ADVANCES IN INFORMATION TECHNOLOGY

From Artificial Intelligence To Virtual Reality

Edited By: Al-Dahoud Ali

UBICC Publisher

ADVANCES IN INFORMATION TECHNOLOGY

From Artificial Intelligence To Virtual Reality

Edited By

Al-Dahoud Ali

Professor, Al-Zaytoonah University, Amman, Jordan

UBICC Publisher

Published by UBICC Publisher

UBICC Journal is a part of UBICC Publishers
Typesetting: Camera-ready by author, data conversation by UBICC Publishing Services
UBICC Journal is indexed by:

- EBSCOhost Online Research Databases
- The Index of Information Systems Journals
- Documents in Computing and Information Science (DOCIS)
- Ulrich's Periodicals Directory
- Directory of Open Access Journals (DOAJ)
- The Index of Information Systems Journals
- Microsoft LIBRA for Academic Search

For permission to use this material from this text, contact us by
Tel +1-347-4149239
Fax +1-212-901-6990
www.ubicc.org
info@ubicc.org

ISBN-13: 978-1468128994
ISBN-10: 146812899X

Contents

Preface

This book contains the extended version of the best published papers in ICIT11, it includes 7 chapters covers the following area: distributed systems, parallel computing, embed artificial intelligence, web services networks, e-learning, discrete transport systems, and routing in wireless sensors network.

Chapter One: This chapter present a new slot selection algorithm for job batch scheduling in distributed computing with non-dedicated resources is presented. The work discloses a scheduling scheme that features multi-variant search. Two algorithms of linear complexity for search of alternative variants are proposed. Having several optional resource configurations for each job makes an opportunity to perform an optimization of execution of the whole batch of jobs and to increase overall efficiency of scheduling.

Chapter Two: This chapter focuses on parallel computing, evaluation of its performance and two kinds of parallel programming examples based on master-worker paradigm. Such programming can be realized and executed on CPUs with multiple processor cores.

Chapter Three: This chapter addresses an importance of the information systems which display and measure person's competences. E-portfolios, as the part of information systems, plays significant role to enhance people's self-confidence and motivate learners to improve their competencies. However, nowadays lifelong learning paradigm requires new operation of thinking and new approach to provide appropriate services to target groups. This study seeks to identify perspective ways to make ePortfolio systems more flexible and intelligent, and also set the objectives to assist learners or job seekers to make odds even in their current competences level and required at the labor market.

Chapter Four: This chapter presents three web services network models based on the notions of dependency, interaction and similarity. Using the WS-NEXT extractor the author's instantiate the models with a collection of Web services descriptions. They take advantage of complex network properties to provide an analysis of the Web services networks. Those networks and the knowledge of their topological properties can be exploited for the discovery and composition processes.

Chapter Five: This chapter highlights the important factor which often cripples e-learning efficiency and is the lack of individual approach to a student. In order to improve knowledge absorption students have be treated accordingly to their specific and preferred learning style. The chapter discusses the results of studies in the literature for modeling learning styles. It presents the basic aspects of the problem, and selected learning models. The chapter ends with a detailed description of the selected model suitable to be applied in e-learning system - including the manner of its recognition and how to store it in a computer system.

Chapter Six: The chapter is focused on the human resource influence on dependability of discrete transportation systems (DTS). The human resource means the driver of the vehicle. We add him/her as a new element of the system description. The dependability means the combination of the reliability and functional parameters of the DTS. This way the analysis of the DTS behavior seems to be more sophisticated. The unified containers transported by trucks with the set of time-type assumptions are the essence of the system discussed. The proposed method is based on modeling and simulating of the system behavior. The income of containers is modeled by a stochastic process. Each container

has a source and destination address. The central node is the destination address for all containers generated in the ordinary nodes. We also propose the heuristic management approach as well as the functional metric for DTS and we test the example system based on the real data.

Chapter Seven: In this chapter; the authors present the answer of the following question: how to realize in WSN networks energy awareness spatial routing employing relations. Considering the routing issue in WSN networks we looked for such a solution that will allow defining the strategy for a global network operation without giving detailed, explicit orders that can interfere with nodes local activities. As the result of research carried out, we proposed novel approach using the relationships. This approach allows for the routing area sizing and implementation costs allocation for a number of nodes in a network.

Professor AL-Dahoud Ali
Editor

Author Biography

Al-Dahoud Ali, PhD., is a full Professor at Al-Zaytoonah University, Amman, Jordan. He took his High Diploma from **FON** University Belgrade 1986, PhD from La Sabianza1/Italy and Kiev Polytechnic/Ukraine, on 1996. He worked at Al-Zaytoonah University since 1996 until now. He worked as visiting professor in many universities in Jordan and Middle East, as supervisor of master and PhD degrees in computer science. He established the ICIT conference since 2003 and he is the program chair of ICIT until now. He was the Vice President of the IT committee in the ministry of youth/Jordan, 2005, 2006. Al-Dahoud was the General Chair of (ICITST-2008), June 23–28, 2008, Dublin, Ireland (www.icitst.org). Al-Dahoud has directed and led many projects sponsored by NUFFIC/**Netherlands**, and **Spanish** Agency for International Development Cooperation. He published many books and journal papers, and participated as keynote speaker in many conferences worldwide. He also organize and participated in many international conference as general chair, International Chair, program chair, session's organizer or in the publicity committee: Dr Al-Dahoud worked as Editor in Chief or guest editor or in the Editorial board in many International Journals:

Professor AL-Dahoud Ali

RESOURCE CO-ALLOCATION ALGORITHMS IN DISTRIBUTED JOB BATCH SCHEDULING

Victor V. Toporkov, Alexander Bobchenkov, Dmitry Yemelyanov and Anna Toporkova

Summary

In this work, we present slot selection algorithms for job batch scheduling in distributed computing with non-dedicated resources. Jobs are parallel applications and these applications are independent. Existing approaches towards resource co-allocation and parallel job scheduling in economic models of distributed computing are based on search of time-slots in resource occupancy schedules. The sought time-slots must match requirements of necessary span, computational resource properties, and cost. Usually such scheduling methods consider only one suited variant of time-slot set. This work discloses a scheduling scheme that features multi-variant search. Two algorithms of linear complexity for search of alternative variants are proposed. Having several optional resource configurations for each job makes an opportunity to perform an optimization of execution of the whole batch of jobs and to increase overall efficiency of scheduling.

Introduction

Job control is among the most difficult problems in the enterprise of distributed data processing in the case of non-dedicated resources that are shared with their owners. One must take into account the in homogeneity, changing composition, different owners of different nodes, and the scale of the computing environment. Economic models for resource management and scheduling are very effective in distributed computing, including Grid[1, 2], utility computing [3], cloud computing [4], and multi agent systems[5]. Pricing problem depending on the desired quality of service is a challenge in economic models of scheduling. There is a good overview of its solution in [6] along with the description of some approaches to forming of different deadline and budget constrained strategies of scheduling in distributed computation. In [7] heuristic algorithms for slot selection based on user-defined utility functions are

introduced.

While implementing some economic policy, resource brokers usually optimize the performance of a specific application [1, 6, and 7] in accordance with the application-level scheduling concept [8]. When establishing virtual organizations (VO), the optimization is performed for the job-flow scheduling [9, 10]. Corresponding functions are implemented by a hierarchical structure that consists of the meta scheduler and subordinate resource managers or local batch-job management systems [8-10]. Within the framework of a model proposed in [2] and based on hierarchical scheme of job flow distribution [9, 10] there is an interaction between users launching their jobs and computational resources owners [8]. It is supposed that resources are non-dedicated (inseparable) that is along with global flows of external user's jobs, owner's local job flows exist inside the CPU node domains. The interests of the said users and owners are often contradictory.

Each independent user is interested in the earliest launch of his job with the lowest costs (for example, the resource usage fee) and the owners, on the contrary, try to make the highest income from their resources. VO administrators are interested in maximizing the whole VO performance in the way that satisfies both users and owners. The meta scheduler [8-10] implements the economic policy of VO based on local CPU schedules. The local schedules are sets of slots coming from local resource managers or schedulers in the node domains.

A single slot is a time span that can be assigned to a task, which is a part of a parallel (multiprocessor) job (Fig. 1). A job schedule is a set of time slots. A batch schedule is a set of slot sets. Every slot set for the execution of the i^{th} job in the batch is defined with the cost and time of resource usage. We assume that scheduling of independent jobs runs iteratively and is based on dynamically updated local schedules [2]. The scheduling is performed cycle by cycle on the sets of preselected slots. During each cycle the sets of available slots are updated based on the information from local resource managers. The launch of any job requires co-allocation of a specified number of slots. Thus, during every cycle of the job batch scheduling two problems have to be solved. Selecting alternate set of slots (alternatives) that meet the requirements (resource, time, and cost).Choosing a slot combination that could be efficient or optimal in terms of complete job batch execution in current cycle of scheduling.

Figure 1. A Slot and Its Attributes

To realize the scheduling scheme described above, first of all, we need to propose the algorithm of finding a set of alternative slot sets**.** Local CPU node schedules in this case are represented as a list of available slots. An optimization technique for the second phase of this scheduling scheme was proposed in [2]. It is implemented by dynamic programming methods using multiple criteria in accordance with the VO economic policy.

The resource requirements are arranged into a resource request [8] containing the usage time, required number and characteristics of computational nodes (clock speed, RAM volume, disk space, operating system etc.) and the maximum price of its use. The launch of a multiprocessor job requires a co-allocation of the specified number of slots. The challenge is that slots associated with different CPU nodes may have arbitrary start and finish points that do not coincide. In its turn, processes of the parallel job must start synchronously. If the necessary number of slots with attributes matching the resource request is not accumulated then the job will not be launched. This job is joined another batch, and its scheduling is postponed till the next iteration (Fig. 2).

In the case of homogeneous nodes, a set of slots for any job is represented with a rectangle window. In the case of CPUs with varying performance, that will be a window with a rough right edge, and the usage time is defined by the execution time of the task that is using the slowest CPU (Fig. 3). We propose two algorithms for slot selection that feature linear complexity $O(m)$, here m is the number of available time-slots.

Existing slot search algorithms, such as backfilling [11, 12], do not support environments with heterogeneous and non-dedicated resources, and, moreover, their execution time grows substantially with increase of the number of slots. Backfilling is able to find an exact number of concurrent slots for tasks with identical resource requirements and homogeneous resources. We take a step further, so proposed algorithms deal with heterogeneous resources and jobs with different tasks.

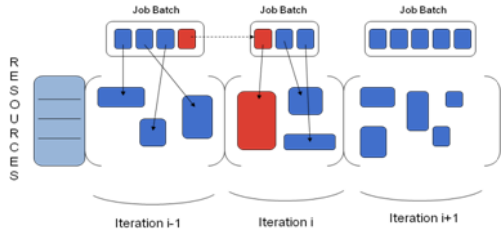


Figure 2. Scheduling Iterations

We propose two algorithms for slot selection that feature linear complexity $O(m)$, here m is the number of available time-slots. Existing slot search algorithms, such as backfilling [11, 12], do not support environments with heterogeneous and non-dedicated resources, and, moreover, their execution time grows substantially with increase of the number of slots. Backfilling is able to find an exact number of concurrent slots for tasks with identical resource requirements and homogeneous resources. We take a step further, so proposed algorithms deal with heterogeneous resources and jobs with different tasks.

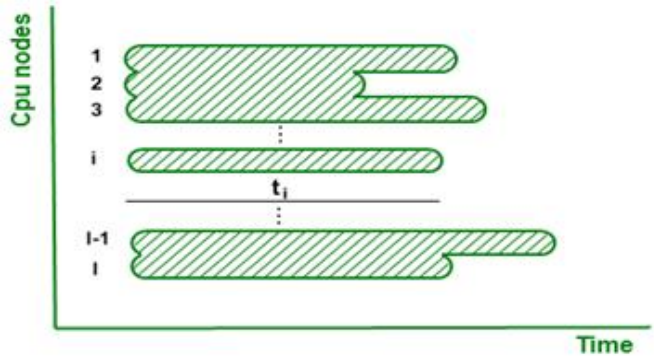


Figure 3. Window with a Rough Right Edge

Slot Sets Searching Scheme

The scheduling scheme works iteratively, during the iteration it consequentially searches for a single alternative for each job of the batch (see Fig. 2). In case of successful slot selection for the i^{th} job (Fig. 4), the list of viewed slots for the $(i+1)^{th}$ job is modified. All time spans that are involved in the i^{th} job alternative are excluded from the list of vacant slots. The selection of slots for the $(i+1)^{th}$ job is performed on the list modified with the method described above (Fig. 5).

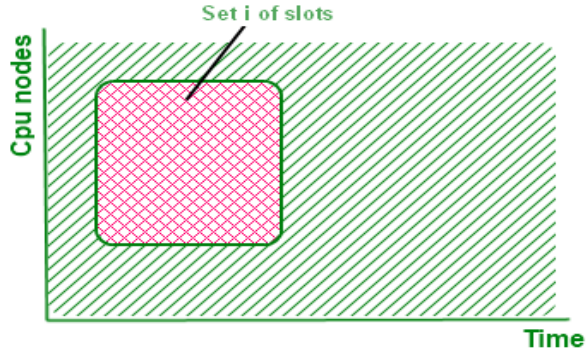


Figure 4: Alternative Searching Example: The i^{th} Set Of Slots

Suppose, for example, that there is a slot K′ among the appropriate window slots (Fig. 6). Then its start time equals to the start time of the window: K′.startTime = window.startTime and its end time equals to K′.end=K′.start + t_j, where t_j the evaluation of a part is job runtime on the CPU node, on which the slot is allocated. This follows from a definition of a window with a rough right edge (see Fig. 3). Slot K′ should be subtracted from the original list of available system slots. First, we need to find slot K – the slot, part of which is K′, and then cut K′ interval from K. So, in general, we need to remove slot K from the ordered slot list and insert two new slots K_1 and K_2.

Their start, end times are defined as follows:

1. K_1.startTime = K.startTime, K_1.endTime = K′.startTime,
2. K_2.startTime = K′.endTime, K_2.endTime = K.endTime.

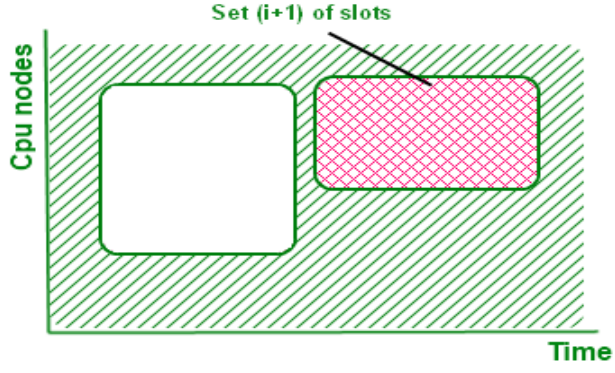


Figure 5: Alternative Searching Example: The $(i+1)^{th}$ Set Of Slots

Slots K_1 and K_2 have to be added to the slot list given that the list is

sorted by non-decreasing start time order (Fig. 6). Slot K_1 will have the same position in the list as slot K, since they have the same start time. If slots K_1 and K_2 have a zero time span, it is not necessary to add them to the list.

