Ahmed Arab, Noureddine Della

Civil Engineering Department
Hassiba Benbouali University of Chlef
Chlef, Algeria
E-mail: abelkhatir@yahoo.com, ah_arab@yahoo.fr, nour_della@yahoo.fr

Hanifi Missoum

Civil Engineering Department
University of Mostaganem
Mostaganem, Algeria
E-mail: hanifimissoum@yahoo.fr

Tom Schanz

Laboratory of Foundation Engineering, Soil and Rock Mechanics
Ruhr University
Bochum, Germany
E-mail: tom.schanz@ruhr-uni-bochum.de

Abstract: Silty sands are the most common type of soil that could be involved in both static and earthquake-induced liquefaction. Most of the recent earthquakes have revealed the liquefaction of silty sands. Therefore, the selection of the appropriate undrained residual shear strength of liquefied soils to be used in the assessment of the post-liquefaction stability of earth dams and other earth structures is becoming a major challenge. A series of undrained monotonic and cyclic triaxial tests were carried out on reconstituted saturated samples of sand with variation in the fines content ranging from 0 to 50% for the monotonic tests and from 0 to 40% for the cyclic ones, in order to study the influence of fines fraction and other parameters on the undrained residual shear strength and liquefaction potential of loose, medium dense and dense silty sand samples (Dr = 12%, 50%, 60% and 90%). The results of the monotonic tests show that the stress-strain response and shear strength behaviour is controlled by the percentage of fines fraction and the samples become contractive for the studied relative densities (Dr = 12% and 90%). The

undrained residual shear strength decreases as the global void ratio decreases and the fines content increases up to 30% fines content. Beyond that, it decreases with increasing the global void ratio and the fines content. Moreover, the undrained residual strength decreases linearly as the fines content and the inter-granular void ratio increase. Cyclic test results show that the increase of the fines fraction accelerates the liquefaction phenomenon for the studied amplitude and the liquefaction resistance decreases with the increase of the global void ratio and the loading amplitude. We notice that the reduction in the liquefaction resistance of Chlef sand-silt mixtures becomes very marked for the smaller cyclic stress ratios CSR = 0.15 and 0.25.

Keywords: silty sand; residual strength; fines content; undrained tests; liquefaction

1 Introduction

On October 10, 1980 at 13:25:23.7 local time (12:25:23.7 GMT) a destructive earthquake took place near Chlef City, Algeria (formerly known as El-Asnam). Chlef is approximately 200 km west of Algiers. The Richter magnitude, M, of this event was 7.2, which corresponds to a surface wave magnitude, M_s, of 7.3. The epicenter of the quake was located at 36.143° N and 1.413° E, 10 km east of Chlef. The focal depth of the earthquake was about 10 km, and the approximate duration was between 35 and 40 sec. The earthquake devastated the city of Chlef, population estimated at 125,000, and the nearby towns and villages. The large loss of life (reportedly 5,000 to 20,000 casualties) and property was attributed to the collapse of buildings. In several places of the affected area, especially along the Chlef river banks, great masses of sandy soils were ejected on to the ground surface level. Great damage to different structures was recorded.

During earthquakes, the ground shakes, causing cohesionless soils to lose their strength and behave like a liquid. This phenomenon is called soil liquefaction and will cause the settlement of buildings, landslides, the failures of earth dams, or other hazards. Liquefaction occurs due to an increase in the excess pore water pressure and a corresponding decrease in the effective overburden stress in a soil deposit. The understanding of the liquefaction phenomena has significantly improved in recent years. Most liquefaction research was carried out on clean sands with the assumption that the behaviour of silty sand is similar to that of clean sands. Recent researches made by [1; 2, 3, 4, 5, 6, 7, 8] indicate that sand deposited with silt content is much more liquefiable than clean sand. Also, strain properties and pore pressure generation in silty sand samples are quite different from clean sand. These new findings emphasize the importance of deposits with mixture of sand and silt. Moreover, the behaviour of silty sand soils such as hydraulic fills is not clearly known during earthquake. Therefore, a deep understanding of silty sand behaviour is needed for liquefaction assessment of silty sandy soils.

Further, the post-earthquake behaviour of silty sand and, consequently, the stability of structures founded on liquefied soil depends on the post-liquefaction shear strength of soil. The strength of soils mobilized at the quasi-steady state has an important implication for engineering practice [9]. Laboratory studies, especially in the last ten years, have contributed much to clarify the parameters that control the residual shear strength and provide some principles for selecting residual shear strength for design. However, the selection of suitable undrained residual shear strength for design and analysis is still not satisfactory [10].

Soil mixtures such as silty clays, clayey sands and silty sands are more commonly found in nature and in earth structures than pure sands, silts and clays. However, the behaviour of soil mixtures is not well understood since soil mechanics has concentrated mostly on pure soils. It has been indicated since the 1960s that the presence of fines will affect a sand's resistance to liquefaction. Nevertheless, a review of studies published in literature shows that no clear conclusions can be drawn regarding how altering the fines fraction affects the liquefaction resistance of a sand under monotonic and cyclic loading. This is particularly true for soils containing non-plastic silty fines. Several laboratory studies have been carried out, and have shown what appear to be conflicting results. Studies have reported that increasing the fines content in a sand will improve the liquefaction resistance of the sand [11,6], decrease the liquefaction resistance of the sand [12, 13, 14, 15, 16, 17, 18], or decrease the liquefaction resistance until some limiting fines content is reached, and then increase its resistance [19, 20, 22]. Moreover, numerous studies have reported that the behaviour of sand mixed with fine grained soils depends mainly on the fines content and the plasticity index [23]. Indeed, to a certain limiting value of fines content, fine-grained soils occupy solely the voids and do not influence the behaviour of the mixture. For this reason, the notion of skeleton void ratio was suggested to characterize those kinds of soils [24, 25].

2 Experimental Investigation

2.1 Materials Tested

Sand samples were collected from the liquefied layer of the deposit areas close to the epicentre of Chlef earthquake (October 10, 1980). Chlef sand has been used for all tests presented in this research. Individual sand particles are subrounded and the predominant minerals are feldspar and quartz. The tests were conducted on the mixtures of Chlef sand and silt. The liquid limit and plastic limit of the silt are 27% and 22% respectively. Chlef sand was mixed with 0 to 50% silt to get different silt content. The dry pluviation method was employed in the present study to prepare the soil samples for the monotonic and cyclic testing. Several laboratory investigations have reported that the behaviour of sandy soils can be

greatly influenced by the sample preparation methods. However, experimental data related to the effect of depositional methods on the behaviour of sand with low plastic fines is very limited because most of the previous studies have concentrated their efforts on clean sand [26]. According to [26], reconstituted samples prepared by dry methods appeared to exhibit a more contractive or unstable behaviour. The index properties of the soils used during this study are summarized in Table 1. The grain size distribution curves for the soils are shown in Fig. 1. The variation of e_{max} (maximum void ratio corresponding to the loosest state of the soil sample) and e_{min} (minimum void ratio corresponding to the densest state of the soil sample) versus the fines content F_c (the ratio of the weight of silt to the total weight of the sand-silt mixture) is illustrated in Fig. 2. We note that the two indices decrease with the increase of the fines content until F_c = 30%, then they increase with further increase in the amount of fines.

Table 1
Index properties of Chlef sand-silt mixtures

Material	Fines Content (%)	G_S	D_{50} (mm)	C_u	e_{min}	e_{max}	I_p(%)
Clean Sand	0	2.680	0.40	2.90	0.535	0.854	-
Silty Sand	10	2.682	-	-	0.472	0.798	-
Silty Sand	20	2.684	-	-	0.431	0.748	-
Silty Sand	30	2.686	-	-	0.412	0.718	-
Silty Sand	40	2.688	-	-	0.478	0.732	-
Silty Sand	50	2.69	-	-	0.600	0.874	-
Silt	100	2.70	0.004	29.4	0.72	1.420	5.0

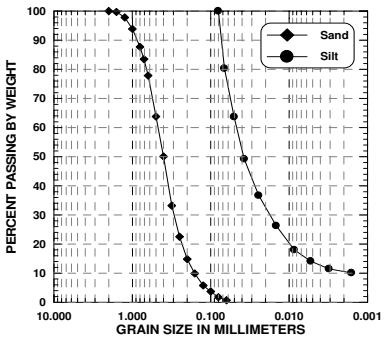


Figure 1
Grain size distribution curves

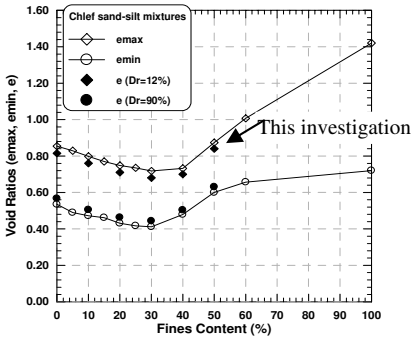


Figure 2
Extreme void ratios of the sand-silt mixtures e_{max} and e_{min} versus fines content

According to some authors, e.g. [24, 25], the behaviour of sand-silt mixture depends on the inter-granular void ratio (e_s):

$$e_s = \frac{(V_v + V_f)}{V_s} \tag{1}$$

where V_v, V_f, and V_s are the volume of voids, fines and sand grains, respectively. Thus $V_v + V_f$ is the volume of inter-granular void space.

When the specific gravity of the silt and sand are very close to each other, the inter-granular void ratio (e_s) can be determined with the global void ratio (e) and the fines content (Fc) using the following expression [4]:

$$e_s = \frac{e + (Fc/100)}{1 - (Fc/100)} \tag{2}$$

Fig. 3 shows the variation of the global and inter-granular void ratios versus fines content for the initial relative densities (Dr = 12% and 90%). As shown in this figure, the global void ratio (e) decreases with the fines content until the value of 30% and then increases; however, the inter-granular void ratio (e_s) increases hyperbolically with the increase of the fines content. This shows that the global void ratio cannot represent the amount of particle contents in silty sands. As the void ratio and proportion of the coarser of fine grains of soil changes, the nature of their microstructures also changes.

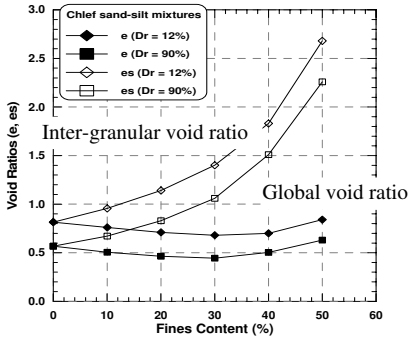


Figure 3
Variation of void ratios with fines content and relative densities

2.2 Experimental Program and Test Procedure

The present experimental study has been conducted to elucidate how the fines fraction affects the undrained residual shear strength and the liquefaction potential of Chlef sand-silt mixtures. For this purpose, a series of undrained monotonic and cyclic triaxial tests were carried out on reconstituted saturated samples of Chlef sand with variation in silt content ranging from 0 to 50%. All specimens were prepared by first estimating the dry weights of sand and silt needed for a desired proportion into the loose and medium dense state (Dr = 12, 50, 60% and 90%) using the undercompaction method of sample preparation which simulates a relatively homogeneous soil condition and is performed by compacted dry soil in layers to a selected percentage of the required dry unit weight of the specimen [27]. The dimensions of the samples were 70 mm in diameter and 70 mm in height in order to avoid the appearance of instability (sliding surfaces) and buckling. After the specimen was formed, the specimen cap was placed and sealed with O-rings, and a partial vacuum of 20 kPa is applied to the specimen to reduce the disturbances. Saturation was performed by purging the dry specimen with carbon dioxide for approximately 30 min. Deaired water was then introduced into the specimen from the bottom drain line. Water was allowed to flow through the specimen until an amount equal to the void volume of the specimen was collected in a beaker through the specimen's upper drain line. A minimum Skempton coefficient-value greater than 0.96 was obtained at back pressure of 100 kPa. All test specimens were isotropically consolidated at a mean effective pressure of 100 kPa, and then subjected to undrained monotonic and cyclic triaxial loading with a constant strain rate of 0.167% per minute (Cell pressure = 600 kPa and Back pressure = 500 kPa).