After the last of the jobs is processed, the algorithm starts next iteration from the beginning of the batch and attempts to find other alternatives on the modified slot list. Alternatives found do not intersect in processor time, so every job could be assigned to some set of found slots without the revision of other jobs assignments. The search for alternatives ends when on the current list of slots the algorithm cannot find any suitable set of slots for any of the batch jobs. Single alternative search algorithm implementation becomes a serious question because characteristics of a resulting set of slots solely depend on it. Doing a search routinely in every scheduling cycle imposes a requirement of an algorithm having complexity as low as possible. An optimization technique for choosing optimal or efficient slot combinations was proposed in [2]. It is implemented by dynamic programming methods using multiple criteria in accordance with the VO economic policy (Fig. 7). We consider two types of criteria in the context of our model. These are the execution cost and time measures for the job batch using the suitable slot combination. The first criteria group includes the total cost of the job batch execution. The VO administration policy and, partially, users' interests are represented with the execution time criterion for all jobs of the batch. In order to forbid the monopolization of some resource usage by users, a limit is put on the maximum value for a total usage cost of resources in the current scheduling iteration. We define as a budget of the VO. The total slots occupancy time represents owners' urge towards the balance of global (external) and local (internal) job shares. If we consider the single-criterion optimization of the job batch execution, then every criterion (the total cost of the job batch execution or the execution time) must be minimized with given constraints for the interests of the particular party - the user, the owner and the VO administrator [2].

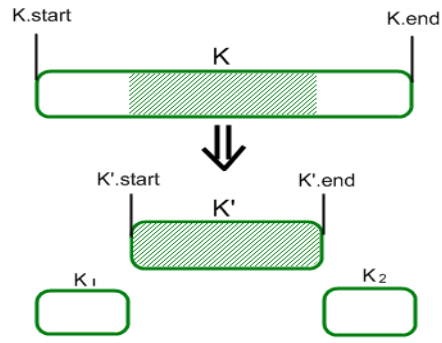


Figure 6: Slot Subtraction

We propose two algorithms that feature linear complexity. Existing slot search algorithms do not support environments with inseparable resources, and, moreover, their execution time grows substantially with increase of the slot number. Assuming that every node has at least one local job scheduled, the backfill algorithm [11, 12] has quadratic complexity in the slot number. Although backfilling supports multiprocessor jobs and is able to find a rectangular window of concurrent slots, this can be done provided that all available computational nodes have equal performance (processor clock speed), and resource requests of tasks of any job are homogeneous. We consider algorithms which deal with heterogeneous resources and jobs, and can form non-rectangular time-slot windows as a result.

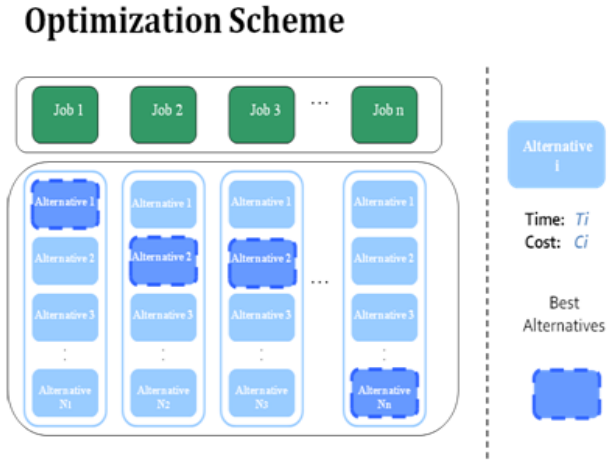


Figure 7: Choosing an Optimal Slot Combination

Slot Search Algorithms

Let us consider one of the resource requests associated with a job in a batch (see Fig. 2). The resource request specifies N concurrent time-slots reserved for time span t with resource performance rate at least P and maximal resource price per time unit not higher, than C.

Class Slot is defined to describe a single slot:

```
public class Slot
{
public Resource cpu;        // resource on which the slot is
allocated
public int cash;            // usage cost per time unit
public int start;           // start time
public int end;             // end time
public int length;          // time span
}
```

Input data include available slots list, and slots being sorted by start time in ascending order (Fig. 8). The search algorithm requires a sorted list to function and guarantees examination of every slot if this requirement is fulfilled. If the necessary number N of slots is not accumulated, then the job scheduling is postponed until the next iteration.

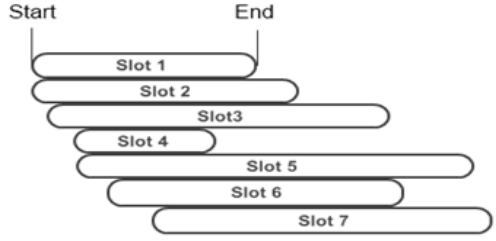


Figure 8: An Ordered List Of Available Slots

Class Window is defined to describe a single window:

```
public class Window
{
      int id;                    // window id
      public int cash;           // total cost
      public int start;          // start time
      public int end;            // end time
      public int length;         // time span
int slotsNumber;          // number of required slots
      ArrayList<Slot> slots;  // window slots
      }
```

Here a slot search algorithm for a single job and resource charge per time unit is described. It is an **A**lgorithm based on **L**ocal **P**rice of slots

(ALP) with a restriction to the cost of individual slots. In process, we construct a slot window as mentioned before.

1. Sort the slots by start time in ascending order (see Fig. 8).
2. From the resulting slot list the next suited slot s_j is extracted and examined.

Slot s_j suits, if following conditions are met:

a) Resource performance rate $P(s_j) \geq P$;

b) Slot length (time span) is enough (depending on the actual performance of the slot's resource) $L(s_j) \geq tP(s_j)/P$;

c) Resource charge per time-unit $C(s_j) \leq C$.

If conditions are met, the slot is successfully added to the window list.

1. We add a time offset d_j of current j^{th} slot in relation to $(j-1)^{\text{th}}$ to the length of the window.
2. Slots whose length has expired considering the offset d_j are removed from the list.

The expiration means that remaining slot length $L'(s_j)$, calculated like shown in step 2°b, is not enough assuming the j^{th} slot start is equal to the last added slot start:

$$L'(s_j) < (t - (T_{last} - T(s_j)))P(s_j)/P,$$

Where $T(s_j)$the slot's is start time, T_{last} is the last added slot's start time.

3. Go to step 2, until the window has N slots.
4. End of the algorithm.

We can move only forward through the slot list (see Fig. 8). If we run out of slots before having accumulated N slots, this means a failure to find the window for a job and its scheduling is postponed by the meta scheduler until the next scheduling cycle (see Fig. 2). Otherwise, the window becomes an alternative slot set for the job. ALP is executed cyclically for every job i in the batch. Having succeeded in the search for window for the i^{th} job, the slot list is modified with subtraction of formed window slots (see Figs. 4-6). Therefore slots of the already formed slot set are not considered in processing the next job in the batch.

In the economic model [2] a user's resource request contains the maximal resource price requirement that is a price which a user agrees to pay for the resource usage. But this approach narrows the search space and restrains the algorithm from construction of a window with more

expensive slots. The difference of the next proposed algorithm is that we replace maximal price C per time-unit requirement by a maximal budget of a job. It is an Algorithm based on Maximal job Price (AMP). The maximal budget is counted as $S = CtN$, where t a time is span to reserve, and N is the necessary slot number. Then, as opposed to ALP, the search target is a window, formed by slots, whose total cost will not exceed the maximal budget S. In all other respects, AMP utilizes the same source data as ALP. Let us denote additional variables as follows: N_S – current number of slots in the window; M_N – total cost of first N slots.

Here we describe AMP approach for a single job.

1. Find the earliest start window, formed by N slots, using ALP excluding the condition 2°c (see ALP description above).
2. Sort window slots by their cost in ascending order. If $M_N \le S$, go to 4, so the resulting window is formed by first N slots of the current window, they are returned to the source slot list. Otherwise, go to 3.
3. Add the next suited slot to the list following to conditions 2°a and 2°b of ALP. Assign the new window start time and check expiration like in the step 4° of ALP. If we have $N_S < N$, then repeat the current step. If $N_S \ge N$, then go to step 2. If we ran out of slots in the list, and $N_S < N$, then we have algorithm failure and no window is found for the job.
4. End of the algorithm.

We can state three main features that distinguish the proposed algorithms. Both algorithms consider resource performance rates. This allows forming time-slot windows with uneven right edge (we suppose that all concurrent slots for the job must start simultaneously).

1. Both algorithms consider maximum price constraint which is imposed by a user.
2. Both algorithms have linear complexity $O(m)$ where m is slot number in the list: we move only forward through the list (see Fig. 8), and never return or reconsider previous assignments.

AMP Search Example

In this example for the simplicity and ease of demonstration we consider the problem with a uniform set of resources, so the slot windows will have a rectangular shape without the rough right edge. Let us

consider the following initial state of the distributed computing environment (Fig. 9). In this case there are six processor nodes cpu1-cpu6. Each has its own unit cost (cost of its usage per time-unit), which is listed in the column to the right of the processor name. In addition there are seven local tasks p1-p7 already scheduled for the execution in the system under consideration.

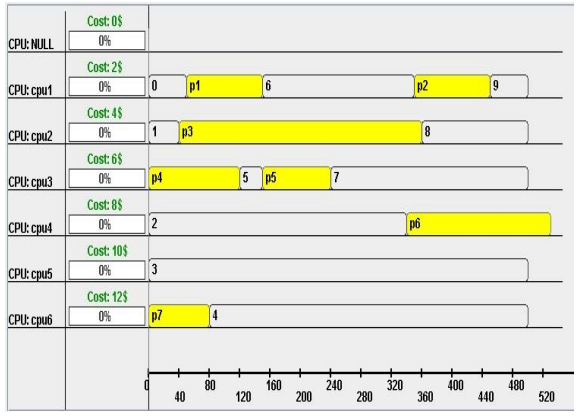


Figure 9: Initial State of Distributed Environment

Available system slots are drawn as rectangles 0...9 (see Fig. 9). Slots are sorted by non-decreasing time of start and the order number of each slot is indicated on its body. For the clarity, we consider the situation where the scheduling cycle includes the batch of only three jobs with the following resource requirements.

Job 1 requirements:

- The number of required processor nodes: 2 runtime: 80
- Maximum total "window" cost per time: 10

Job 2 requirements:

- The number of required processor nodes: 3 runtime: 30
- Maximum total "window" cost per time: 30

Job 3 requirements:

- The number of required processor nodes: 2 runtime: 50
- Maximum total "window" cost per time: 6

According to AMP alternatives search, first of all, we should form a list of available slots and find the earliest alternative (the first suitable window) for the first job of the batch. We assume that Job1 has the highest priority, while Job3 possesses the lowest priority.

The alternative found for Job 1 (Fig. 10) has two rectangles on cpu1 and cpu4 resource lines on a time span [150, 230] and named W1. The total cost per time of this window is 10. This is the earliest possible

window satisfying the job's resource request. Note that other possible windows with earlier start time are not fit the total cost constraint. Then we need to subtract this window from the list of available slots and find the earliest suitable set of slots for the second batch job on the modified list. Further, a similar operation for the third job is performed (see Fig. 10). Alternative windows found For each job of the batch are named W1, W2, and W3 respectively. The earliest suitable window for the second job (taking into account alternative W1 for the first job) consists of three slots on the cpu1, cpu2 and cpu4 processor lines with a total cost of 14 per time unit. The earliest possible alternative for the third job is W3 window on a time span of [450, 500].

Further, taking into account the previously found alternatives, the algorithm performs the searching of next alternative sets of slots according to the job priority. The algorithm works iteratively and makes an attempt to find alternative windows for each batch job at each iteration (Fig. 11). Fig. 11 illustrates the final chart of all alternatives found during search. Note that in ALP approach the restriction to the cost of individual slots would be equal to 10 for *Job 2* (as it has a restriction of total cost equals to 30 for a window allocated on three processor nodes). So, processor cpu6 with a 12 usage cost value is not considered during the alternative search with ALP algorithm. However it is clear that in the presented AMP approach eight alternatives have been found. They use the slots allocated on cpu6 line, and thus fit into the limit of the window total cost.

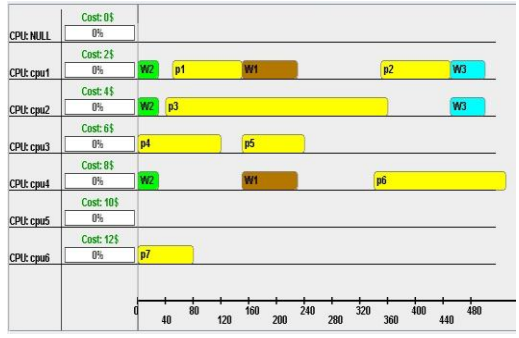


Figure 10: Alternatives Found After the First Iteration

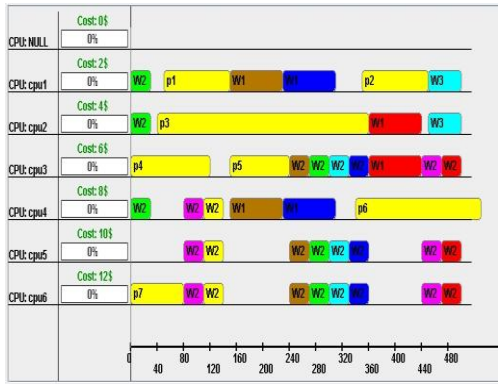


Figure 11: The Final Chart of All Alternatives Found

Simulation Studies

The experiment consists in comparison of job batch scheduling results using different sets of suitable slots founded with described above AMP and ALP approaches. The alternatives search is performed on the same set of available vacant system slots. During the single simulated scheduling cycle the generation of an ordered list of vacant slots and a job batch is performed. The alternatives are found using algorithms ALP and AMP realized in **SlotProcessor** class with the main function public slot Processor Result find Alternatives (ArrayList<Resource Request> requests, VOEnvironment environment, slot Processor Settings)

To perform a series of experiments we found it more convenient to generate an ordered list of available slots (see Fig. 8) with pre-assigned set of features instead of generating the whole distributed system model and obtaining available slots from it.

Slot Processor class was used in order to carry out the experiment series. It realizes described ALP and AMP window search using the list of available slots in no decreasing start time order as an input parameter.

Slot Generator and Job Generator classes are used to form the ordered slot list and the job batch during the experiment series. Here is the description of the input parameters and values used during the simulation.

Slot Generator:

- Number of available system slots in the ordered list varies in [120, 150]
- Length of the individual slot in [50, 300]
- Computational nodes performance range is [1, 3], that is the

environment is relatively homogeneous

- The probability that the nearby slots in the list have the same start time P = 0.4
- The time between neighbor slots in the list is in [0, 10]
- The price of the slot is randomly selected from [0.75p, 1.25p], where p = (1.7) to the (Node Performance)

Job Generator:

- Number of jobs in the batch [3, 7]
- Number of computational nodes to find is in [1, 6]
- Length representing the complexity of the job [50, 150]
- The minimum required nodes performance [1, 2]

All job batch and slot list options are random variables that have a uniform distribution inside the identified intervals. Let us consider the task of a slot allocation during the ***job batch execution time minimization (TM)*** by the technique proposed in [2]. The number of 25000 simulated scheduling cycles was carried out. Only those experiments were taken into account when all of the batch jobs had at least one suitable alternative of execution. AMP algorithm exceeds ALP by 35% with respect to the target optimization. An average batch job execution time for alternatives found with ALP was 59.85, and for alternatives found with AMP: 39.01 (Fig. 12). It should be noted, that an average cost of batch job execution for ALP method was 313.56, while using AMP algorithm average job execution cost was 369.69, that is 15% more (Fig. 13). Scheduling results comparison for the first 300 experiments can be viewed in Fig. 14.

According to the results of the experiments we can conclude that the use of AMP minimizes the total batch execution time though the cost of the execution increases. Relatively large number of alternatives found increases the variety of choosing the efficient slot combination [2] using the AMP algorithm. Now let us consider the task of slot allocation during the ***job batch execution cost minimization (CM)*** [2]. The results of 8571 single experiments in which all batch jobs were successfully assigned to suitable set of resources using both slot search procedures were collected. Average batch job execution cost for ALP algorithm was 313.09, and for alternatives found with AMP: 343.3.

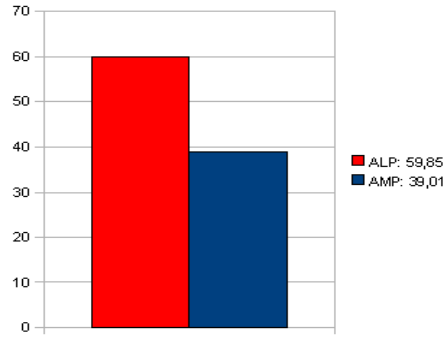


Figure 12: Total Job Execution Time Using ALP and AMP in TM

It shows the advantage in target criterion of only 9% for ALP approach over AMP (Fig. 15).Average batch job total execution time for alternatives found with ALP was 61.04. Using AMP algorithm average job execution time was 51.62, which is 15% less than using ALP (Fig. 16). Average number of slots processed in a single experiment was 135.11. This number coincides with the average number of slots for all 25000 experiments, which indicates the absence of decisive influence of available slots number to the number of successfully scheduled jobs. Average number of batch jobs in a single scheduling cycle was 4.18. This value is smaller than average over all 25000 experiments. With a large number of jobs in the batch ALP often was not able to find an alternative sets of slots for a certain jobs and experiment was not taken into account Average number of alternatives found with ALP is 253855 or an average of 7.28 per job. AMP algorithm was able to found a number of 115116 alternatives or an average of 34.23 per job. Recall that in previous set of experiments this numbers were 7.39 and 34.28 alternatives respectively.

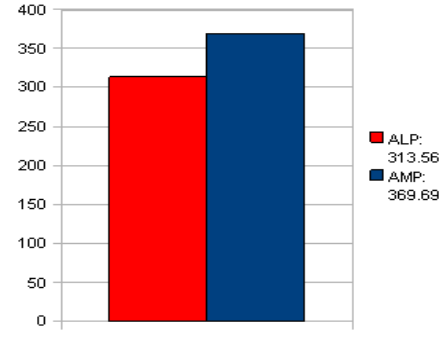


Figure 13: Total Job Execution Cost Using ALP and AMP in TM

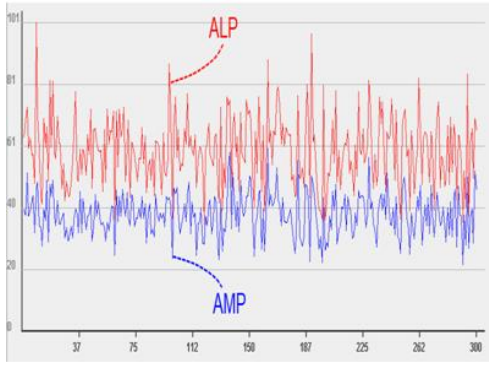


Figure 14: Average job execution time comparison for ALP and AMP in TM for the first 300 experiments

Discussion and Experimental Results

Considering the results of the experiments it can be argued that the use of AMP approach on the stage of alternatives search gives clear advantage compared to the usage of ALP. Advantages are mostly in the large number of alternatives found and consequently in the flexibility of choosing an efficient schedule of batch execution, as well as that AMP provides the job batch execution time less than ALP. AMP allows searching for alternatives among the relatively more expensive processor nodes with higher performance rate. Alternative sets of slots found with ALP are more homogeneous and do not differ much from each other by the values of total execution time and cost. Therefore job batch distributions obtained by optimizations based on various criteria [2] do not differ much from each other either. These factors should explain the results. First, consider the peculiarities of calculating a slot usage total cost $C_t = CtN/P$, where C is a cost of slot usage per time unit, P is a relative performance rate of the processor node on which the slot is allocated, and t is a time span, required by the job (in assumption that the job will be executed on the etalon nodes with $P=1$). In proposed model, generally, the higher the cost C of slot the higher the performance P of node on which this slot is allocated. Hence, the job execution times t/P correspondingly less. So, the high slot cost per time unit is compensated by high performance of the CPU node, so it gets less time to perform the job and less time units to pay for. Thus, in some cases the total execution cost may remain the same even with the more “expensive” slots.

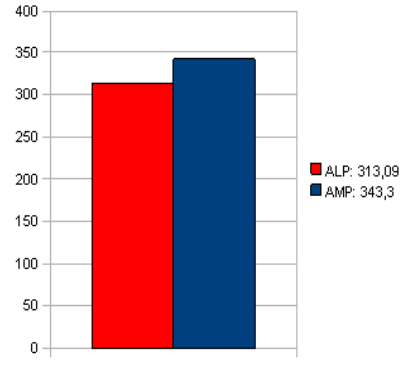


Figure 15: Total job execution cost using ALP and AMP in CM

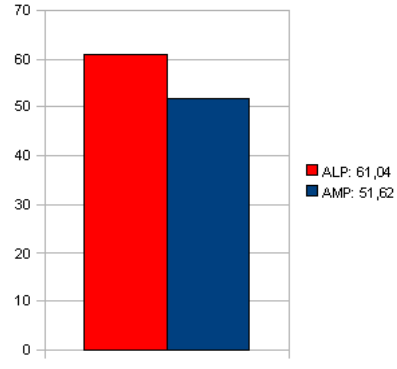


Figure 16: Total job execution time using ALP and AMP in CM

The value C/P is a measure of a slot price/quality ratio. By setting in the resource request the maximum cost C of an individual slot and the minimum performance rate P of a node the user specify the minimum acceptable value of price/quality. The difference between ALP and AMP approaches lies in the fact that ALP searches for alternatives with suitable price/quality coefficient among the slots with usage cost no more than C. AMP performs the search among all the available slots (naturally, both algorithms still have the restriction on the minimum acceptable node performance). This explains why alternatives found with AMP have on average less total execution time. Second, it should be noted that during the search ALP considers available slots regardless of the entire "window". The ALP window consists of slots each of which has the cost value no more than C. At the same time AMP is more flexible. If at some step a slot with cost on δ cheaper than C was added to the desired

window, then AMP algorithm will consider to add slots with cost on the δ more expensive than C on the next steps. Naturally, in this case it will take into account the total cost restriction. That explains, why the average job execution cost is more when using the AMP algorithm, it seeks to use the entire budget to find the earliest suitable alternative.

Another remark concerns the algorithm's work on the same set of slots. It can be argued that any window which could be found with ALP can also be found by AMP. However, there could be windows found with AMP algorithm which cannot be found with a conventional ALP. It is enough to find a window that would contain at least one slot with cost $C_S > C$. This observation once again explains the advantage of AMP approach by a number of alternatives found. The deficiency of AMP scheme is that total batch execution cost on average always higher than the execution cost of the same batch scheduled using ALP algorithm. It is a consequence of a specificity of determining the value of a budget limit and the stage of job batch scheduling [2]. However, it is possible to reduce the job batch total execution cost reducing the user budget limit for every alternative found during the search, which in this experiment was limited to $S = CtN$. This formula can be modified to $S = kCtN$, where k is a positive number less than one, e.g. 0.8. Variation of k allows to obtain flexible distribution schedules on different scheduling cycles, depending on the time of day, resource load level, etc. [2].

Conclusion and Future Work

In this work, we address the problem of independent jobs batch scheduling in heterogeneous environment with non-dedicated resources. The scheduling of the job batch includes two phases. First of all, the independent sets of suitable slots (alternatives of execution) have to be found for every job of the batch. The second phase is selecting the effective combination of alternative slots. The feature of the approach is searching for a number of job alternative executions and consideration of economic policy in VO and financial user requirements on the stage of a single alternative search. For this purpose ALP and AMP approaches for slot search and co-allocation were proposed and considered.

According to the experimental results it can be argued that AMP allows to find at average more rapid alternatives and to perform jobs in a less time. But the total cost of job execution using AMP is relatively higher. When compared to the target optimization criteria during the total batch execution time minimization AMP exceeds ALP significantly. At

the same time during the execution cost minimization the gain of ALP method is negligible. It is worth noting, that on the same set of vacant slots AMP in comparison with ALP finds several time more execution alternatives.

In our future work we will address the problem of slot selection for the whole job batch at once and not for each job consecutively. Therewith it is supposed to optimize the schedule "on the fly" and not to allocate a dedicated phase during each cycle for this optimization. We will research pricing mechanisms that will take into account supply-and-demand trends for computing resources in VO. The necessity of guaranteed job execution at the required quality of service causes taking into account the distributed environment dynamics, namely, changes in the number of jobs for servicing, volumes of computations, possible failures of processor nodes, etc. [13]. As a consequence, in the general case, a set of versions of scheduling, or a strategy, is required instead of a single version [13, 14]. In our further work we will refine resource co-allocation algorithms in order to integrate them with scalable co-scheduling strategies.

Acknowledgement

This work was partially supported by the Council on Grants of the President of the Russian Federation for State Support of Leading Scientific Schools (SS-7239.2010.9), the Russian Foundation for Basic Research (grant no. 09-01-00095), the Analytical Department Target Program "The higher school scientific potential development" (projects nos. 2.1.2/6718, 2.1.2/13283), and by the Federal Target Program "Research and scientific-pedagogical cadres of innovative Russia" (State contracts nos. P2227, 16.740.11.0038, 16.740.11.0516).

References

[1] S. K. Garg, R. Buyya, and H. J. Siegel: Scheduling Parallel Applications on Utility Grids: Time and Cost Trade-off Management, Proceedings of the 32nd Australasian Computer Science Conference (ACSC 2009), Wellington, New Zealand, pp. 151-159 (2009).

[2] V. V. Toporkov, A. Toporkova, A. Tselishchev, D. Yemelyanov, and A. Bobchenkov: Economic Models of Scheduling in Distributed Systems, In: T. Walkowiak, J. Mazurkiewicz, J. Sugier, and W. Zamojski (eds.), Monographs of System Dependability (Vol. 1-3). Dependability of Networks (Vol. 2). – Wroclaw: Oficyna Wydawnicza Politechnki Wroclawskiej, pp. 143-154 (2010).

[3] J. P. Degabriele and D. Pym: Economic Aspects of a Utility Computing Service, Trusted Systems Laboratory HP Laboratories Bristol HPL-2007-101, Technical Report, July 3, pp. 1-23 (2007).

[4] A. Ailamaki, D. Dash, and V. Kantere: Economic Aspects of Cloud Computing, Flash Informatique, Special HPC, , pp. 45-47 (2009).

[5] J. Bredin, D. Kotz, and D. Rus: Economic Markets as a Means of Open Mobile-Agent Systems, Proceedings of the Workshop "Mobile Agents in the Context of Competition and Cooperation (MAC3)", pp. 43-49 (1999).

[6] R. Buyya, D. Abramson, and J. Giddy: Economic Models for Resource Management and Scheduling in Grid Computing, J. of Concurrency and Computation: Practice and Experience, Vol. 14, No. 5, pp. 1507 – 1542, (2002).

[7] C. Ernemann, V. Hamscher, and R. Yahyapour, "Economic Scheduling in Grid Computing", In: Proceedings of the 8th Job Scheduling Strategies for Parallel Processing, D.G. Feitelson, L. Rudolph, U. Schwiegelshohn (eds.), Springer, Heidelberg, LNCS, vol. 2537, 2002, pp. 128-152.

[8] K. Kurowski, J. Nabrzyski, A. Oleksiak, and J. Weglarz, "Multicriteria Aspects of Grid Resource Management", In: Grid resource management. State of the art and future trends, J. Nabrzyski, J.M. Schopf, and J.Weglarz (eds.): Kluwer Academic Publishers, 2003, pp. 271-293.

[9] V. Toporkov: Application-level and Job-flow Scheduling: an Approach for Achieving Quality of Service in Distributed Computing, Proceedings of the 10th International Conference on Parallel Computing Technologies, Springer, Heidelberg, LNCS, Vol. 5698, pp. 350-359 (2009).

[10] V. V. Toporkov: Job and Application-Level Scheduling in Distributed Computing, Ubiquitous Computing and Communication (UbiCC) Journal. Special Issue on ICIT 2009 Conference (selected papers) - Applied Computing, Vol. 4, No. 3, pp. 559-570 (2009).

[11] A. W. Mu'alem and D. G. Feitelson: Utilization, Predictability, Workloads, and User Runtime Estimates in Scheduling the IBM SP2 with Backfilling, IEEE Transactions on Parallel and Distributed Systems, Vol. 12, No. 6, pp. 529-543 (2001).

[12] D. Jackson, Q. Snell, and M. Clement: Core Algorithms of the Maui Scheduler, Springer, Heidelberg, LNCS, Vol. 2221, pp. 87-102 (2001).

[13] V. V. Toporkov and A. Tselishchev: Safety Scheduling Strategies in Distributed Computing, International Journal of Critical Computer-Based Systems, Vol. 1. No. 1/2/3, pp. 41-58 (2010).

[14] V. V. Toporkov, A. Toporkova, A. Tselishchev, and D. Yemelyanov: Scalable Co-Scheduling Strategies in Distributed Computing, Proceedings of the 2010 ACS/IEEE International Conference on Computer Systems and Applications, Hammamet, Tunisia, May 16-19th, 2010. IEEE CS Press, ISBN: 978-1-4244-7715-9 (2010).

Chapter 2

APPLICATION OF PARALLELCOMPUTING TO PRACTICAL COMPUTER EDUCATION

Keiichi Shiraishi, Yoshiro Imai

Summary

Internet and Cloud computing will provide several kinds of services including parallel computing. Parallel computing can be achieved through a computer cluster (Internet itself) and special multiple computer/processor systems. Users of Internet may be more familiar to obtain such parallel computing services in consequences. This chapter focuses on parallel computing, evaluation of its performance and two kinds of parallel programming examples based on master-worker paradigm. Such programming can be realized and executed on CPUs with multiple processor cores. We have compared two types of parallel programming. One is programming for a computer algebra system called Risa/Asir, which is executed on a personal computer of the specific CPU with 4 processor cores. Another is for Cell B.E. with SPE library on Sony-produced PLAYSTATION3. Through real experiment and simple evaluation, the following evidences have been confirmed. Namely, both of two cases are good practices for parallel calculation and in dependence on Amdahl's law and expected performance. And the former is suitable for symbolic and/or algebraic computation, while the latter is efficient for numerical calculation.

Introduction

Parallel computing becomes more and more popular as various kinds of multiple computer/ processor system are available in several engineering fields. Especially, there are rapid and drastic changes even for end users to utilize a computer cluster, multiple processor cores and other multiple computer systems through the Internet. Cloud computing will allow such users to have frequent opportunities to challenge for parallel computing and programming. Computers have made revolutionary progress by means of memory

technology, large-capacity magnetic Disks, and especially powerful microprocessor architecture. And several kinds of computers are considered to be a set of commodity hardware components so that network technologies can easily bring them into the next stage of cluster computing. Some people say that it must become an important task to utilize multiple processors effectively and ecologically in order to obtain fruitful results from the Internet and Cloud computing.

In a focus of microprocessors themselves, there were some complicated limitations for speedup based on uniprocessor architecture. The latest trends of speedup mechanism can be shifted into the architecture of multiple processor cores for high performance computing, namely parallel computing. Such architecture becomes a great idea and breakthrough to improve computing capability. But there are some problems to be resolved in order to achieve efficient programming for parallel computing environment. In other words, it is necessary for users to combine effectively multiple levels of parallel programming, namely not only cluster computing for multiple servers but also parallel programming for multiple processor cores. Education must play an important role to provide a good practice to teach even beginners to obtain useful skills in a short period. In higher education for engineering, it will be much more important than past to educate parallel computing, especially, parallel programming paradigm and its applicable examples for some specific computers with multiple processor cores.