All the undrained triaxial tests for this study were carried out at a strain rate which was slow enough to allow pore pressure change to equalize throughout the sample with the pore pressure measured at the base of sample. All the tests were continued up to 24% axial strain.

3 Monotonic Tests Results

3.1 Undrained Compression Loading Tests

Figs. 4 and 5 show the results of the undrained monotonic compression triaxial tests carried out for different fines content ranging from 0 to 50% at 100 kPa mean confining pressure within two separate density ranges (Dr = 12, 90%). We notice in general that an increase in the amount of fines leads to a decrease in the deviatoric stress. This decrease results from the role of the fines in reducing the soil dilatancy and amplifying the phase of contractancy of the sand-silt mixtures,

leading to a reduction of the confining effective pressure and consequently to a decrease in the peak strength of the mixtures as is illustrated by Figs. 4a and 5a. The stress path in the (p', q) plane shows clearly the role of the fines in the decrease in the average effective pressure and the maximum deviatoric stress (Figs. 4b and 5b). In this case, the effect of fines on the undrained behaviour of the mixtures is observed for the lower fines contents (0% and 10%), and becomes very marked beyond 20%. These results are in good agreement with the observations of [12, 13]. Table 2 presents the summary of the undrained monotonic compression triaxial tests.

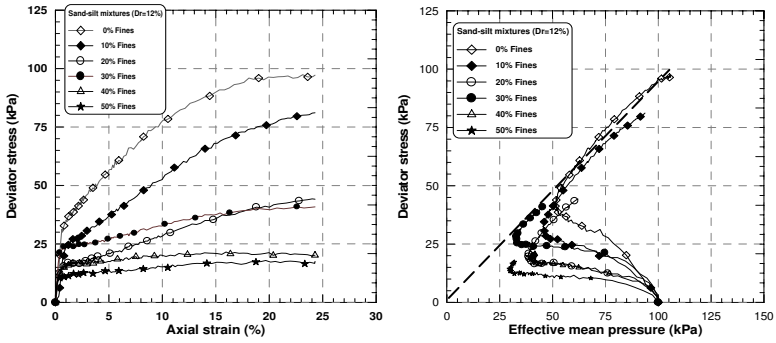
Table 2

Summary of monotonic compression triaxial tests results

Test No	Materials	Fc (%)	Dr (%)	γ_d (g/cm^3)	e	e_s	S_{us} (kPa)
1			12	1.48	0.815	0.815	17.37
2	Clean sand	0	90	1.71	0.567	0.567	20.94
3			12	1.52	0.760	0.956	15.44
4	Silty sand	10	90	1.78	0.505	0.672	19.61
5			12	1.57	0.710	1.140	14.46
6	Silty sand	20	90	1.83	0.463	0.829	18.42
7			12	1.60	0.680	1.400	13.48
8	Silty sand	30	90	1.86	0.443	1.060	17.34
9			12	1.58	0.700	1.830	10.75
10	Silty sand	40	90	1.79	0.503	1.510	15.92
11			12	1.46	0.840	2.680	06.51
12	Silty sand	50	90	1.65	0.630	2.260	12.00

e = global void ratio

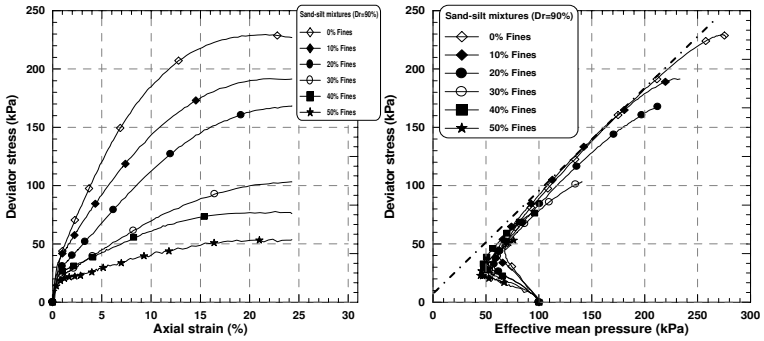
e_s = inter-granular void ratio



(a) (b)

Figure 4

Undrained monotonic response of the sand-silt mixtures (σ_3'=100 kPa, Dr = 12%)



(a) (b)

Figure 5
Undrained monotonic response of the sand-silt mixtures (σ_3'=100 kPa, D_r = 90%)

3.2 Undrained Residual Shear Strength

When loose and medium dense sandy soils are subjected to undrained loading beyond the point of peak strength, the undrained shear strength declines to a near constant value over large deformation. Conventionally, this shear strength is called the undrained steady-state shear strength or residual shear strength. However, if the shear strength increases after passing through a minimum value, the phenomenon is called limited or quasi-liquefaction. Even limited liquefaction may result in significant strains and an associated drop in resistance [9] defined the undrained residual shear strength S_{us} as:

$$S_{us} = (q_s/2)\cos\phi_s = (M/2)\cos\phi_s(p_s') \quad (3)$$

$$M= (6\sin\phi_s)/(3-\sin\phi_s) \quad (4)$$

where q_s, p_s' and ϕ_s indicate the deviator stress (σ_1'- σ_3'), the effective mean principal stress (σ_1'+ 2σ_3')/3 and the mobilized angle of inter-particle friction at the quasi-steady state (QSS) respectively. For the undrained tests conducted at a constant confining pressure and various initial relative densities and fines content, the deviatoric stress (q_s) was estimated at quasi-steady state point along with the mobilized internal friction angle (Fig. 6). Furthermore, the undrained residual shear strength was calculated according to the relation (3).

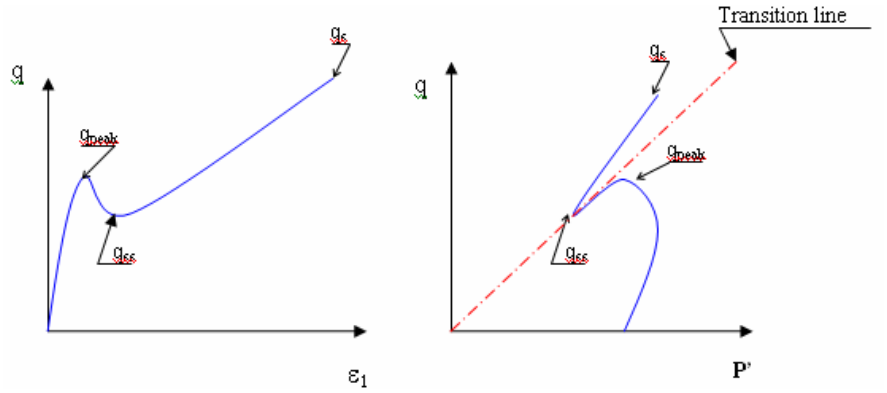


Figure 6
Determination of the phase transition point

Fig. 7 shows the undrained residual shear strength versus the global void ratio and fines content. It is clear from this that the undrained residual shear strength decreases linearly as the global void ratio decreases and the fines content increases for the loose and dense state of the specimen (Dr = 12% and 90%) up to 30% fines content. It means that with a decrease in the global void ratio and an increase in the fines content, the undrained residual shear strength also decreases. In this case we can say that the global void ratio does not represent the real behaviour of silty sand soil in the range of 0-30% fines content. Indeed, the global void ratio appears to be a parameter not as pertinent in sand-fines mixtures as in clean sands for characterizing the mechanical state of these materials. Beyond F_c = 30% the undrained residual shear strength continues to decrease almost linearly with increases in the global void ratio and the fines content for the two relative densities (D_r = 12% and 90%).

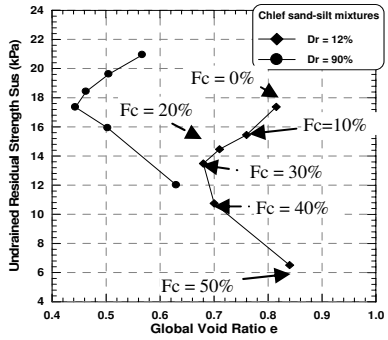


Figure 7
Variation of undrained residual strength with void ratio and relative density (σ_3'=100 kPa)

Fig. 8 illustrates the undrained residual shear strength versus the fines content. Due to obtained results, it can be concluded that we could assume that the undrained residual shear strength of the sand-silt mixtures S_{us} varies linearly and independently of the effective vertical overburden pressure σ_v' for the two relative densities (Dr = 12% and 90%) allowing us to estimate it in the field with no need for in-situ physical parameters measurements in the case of normally consolidated soils (σ_v'= σ_3'). In this laboratory investigation, for the range of 0% to 50% fines content in normally consolidated undrained triaxial compression tests, the following expressions are suggested to evaluate the undrained residual shear strength which is a function of the fines content (Fc):

S_{us}= 17.9552 – 0.1982(Fc) for Dr = 12%

S_{us}= 21.4324 – 0.1625(Fc) for Dr = 90%

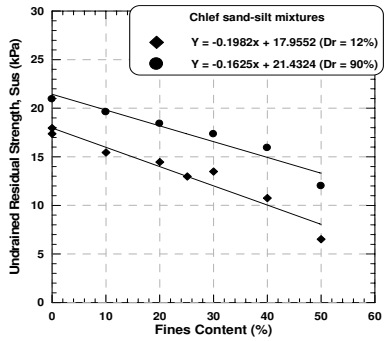


Figure 8
Undrained residual shear strength versus fines content and relative density (σ_3'=100 kPa)

Fig. 9 shows the undrained residual shear strength versus the inter-granular void ratio. It seems that the variation of the undrained residual strength due to the amount of fines is related to the inter-granular void ratio in the range of 0-50% fines content. In this case, the behaviour of silty sand samples is influenced by the contacts of coarser grains, which is quantified by the inter-granular void ratio. By increasing the fines content in the range of 0-50%, the contact between sand grains decreases and therefore the inter-granular void ratio increases and the undrained residual shear strength decreases. Hence, in the range of 0-50% fines content, we could assume that the undrained residual shear strength decreases linearly with the increase of the inter-granular void ratio. In this case, the following expressions are suggested to evaluate the undrained residual shear strength which is a function of the inter-granular void ratio (e_s) for the range of 0% to 50% fines content in normally consolidated undrained triaxial compression tests:

S_{us}= 21.1432 – 5.5378(e_s) for Dr = 12%

S_{us}= 22.9865 – 4.8839(e_s) for Dr = 90%

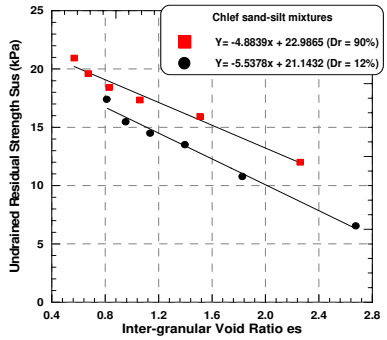


Figure 9

Undrained residual shear strength versus inter-granular void ratio and relative density (σ_3'=100 kPa)

Fig. 10 shows the variation of residual shear strength (S_{us}) with relative density (Dr) at various fines contents. It is clear from this Figure that an increase in the relative density results in an increase in the undrained residual shear strength at a given fines content. [3, 28] report similar behaviour of increasing residual shear strength with increasing relative density. The present laboratory study focuses on the effect of the fines content and the inter-granular void ratio on the undrained residual shear strength of silty sands at various relative densities (Dr = 12% and 90%). It can be noticed from the results of this study that there is a significant decrease in the undrained residual shear strength with increase in the fines content or the inter-granular void ratio, but there is a significant increase in the undrained residual shear strength with an increase in the relative density and a significant decrease of the undrained residual strength with an increase in the fines content. The aspect of the present study is in good agreement with the experimental work reported by [9] on Tia Juana silty sand, [29] on silty sands retrieved from the Lower San Fernando Dam, and by [8] on Adebil sand with different fractions of fines content.