This chapter will focus on providing such a good practice for parallel programming by means of multiple processor cores. In order to show a simple but useful example or methodology for parallel programming education, the chapter describes the following four sections. The next section introduces background of parallel computing and its formulations. The third section explains and compares two types of parallel programming examples with different architectures for multiple processor cores. The fourth section illustrates computing performance for such different programming styles and their evaluation of parallel programming effects. The fifth section mentions our application to practical computer education through parallel programming exercise. Finally, the last section summarizes our conclusions.

Background of Parallel Computing

This section introduces Amdahl's law in the first half and master-worker programming paradigm in the second half as background of our parallel computing. The former is sometimes used with one of limit

arguments to assume that parallel computation may be inefficient, while the latter used to be a contribution to efficient parallel programming.

Amdhal's Law

There are several types of parallel computers, such as a computer cluster, a vector computer, multiple processor cores and so on. For example, Flynn's Classical Taxonomy classifies that a vector computer is classified as SIMD (single instruction, multiple data-stream) computer's category while a computer cluster is done as MIMD one[3]. The former computes vector data, and the latter binds many computers by means of network. Amdahl's law has focused on description of performance for not only vector computers but also parallel computers. And it clearly shows that speed up ratio will be larger if the size of parallelized implementations (namely portion of program to be processed in parallel) grows more. Its characteristics and behavior are simply expressed in the approximate relationship described below [**Error! Reference source not found.**, 3].

It is assumed that Ts is the computing time to run the program on a scalar computer (uniprocessor), while Tp is the computing time to run on a parallel computer (multiple processor cores). α is the parallelization ratio, i.e. the ratio of parallelizable section to program. And finally n is the number of processors. In the Amdahl's law, coefficient P for speed up can be expressed in Eq. (1).

$$P = \frac{Ts}{Tp} = \frac{1}{(1-\alpha) + \alpha / n} \qquad (1)$$

Such a law is well known as a famous formula for determining the maximum expected speedup of an overall system when only part of such a system is improved. It is often used in parallel computing in order to predict the theoretical maximum speed up using multiple processors. But it has been discussed whether it must be treated as an impulsive force or as a regulative rule; namely optimistic view or pessimistic one [6, 7].

Master-Worker Paradigm

Users of parallel computing need more fruitful methodologies in order to perform efficient parallel calculation instead of the above discussion. Master-worker paradigm can play an important role for parallel programming practices. It is known as master-slave approach. And it is a kind of parallel programming model, consists of a master process and worker ones. The master process controls all of tasks and throws the

according tasks to its worker processes under command of it. It also receives the results of the tasks from its workers, combines such results and then produces the final result for the given problem [5]. Each worker receives its assigned task from its above master and runs (i.e. computes) such a task. The worker finishes its own task and returns its result to its master individually and asynchronously. Figure 1 illustrates an example case of master-worker paradigm where one master (process) controls three or more workers (processes) under command of it.

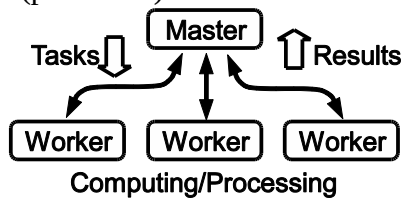


Figure 1: Example of Master-Worker Paradigm.

In master-worker paradigm, if the processing times allocated to all the workers are the same size i.e. the identical time-length, an expected speed up might obey the Amdahl's law. It is one of the most ideal cases for parallel computing. By the way, if such processing times are different one another, the speed up might depend on the order of tasks to be processed. In the other words, it is necessary to discuss how such a suitable schedule is carried out through parallel computing. If there is a smarter scheduling available, parallel computing becomes scalable, more efficient and asynchronously executable in more expected status for parallelization. We will discuss and describe such a scheduling procedure together with some examples of parallel programming in the next section.

Practical Parallel Programming

At first, discussion of scheduling is carried out in the half of this section. And then practical parallel programming is illustrated at the next subsection.

Descriptions and Discussion of Scheduling

It is very much difficult to realize so-called optimal scheduling. So there cannot be a general-purpose scheduling algorithm to build the most suitable procedure to assign the equivalent processing time to each processing unit. Namely there must be no royal road to get semi-optimal algorithm without time-consuming trial-and-error methods. But we have an experience to utilize and evaluate some special scheduling algorithm that “excludes any conscious idle periods” in parallel calculation [5]. The

scheduling only lets the according processors be idling whenever there are no jobs to be executed. Our heuristic results taught that such a scheduling algorithm could do users a favor of fairly shortest processing time, i.e. the algorithm allows us to achieve semi-optimal scheduling. The theorem described below illustrates how such a schedule works correctly [2]. It is assumed that the set of jobs is given, ω is the total elapsed time to finish the jobs, the order of running jobs obeys the schedule and ω_0 is the shortest time which is running and also obeying the optimal schedule. The schedule has been realized by means of a trial-and-error method. The slowness, namely the ratio of ω to ω_0, can be expressed in the following inequality Eq. (2) shown below;

$$\frac{\omega}{\omega_0} \leq 2 - \frac{1}{n} \qquad (2)$$

Where n denotes the number of processors. It seems that the right side expression of Eq. (2) can also give the upper bound of the above ratio, namely the left side of the inequality.

Parallel Programming on CPUS with Multiple Processor Cores

This subsection illustrates two types of parallel programming examples. One is a programming example for Risa/Asir1 on the CPU with 4 processor cores (Xeon 2.13GHz). And another is for CELL B.E. of PLAYSTATION 3. In the former case, programs must be written in a special programming language for Risa/Asir. It is an open source general computer algebra system. OpenXM committee has been developing its Kobe distribution. The original Risa/Asir was developed at Fujitsu Labs LTD. during Japan national project of the fifth generation computer. In the latter case, programs are to be written in the C programming language for the Cell Multiprocessor [4]. These are explained in detail as follows.

Case (I): A Programming Example for Risa/ Asir

Parallel programs for the computer algebra system called Risa/Asir are to be described as follows;

1. Invoke an "asir" process and assign it to one of workers.
2. Load a program code to the according worker process.
3. Execute such a program on the process.
4. Wait for termination of each worker's program and receive the

[1] http://www.math.kobe-u.ac.jp/Asir/asir.html

computing result.

Process assignment is automatically performed by means of load balancing facility of the operating system. Namely users do not need to be aware of such a complicated operation. The computer algebra system Risa/Asir does not request users to revoke (destroy) each worker process. A master process can investigate whether an according worker process terminate or not, so it is very efficient for users to write a good parallel program to reassign a next task onto vacant (idling) worker. A sample program for Risa/Asir is shown in the Appendix A-1 at the end of this chapter

Case (II): A Programming Example for Cell Multiprocessor

Parallel programs for Cell B.E. with SPE library (ver2) on Sony-produced PLAYSTATION3 are to be described as follows;

1. Load a program code for worker process onto memory.
2. Allocate each code onto the according SPE(Synergistic Processor Element)
3. Invoke a thread and give an instruction to execute a program for worker in the thread.
4. Wait for termination of each thread which finishes when the according worker program is terminated.
5. Revoke each worker process.

Users must utilize each worker by means of an invocation of thread. This is why a function gives an instruction to execute programs for workers but it does not send back its return values until programs for workers terminate. In the Cell Multiprocessor, SPE can play a role to perform data passing by means of DMA transfer. So programs for workers, which are running in SPEs, can move several kinds of data from the main memory to local storage by means of DMA operation. After they has processed those data, they can restore data in reverse, namely from their local storages to the main memory, through DMA channel. A sample program for CELL B.E. with SPE library is shown in the Appendix A-2.

Evaluation of Parallel Calculation

This section demonstrates computing performance and its performance evaluation relevant to two different programming styles described above.

Parallel Calculation of Circle Ratio

This subsection describes our practical parallel programming and shows a real result for execution of parallel programs based on master-worker

paradigm. We know that a value of the circle ratio can be computed by means of the following definite integration Eq.(3).

$$\pi = \int_0^1 \frac{4}{1+x^2}\,dx \qquad (3)$$

The real numerical value will be obtained through calculation of the equation (3) by the way of numerical integration based on the following approximation Eq. (4),

$$\pi \approx \frac{1}{N}\sum_{i=0}^{N-1} \frac{4}{1+((i+0.5)/N)^2} \qquad (4)$$

Where; *N* is a division number for integration interval [0, 1] of the equation (3). Users had better choose a larger number as *N* in order to obtain a more precise result for the above numerical integration. Such a calculation can be realized with description of '*i*' related loop operation, namely iteration procedure about control variable of '*i*'. We can achieve practical parallel programming by means of application of each loop operation into the worker process which has been explained in the previous section. In the other words, such a situation is suitable enough to satisfy Amdahl's law with efficient parallelization ratio. We will expect to achieve good speedup-improvement for parallelization in the case of parallel calculation for numerical integration of circle ratio based on master-worker paradigm.

In order to make certain of speedup improvement for calculation through parallel programming, we have applied N=50,000,000 to numerical integration of circle ratio expressed in the approximation Eq. (4). At first, we have computed it with the computer algebra system Risa/ Asir. Our test-bed PC has a CPU chip with 4 processor cores so that we can select parallel programming with the master-worker paradigm based parallel calculation where we can perform assignment of computing task from one worker process assignment case to four. Table 1 shows result of elapsed time for the above parallel calculation. Secondly, we have computed the approximation Eq. (4) with PLAYSTATION3. The Cell B.E. of PLAYSTATION3 has seven sets of SPEs, but SPE library of it can handle only six sets of SPEs. So we can select the same (but not identical) way of parallel programming based on master-worker paradigm from one worker assignment case to six. Table 2 shows another result of elapsed time for such parallel calculation with at most six sets of SPEs.

Table 1: Elapsed time for parallel calculation of circle ratio by the computer algebra system (Risa/Asir)

Numbers of worker process	Elapsed time for calculation (second)	Speedup (Improvement) for only one worker's case
1	37.902	1.000
2	19.627	1.931
3	13.787	2.749
4	11.056	3.428

Table 2: Elapsed time for parallel calculation of circle ratio by Cell B.E. with SPE library on Sony-produced PLAYSTATION3

Numbers of worker process	Elapsed time for calculation (second)	Speedup (Improvement) for only one worker's case
1	6.037	1.000
2	3.023	1.997
3	2.019	2.990
4	1.518	3.978
5	1.217	4.959
6	1.017	5.934

Performance Evaluation of Parallel Calculation

This subsection tries to evaluate two types of execution performance for parallel calculation of circle ratio such as elapsed time of execution and Speedup effect based on parallel programming. An elapsed time for Parallel Calculation is focused and discussed between the case of Risa/Asir and one of PLAYSTATION3. The elapsed times for the CPU

with 4 processor cores with Risa/Asir are clearly larger than ones for the Cell B.E. with SPE library on PLAYSTATION3. Figure 2 illustrates the above discussion. It is estimated that one of such reasons can be based on their description languages. Risa/Asir seems not to be more suitable for numerical calculation than C programming language.

"Speedup" based on parallel programming, namely "Improvement" of execution time by multiprocessors, is then focused and discussed just like the above. In this case, it is almost no effect from their description language. Speedup may be achieved by usage of multiprocessors, and each increasing rate can be expressed to be almost linear.

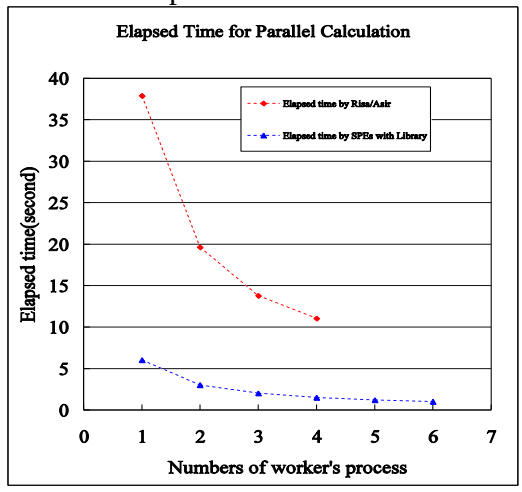


Figure 2: Elapsed time for Parallel Calculation of circle ratio in CPU with 4 processor cores with Risa/ Asir and in the Cell B.E. with SPE library on PLAYSTATION3

Figure 3 shows efficient parallel programming in both of the above cases. Therefore, Speedup from parallel programming, namely increasing rate of execution time by usage of multiprocessors, for the CPU with 4 processor cores with Risa/Asir is relatively similar with one for the Cell B.E. with SPE library on PLAYSTATION3.From comparison of Figure 2 andFigure 3, it is confirmed that elapsed times of calculation by the computer algebra system Risa/Asir are fairly larger than ones by the Cell B.E. with SPE library on PLAYSTATION3. So the former is suitable for symbolic and/or algebraic computation, while the latter is efficient for numerical calculation.

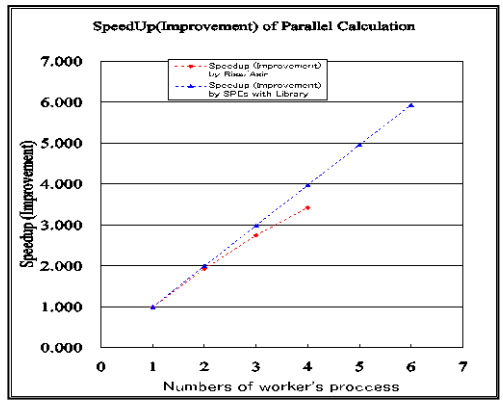


Figure 3: Speedup (i.e. Improvement) of Parallel Calculation of circle ratio in the CPU with 4 processor cores with Risa/Asir and in the Cell B.E. with SPE library on PLAYSTATION3

On the other hand, it is confirmed that Speedup from parallel computing can be achieved in the both cases and Improvement through usage of multiprocessors may not come under the influence of their description languages.

Application to Practical Computer Education

In order to investigate educational effects of parallel programming, we introduce our parallel programming example as teaching materials for engineering education. The detail of our lecture and exercise is as follows;

1. A lecture for 10 students of the third semester of faculty of Engineering in university.
2. At first, short explanation about parallel programming.
3. The following 3 questions before our exercise;
   - Can you explain Parallel computing ? Q(1)
   - Can you write some kind of Parallel program ? Q(2)
   - Can you evaluate Parallel programming in a theoretical way ? Q(3)
4. Secondly, parallel programming exercise with some practical examples described in the above section (about two weeks, 3 hours for a week).
5. The same 3 questions (once again) after our exercise.

The results of answers of the 3 questions from 10 students are as follows. Table 3 shows the answers “before” our exercise and Table 4 shows ones “after” our exercise.

Table 3: Questions and their Answers before Parallel Programming Exercise

	Of course!	Maybe	I'm afraid
Q(1)	1	6	3
Q(2)	0	2	8
Q(3)	0	2	8

Table 4: Questions and their Answers after Parallel Programming Exercise

	Of course!	Maybe	I'm afraid
Q(1)	6	3	1
Q(2)	4	3	3
Q(3)	3	3	4

One student said, "I was afraid of explain parallel programming with example before this lecture, but now I am not. Maybe I can explain suitably." Others clearly said, "We are willing to write parallel programs in fact!" Of course, this time test for only 10 students may be neither suitable nor satisfactory testing group for making conclusion (Also reviewers of our conference have pointed out). They have commented that knowledge about parallel programming must be tested through more than three questions. We should design more precise and detail questions and prepare for investigation. These will be our future problems to be resolved.