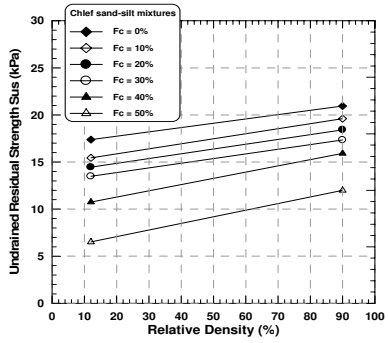


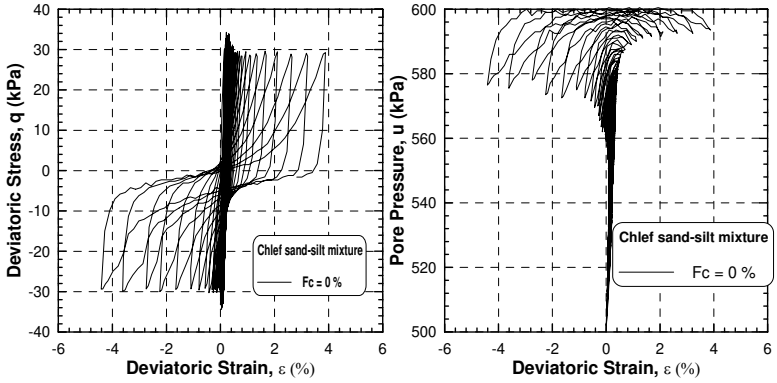
Figure 10

Undrained residual shear strength versus relative density at various fines contents (σ_3'=100 kPa)

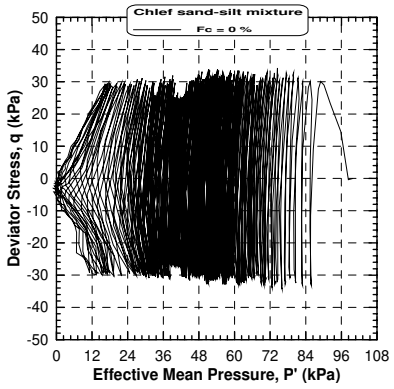
4 Cyclic Tests Results

4.1 Undrained Loading Tests

Three series of stress-controlled cyclic triaxial tests were carried out on isotropically consolidated soil specimens with different fines content ranging from 0 to 40% and alternated symmetric deviator stress under undrained conditions simulating essentially undrained field conditions during earthquakes in order to produce liquefaction potential curves of the sand-silt mixtures. All along the test program, a frequency of 0.5 Hz was used. The first one included three alternated cyclic tests and was realized on clean sand samples (Fc = 0%) with a relative density of 50% and a confining initial pressure of 100 kPa. The loading amplitudes of the cycles (q_m) used are respectively 30, 50 and 70 kPa. The tests of the second one were realized on the mixture sand-silt samples with a fines content of 10% and loading amplitudes of 30, 40 and 60 kPa; while the third series of tests concerned samples with a fines content of 40% and loading amplitudes of 20, 30 and 50 kPa. It is noted that the presence of fines affects considerably the liquefaction of the samples. Fig. 11 illustrates the results of the test carried out on clean sand samples (Fc = 0%) with loading amplitude of 30 kPa. It is clear from the figure that increases in the pore water pressure during the cycles result in a reduction of the average effective pressure. The rate of increase in the pore pressure remains low, because liquefaction is obtained only after 158 cycles (Fig. 11); for the test with qm = 30 kPa and a fines content of 10%, we notice an important increase in the pore water pressure during the 27th cycle with a significant development of the axial strain (2.5%) leading to the liquefaction of the sand-silt mixture sample (Fig. 12).



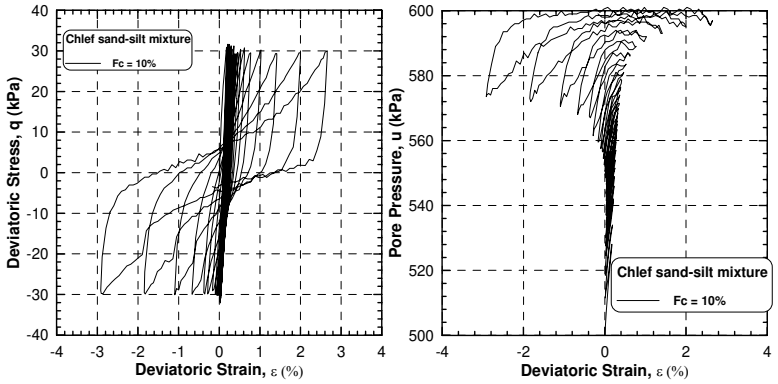
(a) (b)



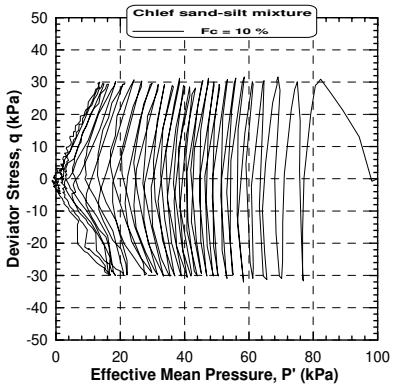
(c)

Figure 11

Undrained cyclic response of clean sand (Fc=0%, q_m=30 kPa, Dr=50%, σ_3'=100 kPa)



(a) (b)



(c)

Figure 12

Undrained cyclic response of sand-silt mixture (Fc =10%, qm =30 kPa, Dr =50%, σ_3'=100 kPa)

While for the test with a loading amplitude of 30 kPa and a fines content of 40%, we notice an important increase in the pore water pressure during the 3rd cycle with a significant development of the axial strain leading to the liquefaction of the sample at the 4th cycle (Fig. 13). This shows that the increase in the amount of fines in the range of (0-40%) leads to an increase in the risk of liquefaction.

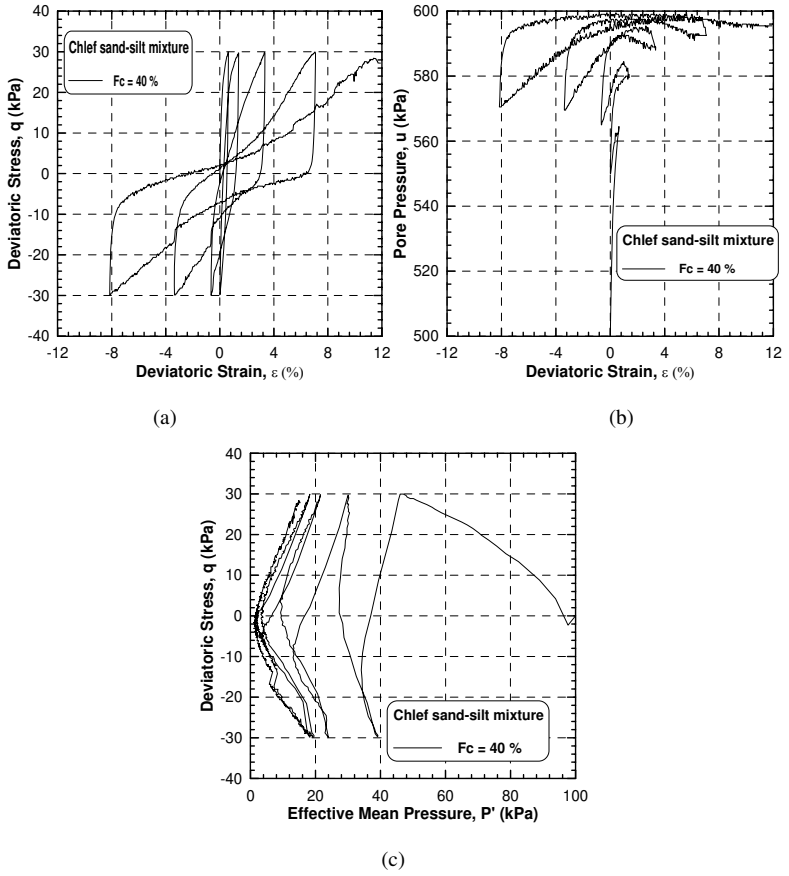


Figure 13
Undrained cyclic test on Chlef sand-silt mixture (Fc = 40%, qm =30 kPa, Dr =50%, σ_3'=100 kPa)

4.2 Effect of Fines on the Liquefaction Potential

Figs. 14a and 14b illustrate the variation of the cyclic stress ratio (CSR = q_m / $2\sigma_3$') and the cyclic liquefaction resistance (CLR) with the number of cycles (N_c) and fines content respectively. The cyclic liquefaction resistance (CLR) is defined as the cyclic stress ratio leading to liquefaction for 15 cycles according to [9]. The test results show that the liquefaction resistance of the sand-silt mixture decreases with increase in the fines content. These results confirm those found during

monotonic tests showing that the increase in the fines content reduces the dilatancy phase and amplifies the phase of contractancy. Consequently, the increase of the contractancy phase induced a reduction in the liquefaction potential with the increase in the fines content. For the mixtures of Chlef sand-silt, the presence of fine-grained soils increases the phase of contractancy resulting in a significant decrease in the liquefaction potential particularly, for the fines contents less than 10%. It should be noted that for the studied amplitude (qm= 30 kPa), the increase in fines content accelerates the liquefaction processes of the Chlef sand-silt mixtures. Fig. 14c presents the loading cycles till the liquefaction versus the fines content. We notice that the liquefaction resistance decreases with the increase in the fines content and loading amplitudes. The soil samples sheared with higher level loading (CSR = 0.25) are more vulnerable to liquefaction than those sheared with smaller loading level (CSR = 0.15).