Parallel Calculation of EXP(X)

This section describes parallel calculation of exp(*x*) to develop teaching material because it is more difficult than the calculation of circle ratio and would be suitable for the next problem. Exp(*x*) can be expanded as a Maclaurin series and approximated with *N* terms. If the approximation is parallelized with Master-worker paradigm, the program will be slower than the sequential one because computation of factorial needs a long time. If you use Stirling's approximation, you should compute (*n*/*e*) to *n*th power and will needs a long time. Eq. (5) shows the idea of parallel computation of exp(*x*), where $L = \left\lceil \frac{N+1}{M} \right\rceil$, *M* is the number of workers. Eq. (5) illustrates the terms are divided to *M* groups that are computed by each worker and the last term of each group appears in the all following groups as coefficient, for example, $\frac{x^{L-1}}{(L-1)!}$ appears in the last term of the

first group and another groups' coefficient, $\frac{x^L}{(2L-1)!/(L-1)!}$ appears in the last term of the second group and another groups' coefficient except first group. So, if the worker process computes the summation of the terms in parentheses and the last term and return them and the master process gather the results in order specified by Eq. (5), the computation time will be short. A sample program for Risa/Asir is shown in the Appendix A-3 at the end of this chapter. **Error! Reference source not found.**, Figure 4 shows an elapsed time and speedup for parallel calculation of exp (1), N=1000 by the computer algebra system Risa/Asir on the CPU with 4 processor cores (Core i7, 2.8GHz).

$$
\begin{aligned}
\exp(x) &\approx \sum_{n=0}^{N} \frac{x^n}{n!} \\
&= \frac{x^0}{0!} + \frac{x^1}{1!} + \cdots + \frac{x^N}{N!} \\
&= \left(\frac{x^0}{0!} + \frac{x^1}{1!} + \cdots + \frac{x^{L-1}}{(L-1)!} \right) \\
&+ \frac{x^{L-1}}{(L-1)!} \\
&\cdot \left(\frac{x^1}{L} + \frac{x^2}{L(L+1)} + \cdots + \frac{x^L}{(2L-1)!/(L-1)!} \right) \\
&+ \frac{x^{L-1}}{(L-1)!} \cdot \frac{x^L}{(2L-1)!/(L-1)!} \\
&\cdot \left(\frac{x^1}{2L} + \frac{x^2}{2L(2L+1)} + \cdots \right. \\
&\left. + \frac{x^L}{(3L-1)!/(2L-1)!} \right) \\
&+ \cdots \\
&+ \frac{x^{L-1}}{(L-1)!} \cdot \frac{x^L}{(2L-1)!/(L-1)!} \cdots \\
&\cdot \frac{x^L}{((M-1)L-1)!/((M-2)L-1)!} \\
&\cdot \left(\frac{x^1}{(M-1)L} + \frac{x^2}{((M-1)L)((M-1)L+1)} + \cdots \right. \\
&\left. + \frac{x^{N-(M-1)L+1}}{N!/((M-1)L-1)!} \right)
\end{aligned}
\tag{5}
$$

Table 5: Elapsed time for parallel calculation of exp(1) by the computer algebra system (Risa/Asir)

Numbers of worker process	Elapsed time for calculation (second)	Speedup (Improvement) for only one worker's case
Sequential	6.24	1.00
1	6.42	0.972
2	1.24	5.03
3	0.562	11.1
4	0.336	18.6
5	0.296	21.1
6	0.248	25.2
7	0.254	24.6
8	0.264	23.6

Figure 4 Show that speedup is more than 2 in the case of 2 workers. It is why in the default case Risa/Asir computes the all of numbers as rational numbers and multiple precision numbers. So, if the number of digits will be double, the complexity will be quadruple.

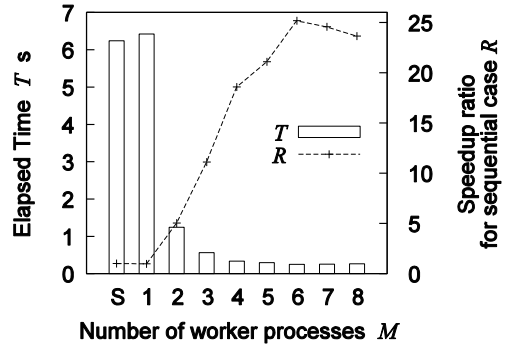


Figure 4: Elapsed time and speedup for Parallel Calculation of exp (1), *N*=1000 in the CPU with 4 processor cores with Risa/Asir. *M*=S means sequential case

Table 6 shows maximum decimal digits of numerators/denominators of return values from worker processes. It shows the number of digits decreases when the number of workers increases. It might be the reason why speedup is more than 2 in the case of 2 workers.

Table 6: Maximum decimal digits of numerators/denominators of return values from worker processes

Number of worker processes *M*	**Digits of return value**
1	2565
2	1437
3	969
4	735
5	594
6	492
7	425
8	375

Conclusion

This chapter has described some samples of parallel programming based on Master-worker Paradigm and their evaluation. We have compared the computer algebra system Risa/Asir to Cell B.E. with SPE library through practical experience. This is why education about parallel programming becomes important as several kinds of multiple computer and processors are available for users by means of the Internet and Cloud computing. An advantage of Risa/Asir is to utilize the same program on multiple processor cores as well as clustered computers. SPE library is a special one for using SPE of Cell B.E. and users need POSIX thread libraries in order to execute multiples of SPEs as "workers" simultaneously. So it is indispensable to realize some kinds of TCP/IP-based communication between each machine in PLAYSTATION3-based cluster system.

With regard to numerical analysis, Risa/Asir cannot be compared to Cell B.E. with SPE library, because it is not designed as efficient software for numerical computation. It is confirmed that Risa/Asir is suitable for symbolic and/or algebraic computation and is not good at high-speed computation. By the way, we have also ascertained the qualitative capability of Risa/Asir to describe parallel programming based on Master-worker Paradigm through our experiment. Through the real lecture with parallel programming exercise, it is confirmed that an application of parallel programming to practical computer education is good for students and it may be suitable for them to build self-confidence by means of writing programming examples.

In the future works, we will begin to investigate much more questions for more numbers of students in order to evaluate our parallel programming education and confirm whether it is really fruitful or not. Additionally, we will try to utilize high-speed computation with 128-bit Vector/ SIMD processing units of SPEs. We have some problems to improve computation performance. Through a real experiment, it is confirmed that one of them is related to realize a smart DMA (Direct Memory Access) between the main memory of Cell B.E. and local memories of SPEs.

Acknowledgement

The authors would like to express sincere thanks to Dr. Ali Al-Dahoud of Al-Zaytoonah University as ICIT2011 Program Chair for his complete supervising, advising and contributing. This study was partly supported by the Special Science Fund from Cmicro Corporation in Japan and also partly supported by Research for Promoting Technological Seeds from Japan Science and Technology Agency (JST).

References

[1] G. Amdahl, "Validity of the single-processor approach to achieving large scale computing capabilities," In AFIPS Conference Proceedings, pp. 483–485, 1967.

[2] C. L. Liu, *Elements of Discrete Mathematics*. 2nd edition, McGraw-Hill, 1985.

[3] J. Hennessy and D. Patterson, *Computer Architecture: A Quantitative Approach*. Morgan Kaufmann, Palo Alto California94303, 1990.

[4] J. A. Kahl, M. N. Day, H. P. Hofstee, C. R. Johns, T. R. Maeurer and D. Shippy, "Introduction to the cell multiprocessor," *IBM Journal of Research and Development*, Vol. 49, No. 4, pp. 589–604, 2005.

[5] B. Wilkinson and M. Allen, *Parallel Programming: Techniques and Applications Using Networked Workstations and Parallel Computers*. 2nd edition, Prentice-Hall, Upper Saddle River NJ07458, USA, 2005.

[6] M. D. Hill and M. R. Marty, "Amdahl's law in the multicore era," *IEEE Computer*, Vol. 41, No. 7, pp. 33–38, 2008.

[7] X.-H. Sun and Y. Chen, "Reevaluating amdahl's law in the multicore era," *Journal of Parallel and Distributed Computing*, Vol. 70, No. 2, pp. 183–188, 2010.

Chapter 3

PROSPECTIVE PROPULSIONS TO EMBED ARTIFICIAL INTELLIGENCE INTO THE E-PORTFOLIO SYSTEMS

Aleksandrs Gorbunovs

Summary

This chapter addresses an importance of the information systems which display and measure person's competences. E-portfolios, as the part of information systems, plays significant role to enhance people's self-confidence and motivate learners to improve their competencies. However, nowadays lifelong learning paradigm requires new operation of thinking and new approach to provide appropriate services to target groups. This study seeks to identify perspective ways to make ePortfolio systems more flexible and intelligent, and also set the objectives to assist learners or job seekers to make odds even in their current competences level and required at the labor market.

Introduction

People have tried to demonstrate to others their skills and achievements for the thousands of years. We can find them in ancient petroglyphs and papyrus rolls, domestic works and art treasures. Masterminds, artists, sculptors and other professionals exposed their imposing accomplishments in paintings galleries, art exhibitions and performances. That might be seen as the germ of the portfolios. Digitalization of society and Internet brought new potentialities in this area. Wider options allowed using the benefits of the portfolios for wider sphere of society, which have flown into e-portfolio systems, impregnating and making richer information systems (ISs). ISs should be considered as the most important tools available to educational managers and specialists and provide higher levels of efficiency in educational and training area, similarly like it applies to managers and their efficiency and productivity in business operations, especially when coupled with changes in business practices and human behavior [1].

For their part, e-portfolio systems become an important part of IS.

Lots of universities and other education institutions (for example, Queensland University of Technology, University of Melbourne, University of New England, University of Wollongong, Clemson University, University of Minnesota, Indiana University-Purdue University Indianapolis, Portland State University, LaGuardia Community College, and many others) use e-portfolios to promote lifelong learning, improve the quality of education and training, and motivate learners to achieve the goals. Some of them just suggest using of e-portfolios during study period, unlike others stipulate active use of e-portfolios and include them as a part of curriculum [2]. Nowadays some employers require from manpower to use e-portfolios to provide evidences of their professional growth. For instance, portfolios as evidence of continued professional development are widely used by nurses and midwifes in Australia and other countries, where is an obligation to self-declare competence and fitness to practice when renewing annual practice certificates [3].

However, students and job seekers sometimes have no needs to attend a university or college and complete their studies of a whole course or program. In fact, if a person completes his/her studies, probably there will be some parts of passed before study materials or subjects already included in a new study program which the student is interested in. Labor market relatively often requires shifting directions of efforts and gaining new qualifications. This is a demand of lifelong learning which in its turn requires persons' readiness to find appropriate courses and complete them. Human resources managers often ask potential employees for the documents which procure a documentary evidence of completed formal education. Informal education and skills as often as not remain unclaimed. However, new initiatives of the European Commission and local governments speak volumes for further steps to recognize informal education and remove obstacles which may delay balanced backing of both formal and informal education and training. For example, in Latvia the special centers for the leveling and attestation of informal education are going to be established in these months. Whole our society, which nowadays is called as the knowledge society, sets a target to people to care about personal and collective developments, be ready to take the challenges and initiative, improve professional skills. That is what we mean under the name of lifelong learning.

Moreover, lifelong learning cannot be compartmentalized into course structures. While lifelong learning is supported by the work done in individual courses, it clearly crosses the boundaries between courses [4].

The problem, which is unsolved yet, is how to organize processing of all data related to person's exiting qualifications and give him/her useful suggestions to obtain an appropriate course or small parts of the courses or subjects. In the light of it, this paper introduces a draft of the new e-portfolio model with some sorts of artificial intelligence (AI) features which will be assigned to solve the abovementioned problems by offering people new services. E-portfolio approach has a number of advantages which are very difficult to provide for traditional classes, such as ability to activate, engage and motivate learners, to favor teaching staff, subject matter experts and advisors to respond and give the answer on the problematical issue, to promote learners of their own. Pursuant to that the new proposed ideas might be discussed to further develop e-portfolio system with AI traits.

Background

Studying available scientific publications and research papers in e-portfolio area, it might be observed that related fundamental in depth studies were not made yet. Although, there are right enough opinions, statements, essays and publications about usability and developments of the e-portfolio systems in individual educational organizations and consortiums. Especially it applies to study results in medicine institutions where it is possible to find so many qualitative materials, suggestions and conclusions, probably, due to specific learning and job requirements for the medical students, medics, and paramedics.

Research activities in education supporting e-portfolio technology area have been going on for more than ten years. Main aim of these activities is to align the technical, managerial and educational developments of the e-studies. Till now the course of e-portfolio development in the world is typically fragmented by diverse tools, functions and implementation levels. Though conceptually e-portfolios may unify support to human resources and lifelong learning development in the world, majority of educational, research and technological institutions endeavor to define their own conceptions and tools. For years portfolios and e-portfolios were used to demonstrate person's achievements, competencies and interests. Main aim of existing e-portfolio systems is to ensure for a user to display his/her previous developments, study and work results and progress. Some others offer users some sorts of interactivity, ability for a student to communicate to others: students, tutors, teaching staff, subject matter experts, and advisers. Some existing e-portfolio systems allows students to make reflections on critical thinking

notes of other students and review own work

To illustrate abovementioned notes in previous paragraph it would be easy to look at the simplest e-portfolio systems or even tools which are used to create offline an evidence document or an image and put it onto Web. Thus, it possible to use, for instance, MS Office package, work with Word or develop presentations in Power Point software about your achievements or things which you are proud of or would like to share to others, and afterwards find appropriate server put them all. Web based applications also offer you right enough ways to speak about yourself in global Web. Next, higher level of e-portfolio systems are represented by offering to users to communicate to each other, exchange with information, work in teams and groups, influence classmates' accomplishments. Reflection, critical thinking process here is mostly accepted. Wikis, GoogleDocs, Web 2.0 and other instruments are very suitable and pretty good for these tasks. The Distance Education Study Centre of the Riga Technical University has good examples of implementation of such tools in students' e-learning process by trying of a new method of the creation and acquisition of knowledge - educational action research in the e-learning group. The method was implemented in Business planning course for the bachelor level students at the Riga Technical University [5].

Some of modern e-portfolio systems ensure interactivity, data management and reporting system for assessment (for example, FolioTek, Task Stream, and others). Mostly, these systems are commercial products and provided as the hosted services. Subscription rates may stand some learners' hands to use such systems. Significant efforts to improve e-portfolio systems and add some extra tools were done by TENCompetence Foundation from December 2005 to December 2009. The Personal Competence Manager (PCM) domain were developed to support learners activities, stimulate their own learning, chose learning route, provide structured, multi-leveled access to learning resources, based upon competences, and ensure communication environment allowing users at each level to discuss, critically think and reflect on these notes [6]. To ensure abovementioned capabilities the various tools both learner and author ones, such as My E-portfolio, The Graphical Planning Tool, User-Profile Editor, and others, were developed [7].

Valuation of prior learning and obtained skills is sometimes observed at human resource management tools embedded in the organization's IS. To verify potential job seeker's eligibility, they ask the candidate to make

a self-assessment by marking his/her competencies/skills in the corresponding boxes. Then the IS's tool checks the data – sets these results against required and gives the suggestions regarding eligibility and appropriate competencies/skills still needed to pretend to the position. What the candidate wants to expect – to get suggestions for further competences development in the way of study materials, links, steps to be taken to obtain related competence, and list of educational institutions – he/she does not receive.

Purposes of e-portfolios vary depending on aims but majority of them have points as follows [8]:

- Document learning.
- Demonstrate accomplishment of specified competencies or criteria.
- Identify strengths and areas for improvement.
- Showcase a career trajectory.
- Demonstrate how learning connects to practice outcomes.
- Plan transition into practice.
- Award financial merit.
- Empower students to take responsibility for their own learning.
- Provide a broader view of learning over time by linking experiences and skills to professional competency.
- Demonstrate knowledge that can be converted to academic credit.
- Serve as a starting point for planning and individualizing learning during orientation.
- Determine advanced placement or shortened orientation time and faster transition to the work force.
- Promote transition from academic to service setting by reviewing the new graduate's documented competencies and tailoring orientation programmers' accordingly.