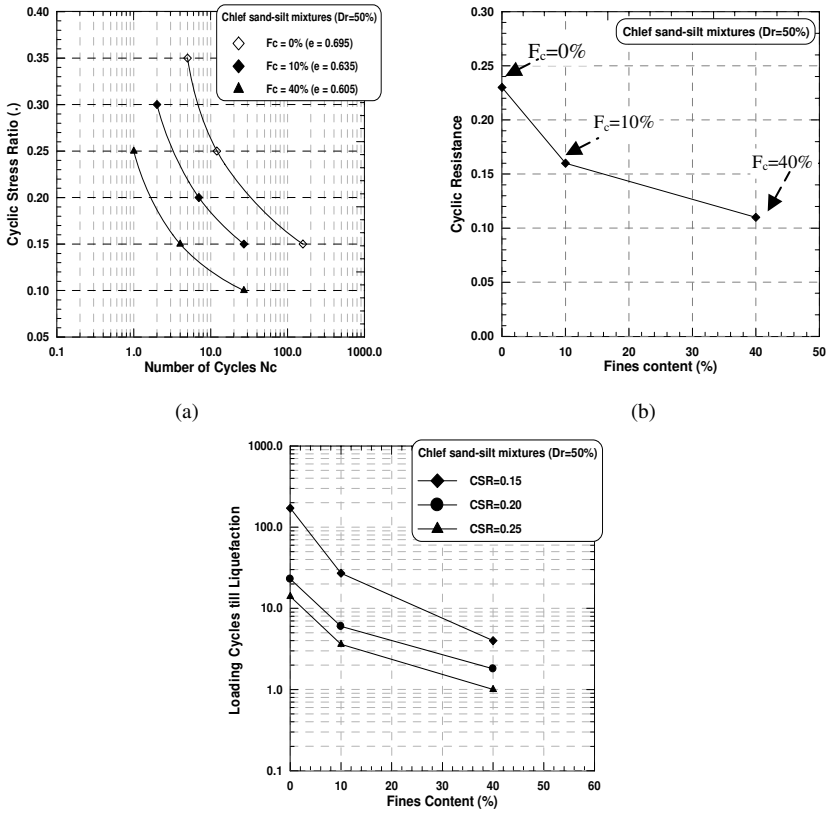


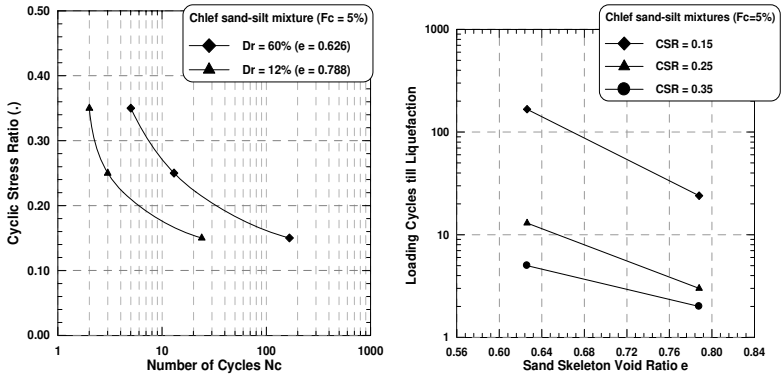
Figure 14

Effect of fines on the liquefaction potential of the Chlef sand-silt mixtures (σ_3'=100 kPa, Dr=50%)

4.3 Effect of the Relative Density

Undrained cyclic tests were performed on Chlef sand–silt mixtures (Fc = 5%) for two relative densities (Dr = 12% and 60%). For each density, we varied the loading amplitude in order to draw the liquefaction potential curve. The tests were carried out for the amplitudes qm = 30, 50 and 70 kPa. We note that the liquefaction was reached quickly for the higher amplitudes: after 2 and 3 cycles respectively for the loading amplitudes of q_m = 70 and 50 kPa, whereas the liquefaction under the loading amplitude qm = 30 kPa for the tests carried out with a relative density Dr = 12% was reached at 24 cycles. For the tests carried out with a relative density of Dr = 60% and for the same loading amplitudes, the liquefaction was reached at 5, 13 and 167 cycles.

Fig. 15 summarizes the results of all these tests. Fig. 15a illustrates the influence of the relative density on the liquefaction potential of Chlef sand-silt mixture (Fc = 5%). It shows clearly that the increase in the relative density leads to an increase in the liquefaction resistance of Chlef sand-silt mixture. Fig. 15b shows the influence of the global void ratio on the liquefaction resistance of the Chlef sand-silt mixtures. This figure shows clearly that the liquefaction resistance decreases with the increase in the global void ratio and the loading amplitude. Note that the reduction in the liquefaction resistance of Chlef sand-silt mixture becomes very marked for the smaller cyclic stress ratios CSR = 0.15 and 0.25.



(a) (b)

Figure 15
Effect of relative density on the liquefaction potential of Chlef sand-silt mixture
(σ_3'=100 kPa, Fc =5%)

Conclusion

A series of undrained monotonic and cyclic triaxial tests were carried out on silty sand collected from liquefied sites at Chlef River, Algeria. The effect of variation in the fines content was studied. In the light of the experimental evidence, the following conclusions can be drawn:

Undrained monotonic tests performed with initial relative densities of 12% and 90% showed contractive behaviour of the Chlef sand-silt mixtures samples at the initial confining pressure in the global void ratio range tested. The undrained residual shear strength decreases as the global void ratio decreases and the fines content increases up to 30% fines content. Beyond that, it decreases with increases in the global void ratio and the fines content. The undrained residual shear strength decreases linearly with an increase in the fines content and the inter-granular void ratio. The peak and residual strengths of the soil are sensitive to the presence of fine grained soils. As the fines content increases, the peak and the residual strengths decrease. The strength of silty sand up to 50% fines content is less than that of the clean sand. This means that the soil becomes weakened with an increase in the fines content up to 50%.

Undrained cyclic tests show that the increase in fines content reduces the liquefaction potential of the sand-silt mixtures and the samples sheared with higher level loading (CSR= 0.25) are more vulnerable to liquefaction than those sheared with smaller loading level (CSR = 0.15). They also show that the liquefaction resistance increases with the relative density but it decreases with the increase in the global void ratio and the loading amplitude. Note that the reduction in the liquefaction resistance of Chlef sand-silt mixture becomes very marked for the smaller cyclic stress ratios CSR = 0.15 and 0.25.

The results of cyclic tests confirm those found during monotonic tests showing that the increase in the fines fraction reduces the soil dilatancy and amplifies the phase of contractancy inducing an important decrease in the liquefaction potential of Chlef sand-silt mixtures. Note that the increase in the fines content accelerates the liquefaction phenomenon for the studied amplitude (qm = 30 kPa).

References

[1] S. Zlatovic, K. Ishihara, “On the Influence of Non-Plastic Fines on Residual Strength” Proceedings of the first International Conference on Earthquake Geotechnical Engineering, 1995, Tokyo, pp. 14-16

[2] P. V. Lade, J. A. Yamamuro, “Effects of Non-Plastic Fines on Static Liquefaction of Sands”, Canadian Geotechnical Journal, 1997, Vol. 34, pp. 918-928

[3] S. Thevanayagam, K. Ravishankar, S. Mohan, “Effects of Fines on Monotonic Undrained Shear Strength of Sandy Soils”, ASTM Geotechnical Testing Journal, 1997, Vol. 20, No. 1, pp. 394-406

[4] S. Thevanayagam, “Effect of Fines and Confining Stress on Undrained Shear Strength of Silty Sands”, J. Geotech. Geoenviron. Eng. Div., ASCE 1998, 124, No. 6, pp. 479-491

[5] J. A. Yamamuro, P. V. Lade, “Steady-State Concepts and Static Liquefaction of Silty Sands”, Journal of Geotechnical and Geoenvironmental Engineering, ASCE, 1998, Vol. 124, No. 9, pp. 868-877

[6] F. Amini, G. Z. Qi, "Liquefaction Testing of Stratified Silty Sands", Journal of Geotechnical Geoenvironmental Engineering, Proc. ASCE, 2000, Vol. 126, No. 3, pp. 208-217

[7] S. A. Naeini, "The Influence of Silt Presence and Sample Preparation on Liquefaction Potential of Silty Sands", PhD Dissertation, Tehran, Iran: Iran University of Science and Technology, 2001

[8] S. A. Naeini, M. H. Baziar, "Effect of Fines Content on Steady-State Strength of Mixed and Layered Samples of a sand", Soil Dynamics and Earthquake Engineering, 2004, Vol. 24, pp. 181-187

[9] K. Ishihara, "Liquefaction and Flow Failure during Earthquakes", Geotechnique, 1993, Vol. 43, No. 3, pp. 351-415

[10] W. D. L. Finn, "State-of-the-art of Geotechnical Earthquake Engineering Practice", Soil Dynamics and Earthquake Engineering, 2000, Vol. 20, pp. 1-15

[11] S. Singh, "Liquefaction Characteristics of Silt", Ground Failures under Seismic Condition Geotechnical Special Publication, 1994, No. 44, ASCE, pp. 105-116

[12] C. K. Shen, J. L.Vrymoed, C. K. Uyeno, "The Effects of Fines on Liquefaction of Sands", Proc. 9th Int. Conf. Soil Mech. and Found. Eng, Tokyo, 1977, Vol. 2, pp. 381-385

[13] J. H. Troncoso, R. Verdugo, "Silt Content and Dynamic Behaviour of Tailing Sands" Proc. 12th Int. Conf. on Soil Mech. and Found. Eng., San Francisco, 1985, pp. 1311-1314

[14] V. N. Georginnou, J. B. Burland, D. W. Hight, "The Undrained Behaviour of Clayey Sands in Triaxial Compression and Extension", 1990, Geotechnique 40, No. 3, pp. 431-449

[15] W. L. Finn, R. H. Ledbetter, G. Wu, "Liquefaction on Silty Soils: Design and Analysis", Ground Failures under Seismic Condition Geotechnical Special Publication, 1994, No. 44, ASCE, pp. 51-76

[16] V. P. Vaid, "Liquefaction of Silty Soils", Ground Failures under Seismic Condition Geotechnical Special Publication, 1994, No. 44, ASCE, pp. 1-16

[17] D. Erten, M. H. Maher, "Cyclic Undrained Behaviour of Silty Sand", Soil Dynamics and Earthquake Engineering, 1995, Vol. 14, pp. 115-123

[18] S. Zlatovic, K. Ishihara, "Normalised Behaviour of Very Loose Non-Plastic Soils: Effects of Fabric", Soils and Foundations, 1997, Vol. 37, No. 4, pp. 47-56

[19] K. T. Law, Y. H. Ling, "Liquefaction of Granular Soils with Non-Cohesive and Cohesive Fines", Proc. of the 10th World Conference on Earthquake Engineering, Rotterdam, 1992, pp. 1491-1496

[20] J. P. Koester, "The Influence of Fine Type and Content on Cyclic Strength", Ground Failures under Seismic Condition. Geotechnical Special Publication, 1994, No. 44, ASCE, pp. 17-33

[21] V. C. Xenaki, G. A. Athanasopoulos, "Liquefaction Resistance of Sand-Silt Mixtures: an Experimental Investigation of the Effect of Fines", Soil Dyn. Earthq. Eng., 2003, No. 23, pp. 183-194

[22] R. Bouferra, I. Shahrour, "Influence of Fines on the Resistance to Liquefaction of a Clayey Sand", Ground Improvement 8, 2004, No. 1, pp. 1-5

[23] K. Ishihara, J. Troncosco, Y. Kawase, Y. Takahashi, "Cyclic Strength Characteristics of Tailing Materials", Soils and Foundations, 1980, Vol. 23, No. 8, pp. 11-26

[24] T. C. Kenny, "Residual Strengths of Mineral Mixtures", Proc. 9th Int. Conf. Soil Mech. And Found. Eng. Tokyo, 1977, Vol. 1, pp. 155-160

[25] J. K. Mitchell, "Fundamental of Soil Behaviour", John Wiley Interscience, New York; 2nd ed, 1993

[26] J. A. Yamamuro, F. M. Wood, "Effect of Depositional Method on the Undrained Behaviour and Microstructure of Sand with Silt", Soil Dyn. Earthq. Eng., 2004, No. 24, pp. 751-760

[27] R. S. Ladd, "Preparing Test Specimen Using under Compaction", Geotechnical Testing Journal, GTJODJ, 1978, Vol. 1, pp. 16-23

[28] T. G. Sitharam, L. Govinda Raju, B. R. Srinivasa Murthy, "Cyclic and Monotonic Undrained Shear Response of Silty Sand from Bhuj Region in India", ISET Journal of Earthquake Technology, June-December 2004, Paper No. 450, Vol. 41, No. 2-4, pp. 249-260

[29] M. H. Baziar, R. Dobry, "Residual Strength and Large-Deformation Potential of Loose Silty Sands", Journal of Geotechnical Engineering, ASCE, 1995, Vol. 121, pp. 896-906

Improving Baggage Tracking, Security and Customer Services with RFID in the Airline Industry

Deepti Mishra, Alok Mishra

Department of Computer Engineering
Atilim University, Ankara, Turkey
deepti@atilim.edu.tr, alok@atilim.edu.tr

Abstract: Radio frequency identification (RFID) has been identified as one of the ten greatest contributory technologies of the 21st Century. This technology has found a rapidly growing market, and an increasing variety of enterprises are employing RFID to improve the efficiency of their operations and to gain competitive advantage. In the aviation industry, major airports/airlines have been looking for the opportunity to adopt RFID in the area of baggage handling for a long time. Many pilot tests have been done at numerous U.S., European, and Hong Kong airports. RFID tags were found to be far more accurate than bar codes, and their performance was also measured to be well above that of bar codes. This paper presents the state of RFID adoption planning, architecture and implementation at a major airline, with a special focus on improved services due to improved baggage handling, on increased airport/airline security and on frequent flier program services. This is accomplished by integrating RFID technology together with networking and database technologies.

Keywords: Radio Frequency Identification; RFID; baggage handling/tracking; airline industry; customer-service

1 Introduction

Developments in logistics have been changing the world faster than ever [1]. Radio frequency identification (RFID) is an emerging technology that is increasingly being used in business and industry. RFID systems have three main components as shown in Figure 1.

- The tag: RFID tags are chips embedded in items which store and transmit information about these items. Most RFID tags store data that identifies a specific item.

- The reader: RFID readers are radio frequency transmitters and receivers that communicates with the tags. Readers, using an attached antenna, receive data from the tag and then pass it to a computer system for processing.
- The computer system: The computer system receives the data from the RFID reader through a cable or wireless connection for storage, interpretation and action.

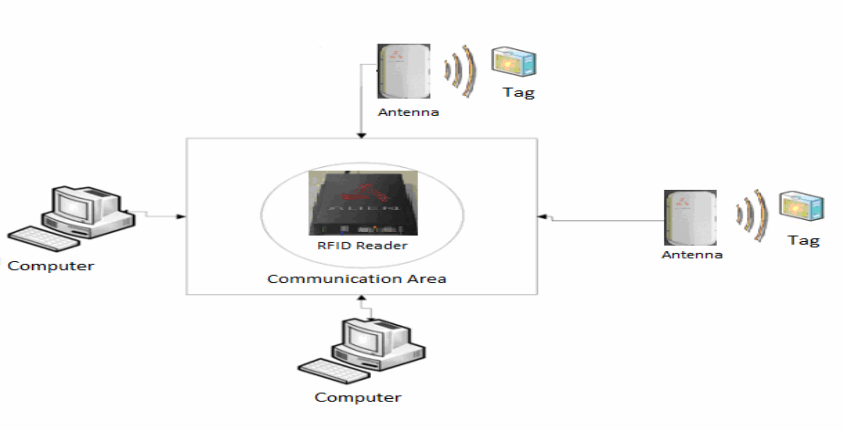


Figure 1
A Simplified View of an RFID system

Owing to its "MOST" (mobility, organizational, systems and technologies) characteristics, RFID has received considerable attention and is considered to be the next wave of the IT revolution [2]. RFID can allow any tagged item to be mobile and intelligent and to communicate with an organization's overall information infrastructure [3]. The main applications of RFID are in aviation, retailing, the food industry, hospitals, libraries, animal detection, building management, waste management, museums, etc.

The global business environment is changing very quickly [4]. The new, global economy is the economy of knowledge and ideas, where innovative ideas and technologies fully integrated in services and products became a key to a generation of new working positions and higher living standards [4].

The aviation industry is just one of many industries that could benefit from RFID technology. In the travel industry, the pressure to provide better customer service has never been greater. Yet the pressure to reduce operating costs is equally strong. This is an industry which operates on average profit margins of less than 4%. RFID is a technology that could help revitalise the airline industry and could be the impetus for the change it needs [5].

"In 2005, the industry lost in the region of $2.5 billion on mishandled baggage, when you take into account the costs involved in reuniting the delayed baggage with its owner, which, happily, is the case over 99% of the time. This year we will

reach the two billion passenger landmark, which on current trends will translate into 30 million pieces of mishandled baggage" [6]. The IATA surveyed airlines on their understanding of the reasons for and proportions of bags being mishandled. Among the main reasons, the airlines identified two areas where RFID can fix the issues [7]:

- Barcode reading problems cause 9.7% of all mishandled baggage and
- failures to receive a baggage status message contribute to a further 11% of mishandled baggage.

Many airlines have run RFID trials over the past few years to prove the efficacy of the systems employed in the air transport environment. RFID benefits over bar code tags, as shown in Table 1, are:

- High reading speed: a considerably higher baggage identification rate than with barcode readers for minimized manual intervention and high savings in terms of time and costs.
- Higher accuracy in reading: Tests have shown first-read rates of over 99% with RFID tags compared to less than 90% for bar code-only tags.
- The writing option opens up the potential for new applications: the data stored on the bag tags can be updated at any time for additional security. Example: inline screening results can be written onto the bag tag.
- Flexible bag-tag choice: the system can read and write on all IATA-specified bag tags from a large range of different manufacturers.
- Efficient baggage reconciliation processes (flight makeup) through simple bag tag reading.
- Robust and easy to integrate: the industrial-design system is based on standard components proven in day-to-day use all over the world.
- Cost Savings [8]:
  - US$760 million per year in industry savings when fully implemented (based on US$ 0.10 per tag cost).
  - Out of 2 billion plus pieces of luggage handled per year, just over 1% are mishandled.
  - Each baggage mishandling costs on average US$90.

Table 1
Comparison of Barcode and RFID Tags (Source: IATA [7])

ATTRIBUTE	BARCODE TAGS	RFID TAGS
FLEXIBILITY Line of sight reading	Required	Not required
ABILITY Number of simultaneous scan	One	Multiple, and can distinguish bags from other items
ACCURACY	Read rate highly variable	Fully automated & accurate Read rate > 99%
DURABILITY	Can be damaged easily	More durable, withstand handling
DATA SUPPORT	No write capability	Possible to update data
MAINTENANCE	High Maintenance	Low Maintenance
COST	Cheap tags but expensive readers	Expensive tags but cheap readers

Both academics and practitioners are keenly aware of how organizations can extract business value from RFID [9], [3]. The main question here is how RFID can be used to reduce costs or whether the investment required in the RFID project can be returned. The cost of investment is apparently the primary reason the airline industry has not yet implemented RFID baggage tagging. Also, such tags end up being disposable and are currently not cheap enough to use, as they cut into already small profits. However, weighed against the fact that airlines lose track of about 100,000 bags each day and must therefore compensate passengers, baggage tagging costs begin to look more attractive for RFID implementation. The prediction is that, as more airlines start to implement radio frequency technology for baggage tagging, the cost of tags and readers will drop, which in turn will likely encourage more airlines to follow suit.

Other issues that need to be worked out, beyond the tag costs, are the infrastructure and the tags themselves. It is unclear who will pay for installing RFID systems because the responsibility for baggage handling varies around the world. Experts recommend that it will be more beneficial if the airports rather than the individual airlines adopt the system [10]. This case study reports that RFID implementation will ensure the effective management of baggage tracking/delivery and will provide customized and personalized services to premium customers. There have been reports of airlines and airports abandoning RFID system implementations and pilot tests due to the lack of demonstrable return on investment, but here tangible and intangible benefits such as efficient baggage handling and improved services to premium customers will outweigh the costs. To the best of our knowledge, this case is the first one which reports improved services to premium customers as an added feature of RFID implementation in the aviation industry.

The paper is structured as follows: Section 2 addresses the literature review of the topic. Section 3 presents the real-life RFID adoption planning and implementation as a case study. Finally, Section 4 concludes with discussions.

2 Literature Review

In the aviation industry, major airports have been looking for opportunities in the baggage handling area since 1999 [11]. Many pilot tests of RFID have been done at numerous U.S. and European airports [11]. In U.S. tests, RFID tags were far more accurate than bar codes when applied to baggage handling operations [11]. Nath et al. [12] advocate embedding RFID tags in luggage labels, as it could eliminate the need for manual inspection and routing by baggage handlers. A network of readers placed along conveyor belts could read the tags' routing information and provide feedback to a system that could then direct the bags onto the correct path [12]. Automatic routing could reduce the number of misrouted bags, lowering costs and improving customer satisfaction [12]. Al-Ali et al. [13] described the design and implementation of a prototype system for baggage handling in airports to enhance the management and tracking of passengers' luggage while, as a side effect, improving airport security. Wyld et al. [14] showed by specifically focusing on Delta Airlines how RFID technology can improve customer service through better operational efficiency in baggage handling, which has been demonstrated to be an integral component of an airline's customer service rating.

Even though the value of RFID-enabled technologies in handling passenger bags is generally accepted in the industry, the adoption of these technologies is hindered by concerns relating to their inadequate return on investment [15]. The reason most of the projects failed to demonstrate the needed financial return was because they focused primarily on increased labour productivity in the baggage scanning process, instead of considering other more valuable applications of RFID-enabled technologies, such as the savings in time, money and effort from the avoidance of costly baggage handling exceptions [15]. Viswanadham et al. [15] attempted to address this issue by highlighting scenarios in the baggage handling process where RFID-enabled technologies may be uniquely positioned to create value.

Sample et al. [16] focused on the use of RFID technology in the US department of Transportation's (DOT) international airport security initiative in Nigeria. One of the uses of RFID baggage tags, in conjunction with RF handheld readers and boarding pass readers, is to verify passenger boarding versus luggage loading for positive passenger baggage matching on flights departing for the U.S. and other international locations [16].

The need for matching passengers with checked luggage has been at the forefront of anti-terrorism concerns ever since the in-flights bombings in the 1980s that took down a Pan American 747 over Lockerbie, Scotland and an Air India jumbo jet over the Atlantic [17]. Out of this concern, airlines must routinely remove bags from the aircraft when a passenger fails to board [18]. Often, this is a time-intensive, laborious process, which can delay flight departures indefinitely, costing the airline countless amounts of goodwill amongst the passengers, even if such measures are done precisely to safeguard their transit and their very lives [17]. A bag with an RFID tag could be pinpointed accurately in a unit loading device or individual aircraft hold, greatly facilitating offloading if necessary.

Wong et al. [19] discussed workers' safety concern due to radiation emissions from a recently installed 900 MHz RFID baggage handling system at Hong Kong International Airport. They concluded that the operation of the RFID system is considered a safe system, as the E-field levels recorded for the whole system is well below the ICNIRP (International Commission on Non-Ionizing Radiation Protection) restricted level.

RFID technology provides enormous economic benefits for both business and consumers, while simultaneously potentially constituting one of the most invasive surveillance technologies threatening consumer privacy [20]. Kelly and Erickson [20] recommended that the use of RFID technology should be legally regulated. Manufacturers and retailers should notify purchasers that RFID tags are being used, and these tags should be automatically disabled at checkout. Also, government authorities should be required to obtain a court order in order to be able to access RFID tags [20]. Juels et al. [21] proposed the use of selective blocking by blocker tags as a way of protecting consumers from unwanted scanning of RFID tags attached to items they may be carrying or wearing.

Generally, the case study method is a preferred strategy when "how" and "why" questions are being posed, and when the researcher has little control over events [22]. The case study method, a qualitative and descriptive research method, looks intensely at an individual participant or a small group of participants, drawing conclusions only about the participants or group and only in a specific context [22]. The case study method is an ideal methodology when a holistic, in-depth investigation is required [23]. Case studies are often conducted in order to gain a rich understanding of a phenomenon and, in information systems research, the intensive nature, the richness of a case study description and the complexity of the phenomenon are frequently stressed in case study reports [24].

3 Case Study

This is a case-study of an airline which is a member of Star Alliance group. The airline needs to maintain high levels of flexibility to face new challenges from competitors around the world, and to identify and create new services to improve customer satisfaction and reduce costs.