Pursuant to this overview, despite on variety of e-portfolio systems available at the moment, there are not enough sophisticated methods and e-portfolio systems which would consider user's existing competencies, on-going learning or practice activities, and make suggestions to user how to improve theses competences. Brief overview of current e-portfolio systems and their characteristics will be given in the next sections of this chapter

Chapter 3

Literature Survey

Educational Organization's Information System and Facets of Data Flow

To start considering the role and functions of the educational institution's IS we should bear in mind that the IS is a part and parcel of an organization. Organizations, as a rule, have more or less similar goals, but different structures. Namely, ISs have common and different functions depending on organization's goals. With reference to Keneth C. and Jane Price Laudons [1] who have described functions of IS, it is possible to adopt their model to IS of educational organizations by adjusting it as it is represented in Fig. 1. Three basic activities – input, processing, and output – produce the organizations' information need [1]. An important data flow of organization's IS must be marked by the feedback as the output returned to appropriate persons or activities in the organization to evaluate and refine the input. The processing activity collects, analyses, classifies, arranges, and calculates data available from the input activity and interaction with internal and external environmental bodies which in the case of educational organization's IS are:

1. Students. They are the main target audience of IS and teachers' efforts. Undoubtedly, information and data given by them in form of essays, tests passed, evidences of achievements, study and practice results, critical thinking notes and so on play a lead of data flow in organization's IS.
2. Teaching staff tutors. They serve as the mentoring and advisory board to ensure the educational process. Collaboration within IS between students and tutors, as well as between students in their groups and peers are essential to direct the study process in right way.
3. Partners. They could be recognized as collaboration organizations: both educational and non-educational. Universities can establish consortiums to synergy their efforts, unite separate ISs, whole or part of them, into a common IS, share knowledge, study materials, and study support, organize collaborative data flow in common part of IS through e-collaboration tools and technologies such as TV-videoconferencing, e-mailing, Web 2.0, etc.
4. Competitors. In case of educational organizations' IS there are numbers of evidences showing competitiveness among educational institutions to attract more students. Governmental schools compete for image, prestige. For private schools this is mainly a question about survival, although image issues remain also actual ones. Curiously, but

at the same time it helps them to improve educational results in emulation with each other and finding innovative approaches.

5. Customers. First of all, that is labor market which consists of: employers, business, governmental and nongovernmental organizations (GOs and NGOs) and prioritizes educational directions and proportion of graduates. Secondly, that is business and variety of project activities connected with business and GOs and NGOs goals.
6. Suppliers. They differ as many as their services vary. This list may include such supply services as hardware and software delivery, electricity, water supply, etc.

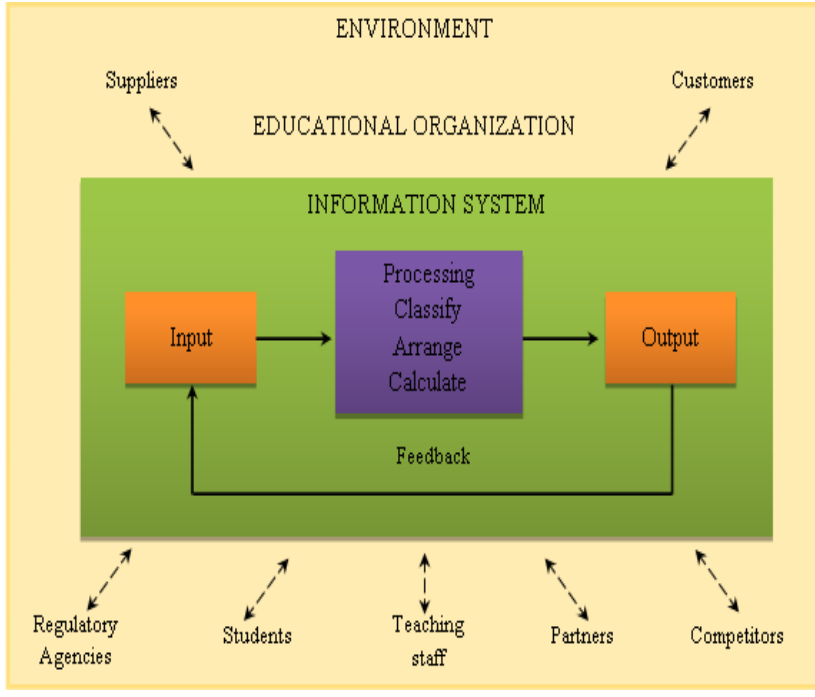


Figure 1: Functions of Educational Organization's IS

7. Regulatory agencies. No doubt, all educational institutions should obey appropriate rules established by State Ministry of Education and Science, State laws and regulations (for European Union (EU) countries – also appropriate EU regulations).

In my proposed new design e-portfolio system model with AI features regulatory agencies also should be recognized as the very important IS environmental actors to ensure direct rapid data interaction between

regulatory agencies and user's e-portfolio which will be outlined in further chapters.

E-Portfolio System as a Part of Information System

The world and even universe must be considered as the gigantic IS that consists of huge amount of ISs some of which are joint in bigger or smaller associations. At the same time, these ISs consist of sub-ISs, a number of which depends on organization's goals and needs.

There are number of e-portfolio definitions. They vary from the simplest to more detailed ones. If Redman, for example, in simply words has defined e-portfolio as the person's data records about his/her accomplishments [9], then Brown has given broader formulation saying that portfolio is an acknowledgement of the evidences and demonstration of the person's knowledge and skills development [10]. Jasper has proposed another statement which is used more than others, emphasizing that e-portfolios are personal collection of evidences which demonstrates continuous acquisition of the person's knowledge, skills, attitudes, comprehension and achievements, and applies on all person's achievements, both past, currant and next ones [11].

Despite on various definitions a person's individual e-portfolio should be considered as an integral part of an educational organization's IS. It is aimed to collect and process data of student's achievements and work-based results, prepared in electronic form, which includes evidences of person's professional competencies or learning outcomes, personal reflection, contemplation and action, work or study activities and progress, knowledge development, proof of accomplishment and ability make achievements, collaborate, communicate, think critically, analyze, and invariably improve themselves [2]. In scientific world a battle of the books still continues about purposes of e-portfolios. Whilst there are some opinions that the only purpose of e-portfolios is assessment, others insist on multiple purposes for developing e-portfolios such as assessment, showcasing skills for potential employment, personal development planning (PDP), reflection on learning, etc. [12].

As reflection is the "heart and soul" of an e-portfolio, and is essential to brain-based learning [13], [14], there is a need to develop appropriate strategies that better support reflection in the learning process, supporting different types of reflection to improve learning [15]. Using reflective processes to learn with and from others, students create opportunities to enhance their interpersonal relationships and gain multiple perspectives. Obtaining new knowledge, students learn about others and themselves in ways that enable them to critically reflect on and critique their

experiences and examine what shaped their perspectives [16].

From lifelong learning prospective an e-portfolio should be developed as an aggregate system to ensure students' learning and new qualification achieving efforts. Thus, there must be place for PDP, reflection on learning, collaboration with classmates and tutors, assessment, presentation of themselves for employers and teaching staff from other educational organizations. One thing which we ought to keep in mind building IS and e-portfolios as IS subsystems, is the security of IS. It occurs at four levels to be effective [17], namely:

a. Physical,
b. Human,
c. Operating system, and
d. Network.

The author is especially concerned about level (b) which besides other issues includes social engineering and social networking because social interaction and collaboration between students and teaching staff in e-portfolio systems is of great importance. Also we ought to keep in mind that security certainly is as weak as the weakest chain [17]. Unless some categories of e-portfolio systems do not support and even do not allow interactivity, others have necessary collaboration Web 2.0 tools. In line with Dr.H.Barett's research, there are six categories of e-portfolios tools from which first three belongs to individual and institutional tools and last three – to institutional ones [18]:

1. Authoring Tools (such as Mozilla Composer, Dreamweaver, Front Page, Apple's iWeb, MS Office and Open Office Word, PowerPoint, Adobe Acrobat, iMovie, etc.). They are used for offline portfolios development with further placement onto the Web server or portable memory discs and do not provide interactivity;
2. Static Web Services (such as GeoCities, GooglePages, Tripod, etc.). Organizations and individual persons can use these services to create and publish a presentation portfolio. These web services provide little or no interactivity (Web 1.0);
3. Interactive Web Services (such as WikiSpaces, GoogleDocs, EduSpaces (Elgg), etc.). As dynamic web services they might be used for creating and publishing of organization's or person's presentation portfolio. They allow interactivity;
4. Software – Server required (such as Blackboard, MS SharePoint, Open Source Content Management Systems (Drupal, Plone), Open Source tools (Elgg, Mahara, OSPI, ePEARL), Embedded in Moodle (Moofolio, MyStaff), etc.). Organizations can install them on their

own servers and provide space for persons' e-portfolios. They allow interactivity but do not provide data management;

5. Hosted Services (such as GoogleAps for Education, Digication, Epsilen, My eCoach, etc.). They can be adopted by organizations. Thus, there is no necessity to use organization's server. Services allow interactivity but do not support data management and reporting system;
6. Assessment Systems - Hosted Services (such as Chalk & Wire, College LiveText, FolioTek, mVentive'sTracDat, Richer Picture, and Task Stream). These are systems which, similarly like services in previous category, are hosted. Thus, organizations would adopt them that will allow hosting e-portfolios, ensure interactivity, data management and reporting system for assessment.

In spite of a great number of e-portfolio systems represented in abovementioned categories there is no any e-portfolio system with artificial intelligence features which could be used for creative selective learning based on person's existing competencies to meet lifelong learning requirements.

Materials, Tools and Methods

New Design e-Portfolio System with Artificial Intelligence Features

To provide new approach to learning, minimize time necessary to find appropriate study courses and complete them, it is necessary to enrich existing e-portfolio systems by new interactions, links, data bases, data processing, services, regulations, and standards. The model of the new design e-portfolio system with AI features (Fig. 2) consists of several IS's environmental actors such as individual e-portfolio owners/users (students), teaching staff of an educational organization, classmates, formal and informal educational institutions (e.g. accredited examination centers), ministries, human resources management organizations, business structures, joint data base holder, other key players, which all are tied with links to ensure their interactions.

A student displays in his/her individual e-portfolio personal profile, study and practice results, and achievements. This information might be available for business and educational organizations, GOs and NGOs. Study support, tutors assistance and classmates collaboration, teamwork and critical thinking, as well student's own critical approach to his/her studies and achievements play essential role in student's growth. Students

ought to interact actively with classmates: put forward remarks, notes and proposals to others about their e-portfolios and displayed information, make necessary assessments of peers work as a part of professors' led assessment when it is required.

Students are encouraged not only to think critically but also to make judgments about their own development and performance [19], thus allowing them to identify their own strengths and weaknesses [20]. Reflection and critical thinking are effective tools to develop students' knowledge using e-portfolio [2]. The individual approach to reflection, which is ultimately about personal growth and development, is limited in its social contextual consideration and also lacks an overt action perspective [21]. Reflection ought to be shown during whole study period instead of students' sometimes short flash activities during examination period or because of requirements of training program [2]. Moreover, reflection in optimum ought to be appeared both immediate (through Web 2.0 tools, discussion forums, blog entries, as well in form which might not appeared in e-portfolio directly, for instance, via e-mailings and TV/Web-conferencing) and retrospectively when old data are added, corrected and improved all the time.

All this must be rendered in an individual e-portfolio. A new featured e-portfolio should be tied with links to in entirety of data bases which may become a totally new joint data base (JDB). JDB will include records of rules related to education, occupational standards, obliged and attested equal for everybody credit system, curriculums, and regulations for leveling.

JDB will actively interact with e-portfolio users, education organizations, labor market, business organizations, GOs and NGOs, Ministry of Education and Science, other ministries and agencies involved in educational issues. JDB can start as a local, state data base, expanding its capacity at European level and even incorporating all leading countries of the world in sphere of education and training.

JDB will be tailored to an advisory purpose to give students or other individual e-portfolio owners' suggestions to choose appropriate educational courses or their parts to get a new qualification. This will be achieved by following activities. Student's personal data and study results (gained qualification), placed in his/her e-portfolio, will be set against data in JDB. Based on a new qualification request, JDB will match available educational or training courses and give immediate response via e-portfolio to the student about necessary educational courses to satisfy his/her needs.

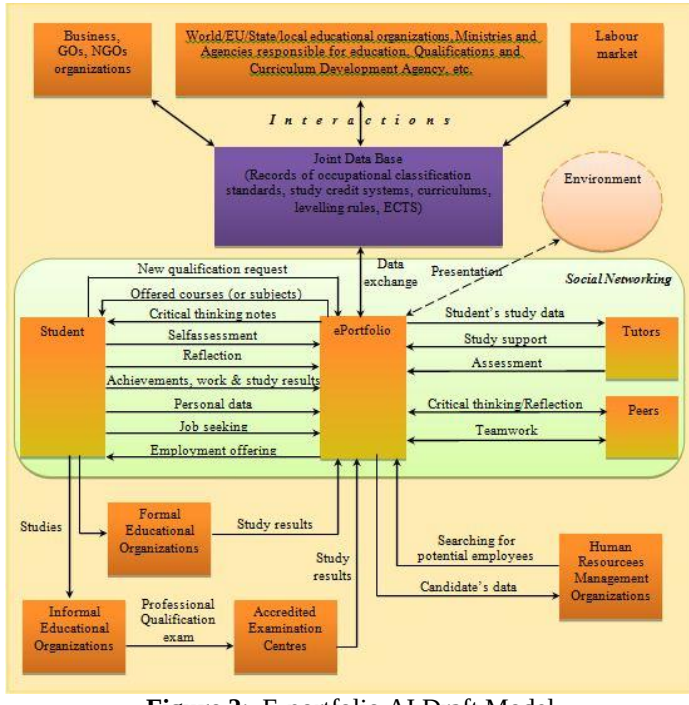


Figure 2: E-portfolio AI Draft Model

Data Analyse Considerations in the Clusters of Competences

To get a sense of the proposed model work, it would be reasonable to see over competence construction in generally. First of all, any of competence consists of knowledge, skills and proficiency, ability to perform duties and apply gained knowledge. It means that we can portion each competence in smaller parts and focus on the problem more detailed. Secondly, in the light of abovementioned considerations, almost each competence might be represented as a cluster of other competencies with their particular number of knowledge, skills and proficiency as it shown, for instance, in Eq. (1): where C is a competence in one of industrial sectors and C1, C2, C3.., Cn – sub-competencies.

$$C = \{C1, C2, C3\ldots Cn\}, \qquad (1)$$

To demonstrate this example, let's assume (just for explanation reasons) that C is aircraft armament engineer's competence. Then C1 might be Higher math, C2 – Higher physics, C3 – English, C4 – Strength of materials, C5 – Theoretical mechanics, C6 – Electrical engineering, C7 – Drawing, C8 – Philosophy, C9 – Economy, C10 – History, C11 – Physical training, C12 – Musketry, C13 – Aircraft armament and

equipment, C14 – Ballistics, and C15 – Explosives. All of these sub-competencies, for their part, also have other level of sub-competencies. It is obvious that the obtaining of competence C1 (higher math) has a prerequisite to have other competencies, such as in algebra and geometry; the obtaining of competence C14 (ballistics) requires to have the competence in physics, but the competence C13 (aircraft armament and equipment) is unthought-of all other competencies and sub-competencies noted before. There are a multitude diversity of competence clusters which have their own variety of clusters and sub-clusters. It is possible to design other competence clusters (A, B, D… Z) and their related sub-clusters. In reality the number of possible competence clusters is almost infinite. This could be seen as a collection of different tailored competencies combined in many possible ways regarding specific professional standards or job descriptions. In other words, competence clusters with complicated combination of competencies will appear. For example, the competence CD8W2Z will consist of all C and Z competencies and sub-competencies, and D8 and W2 competencies and sub-competencies, as it is shown in Eq. (2): Similar sets of competence clusters could be formed for any type of competence issue.

$$CD8W2Z = \{C1, C2\ldots Cn, D8, Z1, Z2\ldots Zn, W2\} \quad (2)$$

Competence Detection and Selection in the Course Self-Managing E-Portfolio Model

Pursuant to considerations noted in previous subsections the exigency of the competence development environment is of the last importance. To simplify the description of operational procedures, initially it would be reasonable to deal with an isolated course. Thus, it could provide eventual possibility for the learners themselves to choose attainable competences order in the course setting. Namely, after choosing of one or another of the course themes/study topics from the list of the themes/topics offered by the e-portfolio system, and completing of the mission, the e-portfolio system will offer to the learner other themes/topics from the compendium of the course.

Obtained competences detection and selection should be organized by analyzing of the obtained competences (both before and during the course) and setting them against competences/sub-competencies list of the course. This model (Fig. 3) prescribes students freedom in the case of setting up their own personal study plans: choosing the order of familiarizing of study topics (here the course designer must define some possible limitations necessary to successful completing of the studies

which the e-portfolio system should maintain; for instance, the student may not start the theme "E" if he/she does not finished themes "A" and "B"), choosing the time of each theme/topic mastering (not exceeding the total hours limits for the course), choosing the classmates to work in teams (depending on their study progress). The e-portfolio system may offer to the student a list of all students who have completed appropriate themes or else automatically knit together determined students in a common team. This approach might be applied also to the whole university, consortium of educational organizations, state, Europe and entire world professional communities.

Results and Discussions

The Distance Education Study Centre of the Riga Technical University has paid attention to the problem of e-portfolio development with AI features. There were a draft e-portfolio model developed and new approaches proposed. It is expected that after adjustments and programming the new e-portfolio system draft model will be put into the practice in 2011/2012 academic year to support the first course university students' activities in their blended learning studies at the business course. It is planned that in the first year of the model's activity it will cover about two hundred learners. Besides, there are competences' self-assessment questionnaires also provided to monitor learners' progress during the course.

They are intended to help the course moderator to track learners' activities and developments, analyze the correlation between competence developments, motivation and usage of e-portfolio system. On other hand, these questionnaires will show to the students their competences growth, and motivate them for further studies. Security issues also affect the e-portfolio and whole IS's run. Student's activities using all features of e-portfolio must be safe. Military and some other ISs require additional security awareness. We should like ours multifunctional IS being cheap, fast and secure. Unfortunately, in reality usually there are only two options available simultaneously. Safety of users' accounts is still actual one. Either to allow external bodies to host e-portfolio services or tend to maintain them yourselves – the choice is yours. The proposed e-portfolio model can solve the problems actualized by labor market urgent needs and lifelong learning necessity. There are many questions to be considered and solved in further research.

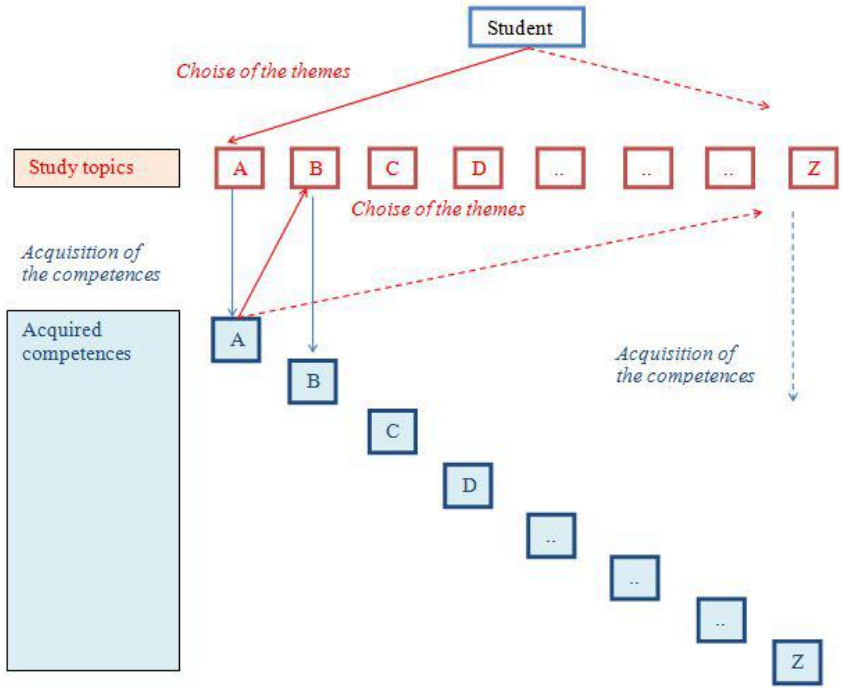


Figure 3: Competence detection and selection scheme

Future Extensions

To satisfy the global labor market and lifelong learning needs, the list/data base of all existing competences ought to be established. The competencies ought to be arranged and assembled into a variety of clusters. Programmers ought to get at the root of things and ensure that proposed e-portfolio system works well. The validation of user's prior competencies and on-going developments, and providing of appropriate competence improvement suggestions by the system, is essential.

As the first further steps we ought to go through the plans to reach established criteria in the recognizing of potential competencies in the world labor market. The JDB requires a set of common credit system, clear rules, definitions and list of educational and professional standards. They can be based on issues of Bologna Agreement. However, it is questionable to count on a reason that all world countries have the same professional standards, curriculums, credit systems and requirements to gain new qualification. If we wish to promote lifelong learning and mobility of manpower, we should tackle the problem of the interoperability and leveling of national standards. For these purposes the

European Credit Transfer and Accumulation System (ECTS), which is a standard for comparing the study achievements and performance of students in European higher education, should be taken into account. This is a challenge to create such a system. There could be discussions about its programming, developing, further maintenance, and costs. An international foundation of the competence clarification, which will deal with competences coding issues, ought to be established. This foundation may consist of number of educational and managerial organizations, and their consortiums.

The establishment of competences list/data base is not a soft snap. It is almost impossible to imagine a creation of the complete competence list at the first onset. Abovementioned proposed foundation may create a task list and distribute them to its members and partners with a task to develop specific competence list/clusters of competences for the specific industry subsector. The common system ought to be based on an idea that knowledge is accessible for everybody. It does not mean that universities should offer classes free of charge. Simply, students must have an opportunity to choose the most suitable option offered from the JDB through their e-portfolios: if it might be for free or else for a valuable consideration.

Review Questions

There might be put a question regarding usability of the proposed e-portfolio system from the credibility position of its competence clusters and their conformity to the labor market requirements. The concern of the detailed list of competence clusters should be activated. Similarly, the ensuring of the competence detection and selection services might be arduous task for the proposed e-portfolio system. Especially it can be faced hardships at the initial time period of the system's run.

Another contention may be related to the necessity of creation of the JDB. The opponents may argue that nowadays all universities have their own portals with embedded data bases which are tailored with the respective university's e-studies section, curriculum, and students' study results. Due to that an assessor can get desiderata information from the university's portal, hereby financial resources could be saved.

Solutions to Review Questions

To achieve the comprehensive aggregate of competences which would belong to variety of professions, we should gather information about the list of the professions, professional standards, industries, sectors

and subsectors of industries. Each industry subsector has a number of professions and exposure of professional standards related to each particular profession.

The gateway to success is, first of all, to study, be aware and take into account professional standards or job/position requirements to certain industry subsector professions. Secondly, it is consequential to endow to all set of professions the precise letter and digit codes. Thirdly, it is of the last importance to compare these findings and make them interoperable which means to make possible to use some parts of individual sub-competencies from the competence clusters to describe other professions.

Heretofore, if the comprehensive competence detection and selection sub-system is not developed yet, initially the proposed model work can be guaranteed by manual data processing in the e-portfolio system. At the first stage of implementation a limited number of related to the study course competencies might be processed. This problem cannot be solved by efforts of the only one university. To further develop the system there is a need to find the partners, confreres who are interested in this issue, and build collaborative foundations. The argument regarding disavowal of the JDB necessity might not be accepted because of the localization policy of many educational organizations, aspiration for the independence and protection of own assets. Besides, usually universities close students' accounts after their graduation and deny their rights to use these accounts.

Acknowledgement

This research has been partly supported by a grant from the European Regional Structural Fund (ERSF), Project No. 2DP/2.1.1.1.0/10/APIA/ VIAA/150 (RTU PVS ID 1534). The author would like to thank the anonymous referees for their very constructive comments.

References

[1] K. C. Laudon and J. P. Laudon: Essentials of Business Information Systems, 7th ed. Person: Prentice Hall (2007).

[2] A. Gorbunovs: Actuality of Interactive E-Portfolio Systems, in Proc. of the International Conference on e-Learning and the Knowledge Society - e-Learning'10, Latvia, Riga, 26.-27. August, 2010, - pp. 148-153 (2010).

[3] C. Emden, D. Hutt, et al: Exemplar. Portfolio learning/assessment in nursing and midwifery: an innovation in progress, Contemporary Nurse, vol. 16(1–2), pp. 124–32 (2003).

[4] E. Heinrich: Contrasting approaches: Institutional or individual ownership in ePortfolio systems, in Proc. Ascilite Melbourne 2008: Concise paper: Heinrich (2008).

[5] J.Kapenieks: E-learning environment for sustainability in educational action

research. 9th International JTEFS/BBCC Conference "Sustainable Development. Culture. Education", Siauliai, Lithuania (2011).

[6] J. Schoonenboom, H. Sligte, A. Moghnieh, D. Hernández-Leo, K. Stefanov, C. Glahn, M. Specht, and R. Lemmers: Supporting life-long competence development using the TENCompetence infrastructure: a first experiment. International Journal of Emerging Technologies in Learning (Special Issue), 3(1), 53-59 (2008).

[7] A. J. Berlanga, P. B. Sloep, F. Brouns, M. Bitter, and R. Koper: Towards a TENCompetence ePortfolio. International Journal of Emerging Technologies in Learning (iJET), 3, 24-28 (2008).

[8] Government of Republic of Ireland: Report of the Commision of Nursing: A Blueprint for the Future. Dublin: The Stationery Office (1998).

[9] W. Redman: Portfolios for Development: A Kogan Page (1994).

[10] R. A. Brown: Portfolio Development and Profiling for Nurses. Lancaster: Quay Publishing (1992).

[11] M. Jasper: Beginning Reflective Practice: Foundations in in Nursing and Health Care, Cheltenham: Nelson Thornes (2003).

[12] H. C. Barrett: Another misunderstanding of term "electronic portfolio" (2010). [Online]. Available: https://www.blogger.com/comment.g?blogID=7079154&postID=2655773264462369508

[13] D. Kolb: Experiential learning: Experience as the source of learning and development, Englewood Cliffs, NJ: Prentice-Hall (1984).

[14] J. Zull: The Art of Changing the Brain, Sterling, VA: Stylus Publishing (2002).

[15] H. C. Barrett: Balancing the Two Faces of ePortfolios (2009). [Online]. Available: http://electronicportfolios.org/balance/balance.pdf

[16] M. Alterio: Collaborative Journaling as a Professional Development Tool, Journal of Further and Higher Education, vol. 28(3), pp. 321-332 (2004).

[17] A. Silberschatz, P. B. Galvin, and G. Gagne: Operating System Concepts, 7th ed., John Wiley & Sons, Inc. (2005).

[18] H. C. Barrett: Categories of ePortfolio Tools (2008). [Online]. Available: http://electronicportfolios.org/categories.html

[19] M. A. Chambers: Some issues in the assessment of clinical practice: a review of the literature, Journal of Clinical Nursing, vol. 7(3), pp. 201–8 (1998).

[20] J. G. Alexander, S. W. Craft, et al: The nursing portfolio: a reflection of a professional, Journal of Continuing Education in Nursing, vol. 33(2), pp. 55–9 (2002).

[21] M. S. Knowles: The Adult Learner: A Neglected Species, Houston, TX: Gulf Publishing Company (1989).

Chapter 4

ANALYZING WEB SERVICES NETWORKS: A WS-NEXT APPLICATION

Chantal Cherifi, Jean-François Santucci

Summary

Web services represent a system with a huge number of units and many various and complex interactions. Complex networks as a tool for modeling and analyzing natural environments seem to be well adapted to such a complex system. To describe a set of Web services we propose three Web services network models based on the notions of dependency, interaction and similarity. Using the WS-NEXT extractor we instantiate the models with a collection of Web services descriptions. We take advantage of complex network properties to provide an analysis of the Web services networks. Those networks and the knowledge of their topological properties can be exploited for the discovery and composition processes.

Introduction

Web service is a set of related functionalities that can be published and discovered in a Web services registry and invoked for remote use. Those modular applications can be programmatically loosely coupled through the Web to form more complex ones. Two of the most popular problems in Web service technology addressed by both industry and academia are discovery and composition [1]. Discovery is the process of locating advertised Web services that can satisfy a service request. Composition arises when several Web services are needed to fulfill a request. The way those processes are achieved depends on how Web services are described. For syntactic Web services descriptions, discovery is performed on registries using keywords. Compositions are manually defined before any submitted request. Semantic descriptions allow automatic discovery and composition processes. Nevertheless finding the right Web services to fulfill a given request is not an easy task. Indeed, Web services are extremely volatile. Their number is continuously growing, and providers may change, relocate, or even remove them.

In this context, the Web services substitution play an important role within the composition process. Substitution consists in using a Web service instead of another. The only constraint is that the replacing one produces the same output and satisfies the same requirements as the replaced one. To perform Web services substitution, the Web services classification process aims at grouping Web services into categories usually called communities. Hence, works in Web services classification aim at grouping Web services according to some similarity criteria [2] [3] [4]. Classification is a step in structuring the Web services space to improve discovery and composition processes. Other criteria can be used to organize a set of Web services like their ability to be composed. In this case communities are formed with Web services that can interact in a composition. On the one hand, Web services represent a system composed by a large number of highly interconnected dynamical units. On another hand, complex networks are a tool specifically dedicated to model such natural and complex systems. They allow studying their structure and dynamics [5]. Hence, a set of Web services can naturally be represented under the form of networks according to different criteria such as their similarity or their ability to be composed. Such kind of structures constitutes a convenient way to represent a set of Web services for visualization and analysis purposes. Moreover they can be stored and serve as a guide for Web services discovery and composition.

In this article, we introduce three models to structure a set of Web services. A dependency and an interaction model materialize Web services composition. A similarity model materializes similarity between Web services. We then provide a topological analysis of the networks structure using a well known benchmark. The rest of the paper is organized as follows. Background key elements are provided in section 2. Section 3 is dedicated to the literature review. Variables used to elaborate networks taxonomy are presented in section 4. In section 5 we introduce networks definitions. The networks taxonomies are presented in section 5. In section 6 we provide an analysis of the structure of some network samples. Finally, conclusions are provided in section 7.

Background

In this section we give some background elements on Web services definition, Web services description languages, Web services discovery and composition and Web services classification.