The increase in passenger and baggage volumes, plus the development of global alliances and dual transfer flights, all create big challenges for airlines and airports. This is especially true for an airline handling over 2 billion passengers per year. This puts an extra load on the existing baggage handling system, which relies on an aging Barcode system. The airline, therefore, requires a highly efficient method to handle the increasing passengers and baggage volumes, and RFID technology has drawn the attention of the airline.

The strategic importance of RFID applications cannot be underestimated and the advancement of RFID creates opportunities for new and innovative services provided through the RFID infrastructure [2]. RFID is expected to drastically impact the organization's strategic management [3]. One of the key priorities for the airline is to offer excellent services to its clientele and, in addition, the airline company appears to target mainly the premium segment of the market. The major advantage of targeting the top end segment of the market is that it gives the airlines the opportunity to maximize revenue and profit generation, a key factor in the highly competitive and not very profitable airline industry. Recently the company started an RFID project with two main objectives:

- To ensure better services, especially in terms of service delivery. Improvement in baggage tracking and baggage delivery has been identified as key business driver.
- To initiate a new Frequent Flyer Program (FFP) experimental project for premium members based on RFID technology, with the main aim of providing customized and personalized services.

By focusing a firm's RFID strategy on customer-facing activities, a firm can use the technology to change its basis of competition from an efficiency oriented strategy to one where RFID has more strategic implications, such as in providing the foundation for new products or services, or by providing the infrastructure to enhance customers' value perceptions in order to strengthen customer loyalty [25].

The additional functionality of RFID allows information to be changed at different points in the airline system. This makes it possible to hold bags for security checking and release them for loading when checked, provided the RFID system is linked to the baggage reconciliation systems. Similarly, RFID will be used to track passenger progress through airports, reducing the number of passengers arriving late at the gate, and in doing so ensuring that planes leave on time.

Business process reengineering (BPR) is described as a fundamental rethinking and radical redesign of business processes in order to achieve dramatic improvements in critical, contemporary measures of performance, such as cost, quality, service and speed [26]. Research in electronic data interchange (EDI) usage has shown that integrating EDI with process redesign results in more benefits than EDI alone [27], [28]. RFID will provide the foundation for a thorough rethinking of vital business processes [29]. RFID will require considerable process redesign at all stages in the value chain where the technology is applied [30]. Some key systems that need to be changed to accommodate RFID functionality:

- check-in systems,
- belt and tray conveyors,
- sorting systems,
- baggage screening,
- baggage reclaim,
- IT control and instrumentation technology.

Three official surveys were written and published before the design process for three different classes of target groups: passengers, managers and stakeholders. Additionally interviews were conducted with these target groups. The following issues and requirements were found to be necessary from user's perspective:

- Security / Privacy policy,
- Scalability, Reliability, Availability,
- System performance and speed,
- Mobility of hardware parts,
- Initial Price and total ownership cost,
- Self managed system, Automatic failover,
- Using contactless cards which work remotely,
- ISO-18000 compliant.

Keeping the above facts in mind, the design of the system was made. The company chose the system which follows the 'EPC global UHF Gen 2' standard protocol, the most reliable and accepted standard in implementing RFID projects. The specifications of the system are explained in three sections: Architecture, Hardware, and Software. Users were indirectly involved in the system architecture design process, from surveys and initial interviews, as well as directly via interviews and meetings throughout the design process. However, due to the nature of other parts like hardware and software specifications, users were not directly involved in these parts. However, the selected technologies were chosen based on the requirements mentioned by users as well as on compatibility with the other parts of airlines IT infrastructure and on cost implications.

3.1 System Architecture

The system is made up of a RFID passive card which transmits radio frequency data collected by an RFID wired reader, which in turns transmits data to a filtering system. This filtering system filters data based on business rules and transmits that to the appropriate application, which is connected to a host of database, via the network. The database processes the data and sends it back to the appropriate application. This process works in the same way for baggage tags as well as for frequent flyer tags. While it informs airport staff of the presence of a premium member via the frequent flyer tags, it also ensures that the bag is directed to the correct belt by the baggage tag.

The main critical point of the system is the sophisticated filtering algorithm which improves the performance of the system and makes it highly scalable as well. In other words, the system works with paralleling ability, and if the number of bags or passengers increases, there will only be a need for adding another filtering system, without any drastic change in the system architecture. While the tags on the baggage will be for a single use only, the cards for the members should theoretically be long lasting.

Just to focus on how the RFID chips for Frequent flyer members are programmed, the system will not be confused by different RFID cards because the system will respond based on the data written in a card which is the username and user id (which is the frequent flyer number) and response will be based on a set of clearly laid out business rules.

3.2 Hardware Specification

In this phase, surveys were used for making decisions regarding the special consideration as to “Price / Total cost of ownership”, and being ISO-18000 compliant.

RFID tags: The tags are based on the IATA standard of RP1740c and use 850 MHz to 950 MHz frequency, also known as UHF, which is licensed by different countries at different bands and powers. The air interface protocol employed is ISO-18000-6-C, which is an open standard that defines the way in which the reader talks to the tag and the way the tag responds.

RFID cards: The cards are the same size of Visa cards. They follow ISO-7813 with “Tag Model: 116501 GAO”, which follows the ISO-18000 6-C (the IATA accepted standard). It is very light in weight and is contactless, with the ability of 5-500 tag reads per second. For the non-technical design of the card, such as colour and images, customer opinions were widely used. The tag is passive, thus it follows the security and privacy policy regulation.

The RFID reader: Fixed readers were selected due to the fact that they are used mainly for monitoring. The reader chip is ‘Intel® UHF RFID Transceiver R1000,

which meets the EPCglobal Gen 2 and ISO 18000-6C specifications (the IATA mandatory standard) and is appropriate for the purpose of the project (Intel® UHF RFID Transceiver R1000 is Intel's newest RFID reader, which has a reasonable price and can read UHF RFID tags). Readers come in a wide range of sizes, offer different features and start at $500. They can be affixed in a stationary position, integrated into a mobile computer that is used for scanning bar codes, or even embedded in electronic equipment, such as label printers [31].

The Network: It is a high speed wired fibre optic LAN. The reasons for choosing this are network area, resistance to noise, high security and resistance to security breaches, high performance and very high speed. Though this technology is expensive, the advantages outweigh the costs. Due to security concerns, all parts of the network have been wired and all the transmitted messages are encrypted. The network is segmented and separated with firewalls, which makes the system robust against security breaches.

Servers: The servers are comprised of Intel Xeon 2.8 GHz processors and Intel Xeon 3.4 GHz processors with minimum 2 GB RAM and required hard disk volume based on the volume of data. These servers have been used for hosting database, application server and load balancing purposes.

3.3 Software Specification

Oracle application servers are used along with an Oracle sensor edge Server as the middleware between the Oracle Database and the application. The main reason for using this technology is that the product has the proven ability to filter the data from sensor based technologies such as RFID, which collect unstructured data from the environment. One of the main factors for selecting Oracle products is that the airline is already using many Oracle products such as the Oracle ERP system e-business suite; therefore the new system is compatible with the existing ones. The operating system is Oracle Enterprise Linux 4 which is free, open source and highly compatible with Oracle products.

3.4 Implementation

Based on the nature of the project and company requirements, the project will be implemented in phases based on the functionalities and logistics issues as shown in Table 2.

Table 2
Outline of Implementation

Phase	Baggage Tags	Passengers Tags
One	Will be implemented between the airport where the airline is based and key airports which already have RFID capability.	Will be implemented for premium passengers using the Premium Terminal at the airport where the airline is based, to provide more personalized service in the lounge.
Two	Will be implemented between airport where the airline is based and high value/volume destinations.	Will be used for the Decision Support System (DSS) for analysing customer behaviours, travel trends and the booking /travelling habits of customers for better Customer Relationship Management (CRM).
Three	Will be implemented for all destinations in the airline's network.	Will be implemented for all the airline's frequent flyer members at all stations in the network.

3.5 Testing

Testing has been done at all three levels:

- Unit Testing: The system has been tested to ensure that it is error-free. This test has been done by the system creator and his/her related team. In this stage, each part of the system is tested separately; for example, RFID readers have been tested to ensure that they work properly in a real environment. The other parts of the system have been tested by following the same approach.
- System Testing: In this stage, the system has been tested as an integrated unit (testing the system as a whole). The main concept of this test is to check whether all parts of the system are compatible and work in harmony with each other. This test has been performed by a special team of system inventors. Penetration test has been done in this stage to check the system scalability, performance and the maximum workload at which the system can operate without problems.
- Acceptance Testing: This has been final test and as suggested by its name, stakeholders, managers and passengers play main roles in this stage. The main question which needs to be answered in this stage is if the system fulfils all the needs of the target groups. This has been a broad, precise test and has a documented test plan which is written by system creator and his/her clients.

For the airport tests, the focus has been on system durability, user friendliness (stakeholders are able to work with applications easily) and whether the system users can manage other related tasks without any additional work load. Another focus has also been to check if there are any bugs which still exist in the system.

4 Discussion and Conclusion

As it is among the top 50 airlines world-wide, this airline needs to stay competitive by offering the highest quality passenger services and service levels. One of the key priorities for the airline is to offer excellent service to its clientele and to differentiate itself by implementing RFID technology not only to tag the passenger's baggage but also the passengers themselves.

The objectives of the RFID business case study were cost savings and other benefits such as enhanced safety and quality control, increased customer satisfaction, etc. The adoption of the RFID technology for the sorting and handling of baggage along the global supply chain provides a Win-Win-Win for the three main stakeholders, the airlines, the airports and the passengers [7].

According to the survey performed by the IATA [7], the RFID project is expected to offer a 9.7% savings opportunity. While barcode reading problems attribute 9.7% of the total baggage mishandlings to the airline, airports deal with a much larger baggage volume. An airport requires manual interaction to ensure that the baggage makes the intended flight. If RFID technology is put into use, it allows the airport to increase efficiency in the baggage handling operation. It is estimated a 12.5% savings opportunity, based upon analysis of airport baggage statistics. Another factor contributing to baggage mishandling is the failure to receive a baggage status message. The airline's survey results state that this factor alone will contribute to savings of around 11% [7].

While cost savings is the key business driver, there are other benefits that the RFID project can create. By implementing the latest RFID technology, it can improve the overall passenger service level by improving the tracking of mishandled baggage. The RFID technology tracks and records the baggage location when it is delivered to the wrong terminal or flight. This will make for more responsive and faster delivery of mishandled baggage.

On top of these benefits, the new technology will also enhance the speed and accuracy of baggage handling, especially when dual flight transfer is required. Passengers will be better informed, and the baggage delivery status can be tracked easily. All these functions will set passengers' mind at ease and in return increase the passengers' satisfaction toward the airline.

With the increasing transparency of the baggage handling process and a reduced baggage claims record, the RFID project will help the airline to build a strong brand image and set new standards of passenger service.

In addition, the baggage tagging initiative is also expected to reduce the overall operating cost, beyond the savings associated with baggage claims. With improved efficiency in baggage handling and tracking, the reduced number of baggage claims will free up customer service resources to carry out higher impact activities like customized services for frequent travelers. Enhanced baggage

handling will improve the resource planning capability and strengthen the decision making capabilities of the airline. The quick tracking function of the RFID system will help the airline to identify areas of failure, and can help to identify the problem caused by other carriers or airports. By feeding the information into IT tools like dashboards, it can help the airline to manage its global productivity and performance, and allow it to make necessary changes and decisions efficiently and effectively.

The result and logging of performance can contribute to a key service level measurement. It can assist the airline in planning for the resource allocation on one hand and become a key reference for negotiating contracts with Ground Handlers on the other.

The RFID project can improve the security management of the airline and airport. After the 9/11 incident, the Federal Aviation Administration (FAA) put great pressure on the airlines to ensure proper baggage-to-passenger matching. This in return increased the requirements for the tracking and visibility of all bags. Airlines can add different security levels into the baggage tag, and this together with the tracking mechanism can immediately identify and locate any misplaced bag. The system and information logged could become key references for security audits. The overall transparency of information can help both airlines and airports to enhance and optimize the baggage handling process. In this way, RFID can be instrumental in helping airports and airlines deal with security issues. Thus in an effort to increase safety standards, airports can get financial support from management and authorities.

After the successful implementation of the RFID baggage tagging system, the next phase of the project will be to extend the system to passenger tagging. By tracking the location of passengers with the RFID cards, customized services can be offered to boost the customer satisfaction and create a positive impact on the sales turnover. Via the information on the RFID embedded card of the premium passengers, they can be greeted in the language they prefer, and they can be offered their favourite newspaper and drinks once they enter the premium passenger lounge. This RFID card can also be a tracking device to help the airline better understand its passengers' profiles. By tracking which duty free shops passengers visit, which restaurant they go to, etc. the cards can be a source of information for the Customer Relationship Management system (CRM), which in turn can allow for custom-made programs to enhance business performance and improve customer loyalty.

Acknowledgement

We would like to thank executive editor Dr. Péter Tóth and referees for their valuable comments to improve the quality of this paper. We would also like to thank the Academic Writing & Advisory Center (AWAC) of Atilim University for nicely editing the manuscript.

References

[1] Koczor Z., Takács A., "Engineering Evaluation about the Role of Innovation in a Globalized Economy, Acta Polytechnica Hungarica, Vol. 5, No. 3, 2008, pp. 65-73

[2] Tzeng Shiou-Fen, Chen Wun-Hwa, Pai Fan-Yun (2008) Evaluating the Business Value of RFID: Evidence from Five Case Studies, *International Journal of Production Economics*, Volume 112, Issue 2, April 2008, pp. 601-613

[3] Curtin, J., Kauffman, R. J., Riggins, F. J. (2007) Making the ''MOST'' out of RFID Technology: A Research Agenda for the Study of the Adoption, Usage, and Impact of RFID, *Information Technology and Management*, 8 (2), pp. 87-110

[4] Lesakova L., "Innovations in Small and Medium Enterprises in Slovakia", Acta Polytechnica Hungarica, Vol. 6, No. 3, 2009, pp. 23-34

[5] Collins, J. (2004) Delta Plans US-Wide RFID System: the Airline Carrier will Spend up to $25 Million during the Next Two Years to Roll out an RFID Baggage-Handling System at Every US Airport it Serves. *RFID Journal*, 2 July, available at: www.rfidjournal.com/article/articleview/1013/1/1

[6] SITA (2006) Mishandled Baggage Costs Air Travel Industry $2.5 Billion a Year. *WIRELESS NEWS*, 10Meters, http://www.10meters.com

[7] IATA (2007) RFID Business Case for Baggage Tagging. available at http://www.iata.org/NR/rdonlyres/99091491-CB49-4913-BAB4-EA578CA814CC/0/RFIDforbaggagebusinesscase21.pdf

[8] IATA (2008) Fact sheet: Radio Frequency Identification (RFID) for aviation. available at http://www.iata.org/pressroom/facts_figures/fact_sheets/RFID.htm)

[9] Weinstein, D. (2005) RFID: A Technical Overview and its Application to the Enterprise, *IT Professional*, Vol. 7(3) pp. 27–33

[10] DeVries, Peter D. (2008) The State of RFID for Effective Baggage Tracking in the Airline Industry, *International Journal of Mobile Communications*, Vol. 6, No. 2, pp. 151-164

[11] Chang, Y. S., Oh, C. H., Whang, Y. S., Lee, J. J., Kwon, J. A., Kang, M. S., Park, J. S., Ung, Y. P. (2006) Development of RFID Enabled Aircraft Maintenance System. Industrial Informatics, 2006 IEEE International Conference on, Aug. 16-18, 2006, pp. 224-229

[12] Nath, B., Reynolds, F., Want, R. (2006) RFID Technology and Applications, *IEEE Pervasive Computing* 5, 1 (Jan. 2006) 22

[13] Al-Ali, A. S. A., Sajwani, F., Al-Muhairi, A., Shahenn, E. (2007) Assessing the Feasibility of Using RFID Technology in Airports. *RFID Eurasia*, 2007 1st Annual, Sept. 5-6, 2007, pp. 1-5

[14] Wyld, David C., Jones, Michael A., Totten, Jeffrey W. (2005a) Where is my Suitcase? RFID and Airline Customer Service, *Marketing Intelligence & Planning*, 2005, Volume 23, Issue 4, pp. 382-394

[15] Viswanadham, N., Prakasam, A., Gaonkar, R. (2006) Decision Support System for Exception Management in RFID Enabled Airline Baggage Handling Process. Automation Science and Engineering, 2006. IEEE International Conference on, CASE '06, Oct. 8-10, 2006, pp. 351-356

[16] Sample, K. B., Taylor, D. K., Rao, E. (2004) High Tech Aviation Security Program in Africa - a Model for Technology Transfer. Security Technology, 38th Annual International Carnahan Conference on, Oct. 11-14, 2004, pp. 270- 277

[17] Wyld, David C., Jones, Michael A., Totten, Jeffrey W. (2005b) I'm on Beale Street, but my Luggage is in Memphis … Egypt?: Deploying RFID-enabled Baggage Tracking Systems to Improve Airline Customer Service, *Academy of Marketing Studies Journal*, Vol. 9, Number 1, Jan-July, 2005

[18] AIM Global (2004) Flying high, RFID Connections. January 2004 from www.aimglobal.org/technologies/rfid/resources/articles/jan04/0401-bagtag.htm

[19] Wong, Y. F., Wu, P. W. K., Wong, D. M. H., Chan, D. Y. K., Fung, L. C., Leung, S. W., (2006) RFI Assessment on Human Safety of RFID System at Hong Kong International Airport. Electromagnetic Compatibility, 2006, EMC-Zurich 2006, 17th International Zurich Symposium on, Feb. 27-March 3, 2006, pp. 108-111

[20] Kelly, Eileen P., Erickson, G. Scott (2005) RFID Tags: Commercial Applications v. Privacy Rights, *Industrial Management & Data Systems*, Vol. 105, Issue 6, pp. 703-713

[21] Juels, A., Rivest, R. L., Szydlo, M. (2003) The Blocker Tag: Selective Blocking of RFID Tags for Consumer Privacy. In V. Atluri, ed. 8th ACM Conference on Computer and Communications Security, pp. 103-111, ACM Pres 2003

[22] Yin, R. K. (2003) Case Study Research: Design and Methods, 3rd Edition, Sage Publications, Thousands Oaks, CA

[23] Feagin, J., Orum, A., Sjoberg, G. (Eds.) (1991) A Case for Case Study, University of North Carolina Press, Chapel Hill, NC

[24] Van Der Blonk, H. (2003) Writing Case Studies in Information Systems Research, *Journal of Information Technology*, 18, pp. 45-52

[25] Lee, L. S., Fiedler, K. D., Smith, J. S. (2008) Radio Frequency Identification (RFID) Implementation in the Service Sector: A Customer-Facing Diffusion Model, *International Journal of Production Economics*, Vol. 112(2), pp. 587-600

[26] Hammer, M., Stanton, S. (1999) How Process Enterprises Really Work. *Harvard Business Review*, 77 (6), pp. 52-57

[27] Clark, T. H., Stoddard, D. B. (1996) Interorganizational Business Process Redesign: Merging Technological and Process Innovation, *Journal of Management Information Systems*, 13 (2), pp. 9-28

[28] Riggins, F. J., Mukhopadhyay, T. (1994) Interdependent Benefits from Interorganizational Systems: Opportunities for Business Partner Reengineering, *Journal of Management Information Systems*, 11 (2), pp. 37-57

[29] Chao, C. C., Jen, W. Y., Chi, Y. P., Lin, B. (2007) Improving Patient Safety with RFID and Mobile Technology. *International Journal of Electronic Healthcare*, Volume 3, Number 2, pp. 175-192

[30] Chuang, Y. W. (2005) An Exploration of Application of Radio Frequency Identification Technology in Hospitals. Working Paper, National Taiwan University

[31] Homs, C. (2004) Exposing the Myth of the 5-Cent RFID Tag, Forrester Research Inc., Cambridge, MA

Could You Check This, Please? Experiences in a Bilingual Environment

Lívia Szedmina

Subotica Tech, College of Applied Sciences
Subotica, Serbia
slivia@vts.su.ac.rs

Abstract: This paper describes some experiences of a proof reader and translator working in a very specific bilingual linguistic environment. Bilingualism has a peculiar effect on the use of technical English, as seen with language users at Subotica Tech. Vajdaság is a unique location in terms of language composition. While for many, Hungarian is the first language, there is some language transfer from Serbian, especially concerning the use of English technical terminology in native language texts. A number of examples are offered to show different patterns that may be explained by the overwhelming influence of the first language.

Keywords: technical English; bilingualism; Vajdaság; proofreading; correction

1 Introduction

This article is an overview of the experiences of a professional proof reader regarding the use of technical English by native speakers of Hungarian and native speakers of Serbian from the territory of Vajdaság. All linguistic examples were taken from papers written at Subotica Tech, where the author is currently employed. The author has been working as a translator and proof reader for technical English, and more specifically, for the fields of electrical and mechanical engineering, as well as for informatics. The English language encountered by the author bears some specific markers stemming from the influences of the first language, or conversely, they may appear due to the fact that the users are bilingual. The samples presented in this paper are taken from young professionals with university degrees in engineering (electrical and mechanical engineering, or computer sciences) except in one instance, where it is a doctoral degree in informatics.

With the unstoppable spread of IT technologies, and primarily the computer, English as a global technical lingua franca has toppled the previous primary technical language, German. Consequently, this means that hardly anyone can vie

for a career in the domain of engineering without a passable knowledge of the English language. The majority of people who have agreed to have their works analyzed from a language point of view for this study are involved in doctoral studies, where there is a constant need for presenting new research results, writing papers and giving presentations at conferences. While all of the participants did take English as a compulsory subject during their university studies, since then they have mainly used English passively, in written articles. For the sake of clarity, the term 'writer' will be used to denote the participants whose samples are presented in this work, while 'author' will refer exclusively to the author of this present article.

This work is not written with the intention of making concrete generalizations that hold true for the use of technical English of all speakers of Hungarian and Serbian in Vajdaság. The aim is to present some of the linguistic problems and repeated patterns that the author has met in her work based on the samples listed below.

2 Method

The method of proof reading has become a great deal easier with the widespread use of the word processor's 'track changes' option. However, the author's preferred method of correction is more old-school: to put the incorrect or unnecessary words in brackets, insert the correct version so both the incorrect and the now corrected text can be seen, and finally modify the font color of these to some shade of blue or turquoise. The author prefers the use of cold colors (softer shades of blue, green or turquoise) as these will appear less "offensive" to the original writer of the proofread text. This has been repeatedly confirmed in follow-up conversations regarding the proofread text, when the writers would point out that the usual color for correcting, red, made their own work seem much worse, much more "full of mistakes" than if the same number of mistakes were highlighted in turquoise. For the sake of visibility in a monochrome color setting, a different font was used, **Britannica Bold**, to highlight the corrections.

3 Technical English

Time is a crucial factor in technical English since the need for up-to-date knowledge does not allow users of manuals or technical descriptions of novel technologies to wait for the translation into a national language. Engineers need to be able to read, comprehend, and as the ultimate test, implement the new material written in technical English. While certainly a large corpus of the technical vocabulary dates back centuries and is morphologically firmly based on the

classical languages, novel scientific fields such as IT or robotics present a new challenge for professionals and proof readers alike.

Translations lead to another problem: does it make sense to translate generally known terms such as “download”, “online” or “relay cable” into national languages? Do translations such as “herunterladen” (German for “to download”) or “prenosni kabel” (Serbian for “relay cable”) promote national language identity or hinder international communication, or possibly both? Who makes a valid translator or proof reader for technical English? Is the “expert of English” required to have an engineering degree as well as advanced knowledge of English? Or can the person be a graduate in the English language and yet not need to have an in-depth knowledge of the technical, technological aspects?

All these are questions that the author has been faced with since the very beginning of her career. As a graduate in English language and literature with teaching qualifications, engineering and computer science were a world away. And yet, very soon upon taking a full time job as a teacher of English at Subotica Tech – College of Applied Sciences, the tasks included translating and proofreading technical English texts. The articles sent to the author came from a wide range of subjects, most that the author had never heard of before: fuzzy mathematics, heat engines, personalization of interactive learning materials, intrusion detection, product management, hexapod robot’s walk algorithms, or fixture-planning systems for box-shaped parts, just to name a few.

Proofreading an English text was the easier part, but translating a text written in Serbian (not the native language of the author) on a heavily technical topic was quite a challenge. Where does one turn to? Printed technical English dictionaries can only do so much since they tend to become outdated at an incredibly rapid rate. The best way to keep up was through online dictionaries and various forums. However, there were countless instances where the dictionary offered several translations for a given word (the word “pontosság” in metrology can be translated both as “precision” or “accuracy”, yet only one of the two, namely “accuracy”, is used). How can a non-engineer know this or make a competent decision?

4 Vajdaság

Subotica Tech is located in Szabadka (Subotica) in the northern part of Vajdaság, Serbia. There are two languages used for teaching at the College, Hungarian and Serbian. It is undoubted that bilingualism has a positive effect on any further learning of a new foreign language and this puts the people from Vajdaság in a peculiar situation. Nevertheless, there is still a strong influence of the mother tongue mainly due to the good (or not so good), largely passive knowledge of English and the lack of day-to-day active language practice, resulting at times in English texts that bear the linguistic structure of either Hungarian or Serbian.

However, it is not only a matter of transforming a Hungarian- or Serbian-influenced English text into correct technical English. Often problems arise due to the unique lingual and geo-political situation of the College and its wider surroundings, Vajdaság. While a large number of people, and specifically employees of Subotica Tech, speak Hungarian as their first language, or are bilingual, this 'local' variant of the Hungarian language is bound to be somewhat distinct from the Hungarian spoken in Hungary. There have been examples when this dialectal differences have been the reason for the rejection of a paper. The writers were told that there were numerous language instances in the paper which were not acceptable in "proper" Hungarian technical language, but are in common use in the technical Hungarian language used in Vajdaság. These instances referred mainly to the use of some English or international terms for which a definite Hungarian translation exists (even though these may sound cumbersome to the ears of vajdasági Hungarians). For example, the expression 'real time' used in a Hungarian text in the context of data processing was frowned upon, and the reviewers demanded the "correct" Hungarian expression, "valós időben", be used instead. Bilingualism may offer an explanation for this phenomenon. As mentioned before, most Hungarian lecturers at Subotica Tech are bilingual and hold their lectures and practices in both languages; thus linguistic transfer is hardly avoidable. Also, the majority of the teaching staff finished their studies in Serbia (or before the disintegration, in former Yugoslavia) where the language of education was Serbian. In addition, the Serbian language is much more relaxed about importing English technical terms and using them in a Serbian language environment. This tendency was evidenced in the vajdasági Hungarian use of the expression "real time".

5 Language Accuracy

In the section below there will be several examples regarding the question of accuracy in the scientific texts that the author has been sent for proofreading. Many of the writers whose works are referred to in this article have focused more on the practical implementation of technical English as opposed to accuracy. During the proofreading process the obvious grammar mistakes are corrected and should the author meet some ambiguous expressions, or where the content is not clear, comments are inserted. Occasionally proofreading is only a first step, followed by a personal consultation regarding the particular meaning that the writer wanted to use in his work.

5.1 Language Transfer

The following example is taken from a journal paper written by one bilingual writer (speaker of Serbian and Hungarian) and one monolingual writer (speaker of

Serbian), both mechanical engineers. The topic of the paper is the description of an adaptive product configurator regarding the thermal insulation of customers' homes taken from [1].

Example 1: "The reason for doing so is that customers usually, based on their belief, **sooner answer (earlier to)** questions that are of higher importance to them than **(to)** questions that are not."

The expression "earlier" is a direct transfer from Serbian as in the expression "potrošači bi ranije odgovarali na ovakva pitanja…" where the word "ranije" means "earlier" in the temporal sense, and not in the comparative sense, as e.g. the word "rather."

5.2 Passive Voice

One typical mistake that is present is almost every English language technical text that the author is given for proofreading is the lack of the use of the Passive Voice. This feature is not entirely surprising, as both the Hungarian language and the Serbian language are a lot less prone to using the Passive Voice than English. In Hungarian and Serbian technical texts the first person plural is a perfectly acceptable form, while English requires the "drier" version of the Passive Voice. All instances below are taken from [2], a paper written by a native Hungarian but bilingual writer for the conference Symposium on Intelligent Systems, SISY 2008 on the topic of e-learning and interactive learning:

Example 2: "**(By a method we mean a) A method involves the** notion of adaptation that can be presented at **(the) a** conceptual level."

Example 3: "**These styles were then grouped (He grouped these styles)** into the "families of learning styles".

Or, in yet another, Example 4: "**(On t)The next picture (we) presents** Peter Brusilovsky's taxonomy of adaptive hypermedia technologies [Brusilovsky, 2001]."

From the proof reader's point of view this means that usually the entire sentences has to be rephrased to comply with the English "requirement" for the Passive Voice. However, in Example 2 it can be seen that a significantly simpler expression ("by a method we mean") was used which was then replaced by a more professional-sounding word ("indicates"). The proof reader's task at times involves less error correction, but rather turning *what* the writer wanted to say into *how* the writer of the text wanted to say it.

Example 5 is quoted from [3] and it presents an instance where the Passive Voice was used with a Serbian sentence structure. The writers are native speakers of Serbian. While one of the writers is bilingual, Serbian is still the first language.

Example 5: "**With the theme of statistical intrusion detection have dealt the authors in publications [1]-[8].**"

which was then corrected according to the English sentence order of the Passive Voice:

"**The authors have dealt with the (theme) topic of statistical intrusion detection (have dealt the authors)** in publications [1]-[8]."

5.3 Technical English vs. Colloquial English

English is undeniably very much present in any science- or IT-oriented person's life; it comes in the form of movies shown in the cinema, songs played on the radio, newsletters subscribed to, etc. Thus it is not always an easy task to discern the English language of popular culture from the English language of the scientific world. This is underlined by Example 6. from [4] shown below, where a grammatical mismatch and an abbreviated form of the auxilliary verb had to be corrected.

Example 6.: On the other hand, obstacles that are illuminated with infrared light **(doesn't) do not** reflect light from **the** visible color spectrum.

5.4 Articles

One of the most common sources of grammatical mistakes is the use of articles. In this field a specific distinction between native speakers of Serbian and native speakers of Hungarian can be determined. The Hungarian language contains both definite and indefinite articles and makes ample use of them. This makes it easier for Hungarian native speakers to "sense" where articles are needed in the English text. (For the sake of argument, an obvious fact will be disregarded, namely that intermediate or advanced speakers of English should *know* and not only *sense* where articles are used.) As opposed to this, Serbian native speakers will leave out the article more often, as it does not exist in the Serbian language. Two examples will lend some support for the statement above:

Example 7: In both cases **the** video chips are sensitive not only to rays from **the** visible spectrum of light, but also to rays from **the** IR spectrum of light. (Serbian native, with every article needed in the text missing, taken from [4].)

Example 8: From the perspective of social constructivism, the function of individual differences on skills, aptitudes and learning preferences could have **an** impact **on** the application of technology in classroom settings. (Hungarian native, taken from [5], where only the indefinite article is missing, which corresponds to the lack of an article in the Hungarian original.)

5.5 Generalization

This section refers to grammar mistakes made by not so advanced users of English who make generalizations based on assumption. This type of language learner behavior becomes less frequent with an increase in language proficiency. However, crucial from a language learning point of view, it shows a degree of independence in the use of the language that will enable the user to develop their language knowledge with less reliance upon a language teacher. The writer of the example presented below is a very independent language learner, having developed most of her skills by herself, relying heavily on the knowledge of fuzzy mathematics and constantly working with English language scientific texts. In this particular case [6] a generalization of noun formation is shown. While the writer must have been aware that many nouns are formed with the suffix -ness, in the given transformation, from the adjective 'uncertain', the suffix –ty is needed.

Example 9: "Let **(we)us** assume that we have T2 fuzzy sets as the rule premise of the rules and as the system input, described respectively by the following equations where the fuzziness represent the **(uncertainness) uncertainty** of T1 fuzzy membership values of these sets."

Conclusion

This article tries to present some of the issues that a proof reader has encountered in a bilingual, Hungarian-Serbian environment within a specific institution of higher education, Subotica Tech. These examples are naturally not meant to make generalizations encompassing all speakers of Serbian and Hungarian, but certainly they support a pattern that English language teachers working in the field of these languages have experienced and that English language proof readers have met with repeatedly. There are constant challenges for proof readers and translators as the technical English language continues to evolve and keep up with technical innovations. On the other hand, it is also important to be able to present one's knowledge on a subject, one's research and contribution to science in their own language, keeping the overwhelming influence of English at bay. The Hungarian as used in Vajdaság may to a certain degree be different from the Hungarian spoken in Hungary, with some transfer from Serbian, the language of the environment. However, while language must be used correctly, it must also be used authentically, as it is understood by its speakers. In this sense, there may well be some instances of English technical terms within a Hungarian text, and there may be international words used within the bounds of Hungarian grammar, but this will make a distinct, vajdasági Hungarian text.

References

[1] Zoran Anišić, Igor Fürstner, Ilija Ćosić: Mass Customization: Some Trends and Research, DAAAM International Scientific Book 2009, B. Katalinic (Ed.), Published by DAAAM International Vienna, ISBN: 978-3-901509-71-1, ISSN: 1726-9687, Vienna, Austria

[2] Robert Pinter: On the road to Adaptive and Intelligent Webbased Educational Systems, 6th International Symposium on Intelligent Systems and Informatics, September 26-27, 2008, pp. 318-320, Subotica, Serbia. IEEE Catalog Number: CFP0884C-CDR, ISBN: 978-1-4244-2407-8, Library of Congress: 2008903275

[3] P. Čisar, S. Maravić Čisar: Skewness and Kurtosis in Function of Selection of Network Traffic Distribution, Acta Polytechnica Hungarica, accepted for publication

[4] Bojan Kuljić, János Simon, Tibor Szakáll: Pathfinding Based on Edge Detection and Infrared Distance Measuring Sensor, Acta Polytechnica Hungarica - Journal of Applied Sciences, Volume 6, Issue Number 1, pp. 103-116, Budapest, Hungary, 2009, ISSN 1785-8860

[5] Robert Pinter: Measuring The Preferred Learning Style**,** Proceedings of the 9th Wseas International Conference on Signals, Speech and Image Processing/9th Wseas International Conference on Multimedia, Internet & Video Technologies, Vol. br., str. 132-134, Budapest, Hungary, September 3-5, 2009

[6] Márta Takács: Fuzziness of Rule Outputs by the DB Operators-based Control Problems, in Book Series: Studies in Computational Intelligence, Book: Towards Intelligent Engineering and Information Technology, Volume 243/2009, ISBN: 978-3-642-03736-8, Publisher: Springer Berlin / Heidelberg, pp. 653-663