Web Service Definition

Different kinds of information are linked to the notion of Web service. Some non functional properties (service provider, quality of service, service location) are present aside the Web service functionalities. In this work we focus on the functional aspect of Web services. Hence, we consider a Web service as an interface. A Web service interface is defined as a set of operations. An operation i represents a specific functionality. It is characterized by one set of input parameters noted I_i, and one set of output parameters noted O_i. I_i is the required information in order to invoke a Web service operation. At the Web service level, the set of input parameters of a Web service k is $I_k = \cup I_i$ and the set of output parameters $O_k = \cup O_i$. Fig. 1 represents a Web service numbered 1 with two operations, Operation 1: ($I_1 = \{a, b\}, O_1 = \{d\}$), Operation 2: ($I_2 = \{c\}, O_2 = \{e, f\}$), Web service 1: ($I_1 = \{a, b, c\}, O_1 = \{d, e, f\}$).

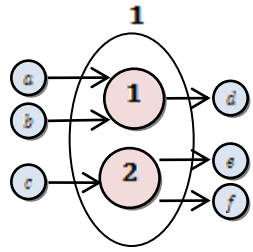


Figure 1: Schematic representation of Web service 1, with two operations 1 and 2

Description Languages

Production Web services are mostly expressed with Web Service Description Language (WSDL), a syntactic Web services description language [6]. This XML-based language has been proposed in the context of the W3C. More recently, the research community followed the current semantic Web trend by introducing semantics in Web services descriptions. Semantic Web services aim at augmenting Web services with rich formal descriptions of their capabilities. Several initiatives for semantic description languages exist and we can distinguish two main conceptual approaches. The first one aims at semantically annotating existing WSDL descriptions of Web Services. WSDL-Semantic (WSDL-S) [7] or Semantic Annotation for WSDL (SAWSDL) [8] are two semantic extensions of WSLD. The second approach aims at developing pure semantic Web services. The field includes substantial bodies of

work, such as the efforts around Ontology Web Language for Services (OWL-S) [9]. OWL-S is ontology of Web services specified by the W3C.

Composition

Web services composition addresses the situation when a request cannot be satisfied by any available atomic Web service. In this case, a composite Web service is synthesized to fulfill the request. A composite Web service is obtained by combining existing available atomic or even other composite Web services. The composition synthesis thus produces a specification of how to link the available Web services to realize the request.

Classification

Considering a set of Web services, the classification process aims at grouping them into categories. These categories are usually called communities. As in the literature classification is mainly performed according to the similarity between Web services, we will focus in work based on this definition. In this case there are two approaches to define communities i.e. top-down or bottom-up. In the former, abstract communities are designed a priori, and Web services are then defined in order to fit these categories [2] [3] [4]. In the later, communities are mined from an existing Web services collection [10] [11].

Literature Survey

Despite the great potential they offer in terms of analysis tools, complex networks have not been widely used in the Web services area so far. Nevertheless, some authors already followed this recent trend to structure a set of Web services. In [12] the authors define three composition network models according to the node types that can be parameters, operations or Web services. They use syntactic Web services to build networks considering either a full or a partial interaction between the nodes. Two types of syntactic matching i.e. equal and flexible are used to compute the links between networks nodes. Using complex network theory, they provide an analysis of the topological landscape of Web services networks formed by a real-world data set.

In [13] and [14] the authors provide an interaction network model with Web services as nodes. They deal with semantic Web services considering a partial interaction mode. Equivalence and subsumption ontological concepts relationships are considered to compute the links between networks nodes. In [13], complex network theory is also used to

rank the Web services according to their connectivity. Experiments are performed on an automatically generated and simulated Web services network. A composition algorithm is applied to the networks while being dynamically guided by the ranking. In [14] the network is built from a set of artificial Web services descriptions. To synthesize a composition plan the network is explored with a backward chaining discovery.

In [15] the authors propose a dependency network model with parameters as nodes. The model is based on semantically described Web services. The network is used to derive composite Web services with a breadth first search algorithm. In [16] an interaction Web services network is proposed. The goal of this work is to classify Web services. The authors provide a graph based method for composition oriented Web services classification using a b-coloring approach. From all these works we can observe that there are various ways to represent a Web services set as a network. We can identify some variables to distinguish the proposed models. The Web services description, the network nodes, the relationship between nodes, the amount of information considered to establish a relation between two nodes and the matching are among the variables that allows building different types of networks. To evaluate all the possible models based on these variables we derived a tool, WS-NEXT, that allows building associated networks from a given set of Web services [17].

Network Variable

In this section we give an accurate meaning of the previously identified Web services network variables that can be used to modulate a Web services network.

Description

The description variable represents the Web service description type. Those two types are syntactic and semantic descriptions. Corresponding variables values are respectively noted syntactic and semantic. In a syntactic description, each parameter has a name and an XML type. In a semantic description the name and the type are also generally specified and an additional ontological concept is associated to the parameter. Ontological concepts are domain specific and consensual terms. They give parameters a contextual and precise meaning.

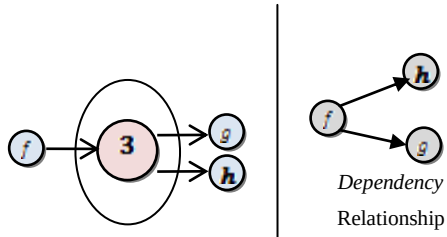
Granularity

The granularity determines the nodes entities i.e. the nature of the

nodes in a network. From coarser to finer, we consider Web services, operations or parameters as node entities. We note the corresponding variables values as service, operation and parameter.

Model

The model expresses the nature of the links i.e. the type of relationship between nodes. This relationship depends on the granularity. Considering parameters as nodes, if one is an input parameter of a Web service (or of an operation) and if the other is an output parameter of the same Web service (or operation), there is a dependency relationship between them. Indeed, the production of the second parameter depends on the provision of the first one through the invocation of the Web service (or of the operation). This model is noted dependency and is illustrated by Fig. 2. One Web service numbered 2 is considered with one operation numbered 3 and its parameters as follows: 3 ($I_3 = \{f\}$, $O_3 = \{g, h\}$). Parameters h and g depends on the provision of parameter f, hence there is a dependency relationship between them f and h and g.



(a) one Web service with one operation and its parameters

(b) corresponding dependency relationships between parameters

Figure 2: Example of the Dependency Model

Considering Web services or operations as nodes, a relationship between two nodes corresponds to the information flow between them. In other words the first one is able to provide the information needed by the second one in order to invoke it. This model is called interaction. It is illustrated by Fig. 3. Two Web services 1 and 2 are considered. Web service 1 has two operations as follows: 1 ($I_1 = \{a, b\}, O_1 = \{d\}$) and 2 ($I_2 = \{c\}, O_2 = \{e, f\}$). Web service 2 has one operation as follows: 3 ($I_3 = \{f\}, O_3 = \{g, h\}$). Web service 1 can provide the information in order to invoke Web service 2; hence there exist an interaction

relationship between them. Operation 2 can provide the information in order to invoke operation; hence there exist an interaction relationship between them.

Dependency and interaction models are different ways to materialize Web services composition. Considering operations as nodes, a relationship between two operations corresponds to a certain type of similarity between them. The similarity relation can be either symmetrical or asymmetrical. In the first case, the two operations are said to be similar to each other. In the second case, the second operation is said to be similar to the first one according to some criteria. This model is noted similarity and is illustrated by Fig. 4. Three operations with their respective parameters are considered as follows: 1 ($I_1 = \{a, b\}, O_1 = \{d\}$), 4 ($I_4 = \{a\}, O_4 = \{b, d\}$), 5 ($I_5 = \{a, b\}, O_5 = \{d, e\}$). The similarity relation between operations 1 and 4 is symmetrical.

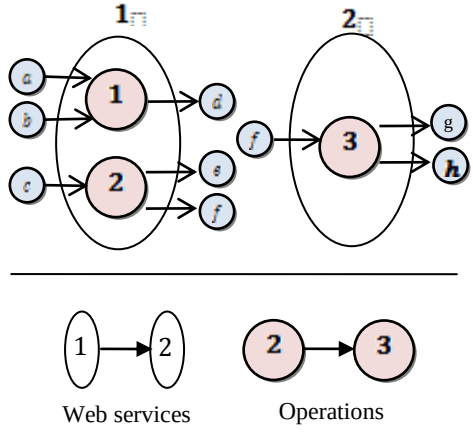


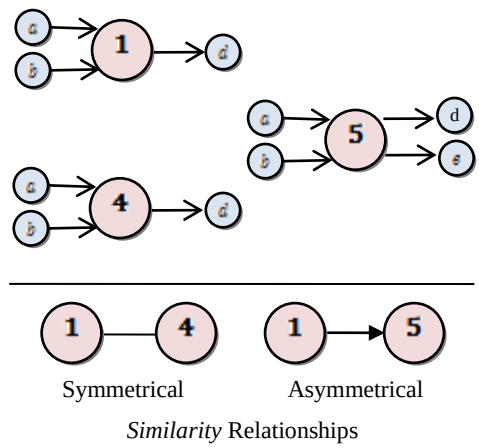
Figure 3: Example of the Interaction Model. Top: two Web services with their respective operations and parameters. Bottom: corresponding interaction relationships between Web services and between operations.

Indeed the two operations are symmetrically similar because they have the same output parameters. An asymmetrical similarity relationship exists between operation 1 and operation 5. Indeed, operation 5 has more output parameters than operation 1.

Mode

The mode represents the amount of information used to link two nodes in a network. This variable is related to the interaction model. Two cases must be considered. Either all the information is provided or only part of

this information exists. If a Web service or an operation can provide all the parameters values needed to invoke another one, we will denote this case as full interaction mode. Fig. 3 is an example of the full interaction mode. A full interaction exists between Web service 1 and Web service 2. Indeed, Web service 2 needs only parameter f to be invoked and Web service 1 can provide this information. A full interaction also exists between operation 1 and operation 3. If a Web service or an operation cannot provide all the input parameters required by a second one, this mode is denoted by partial. Such a case is illustrated by Fig. 4. Two Web services 2 and 3 are considered. Web service 2 has one operation 3 ($I_3 = \{f\}, O_3 = \{g, h\}$). Web service 3 has one operation 6 ($I_6 = \{g, h, i\}$, $O_6 = \{j, k\}$). There is a partial interaction between Web service 2 and Web service 3. Indeed, Web service 2 can provide only part of the information needed by Web service 3 which is parameters g and h.



Similarity Relationships

Figure 4: Example of the Similarity Model. Top: three operations with their respective parameters. Bottom: corresponding similarity relationships between operations.

Matching

The matching variable is related to the similarity measures between parameters. It is computed differently for syntactic and semantic descriptions. For syntactic descriptions, matching consists of comparing two Web services parameters names using similarity metrics. We distinguish two cases. The first case considers two parameters as similar if their names are exactly the same string. It is called equal. The second case considers two parameters as similar if their name presents a certain level of similarity. It is called approximate. Different similarity metrics

can be used. Classical similarity metrics have been considered in WS-NEXT (Levenshtein, Jaro and Jaro-Winkler). These metrics are denoted as Levenshtein, Jaro and Winkler. We also developed a smoothed metric based on Levenshtein distance between filtered strings denoted Smoothed.

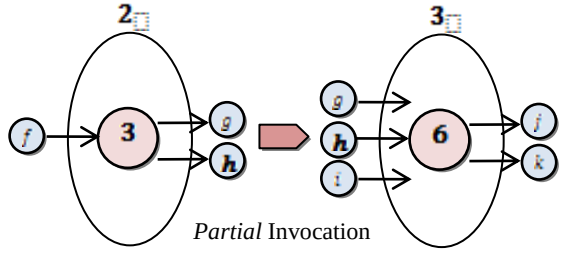


Figure 5: Example of Partial Interaction Mode.

For semantic descriptions, matching consists in comparing ontological concepts associated to the parameters. This is done by the classical operators (exact, plugin and subsume) that have been developed in previous work in the matchmaking area [18]. Exact corresponds to a perfect matching, i.e. both concepts belong to the same ontology and are exactly identical. When the concept associated to the first parameter is strictly more specific than the other one, plug-in is used. Subsume is used when the first concept is strictly more general than the second one. We add a fourth matching operator called fitin which consider the case where there is simultaneously plug-in and exact similarities between two nodes. This operator leads to a more flexible semantic interaction representation. The matching variables values are denoted by exact, plug-in, subsume and fitin.

Networks Definition

Dependency, interaction and similarity networks can be used to represent a set of Web services. In dependency networks nodes are parameters while interaction networks can use either operations or Web services as nodes. In similarity networks nodes are operations. In the following we define the three corresponding network models.

Dependency Network

A dependency network is defined as a directed graph whose nodes correspond to depending parameters and links indicate the head parameter depends on the tail parameter (as illustrated by Fig. 2, g depends on f) [19]. In the context of dependency networks, each Web

service w is formally defined as a triplet (I_w, O_w, K_w). K_w denote the set of dependencies defined by w. Each operation i is formally defined as a triplet (I_i, O_i, K_i). K_i denotes the set of dependencies defined by i. We consider each output parameter depends on each input parameter. To build such a network, we first associate one node to each parameter present in the whole collection. Then, links are drawn by considering each Web service (or operation) separately. A link is added between each one of its input parameters and each one of its output parameters. Additionally, one parameter may be used by several Web services or operations, either as an input or an output. Consequently, we have to decide if two parameters are similar. This is done trough the matching functions described in section 4. In the case of syntactic dependency network, equal matching is applied. For a semantic description exact matching is applicable.

Interaction Network

We define an interaction network as a directed graph whose nodes correspond to interacting Web services and links indicate the possibility for the tail Web service to act on the head Web service [20]. To represent a collection of Web services descriptions as an interaction network of Web services, we first define a node to represent each Web service in the collection. Then, a link is drawn from a Web service 1 to a Web service 2 if for each input parameter in I_2, a similar output parameter exists in O_1. In other words, the link exists if and only if Web service 1 can provide the information requested to invoke Web service 2. In the interaction network, a link between two Web services therefore represents the possibility to compose them. Similarly, we can define an interaction network at the operation level. The matching functions described in section 4 are used to determine the similarity between two parameters.

Similarity Network

We define a similarity network [21] as a graph whose nodes correspond to possibly similar Web services operations. To represent a collection of Web services as a similarity network of operations, we first associate a node to each operation in the collection. Then, a link is added between two nodes if the corresponding operations are similar. The similarity relation between two sets of parameters can be approached in several ways. To that end, we defined four similarity functions. They are respectively named Full Similarity (FullSim), Partial Similarity

(PartialSim), Excess Similarity (ExcessSim) and Relation Similarity (RelationSim). These functions are defined in terms of set relations between the input and output parameters sets of the compared operations. Let I_i and O_i be the sets of input and output parameters for operation i respectively. Suppose we want to compare operation 1 and operation 2. FullSim states both operations are fully similar if they provide exactly the same outputs ($O_1 = O_2$) and if they have overlapping inputs ($I_1 \cap I_2 \neq \emptyset$). PartialSim states 2 is partially similar to 1 if some outputs of 1 are missing in 2 ($O_1 \supset O_2$) and if they have overlapping inputs ($I_1 \cap I_2 \neq \emptyset$). ExcessSim states 2 is similar to 1 with excess if 2 provides all the outputs of 1 plus additional ones ($O_1 \subset O_2$) and if 2 has at most the inputs of 1 ($I_1 \supseteq I_2$). The RelationSim function states both operations have a relational similarity if they have exactly the same outputs ($O_1 = O_2$) and if they do not share any common input ($I_1 \cap I_2 = \emptyset$). To determine the relations between two sets of parameters, one needs to be able to compare the parameters themselves. Hence, the similarity functions are based on the equal matching described in section 4.

Web Services Networks

In order to build Web services networks from a set of Web services descriptions, we used WS-NEXT (Web Services Network Extractor). WS-NEXT allows building networks from a collection of Web services descriptions files, according to the network models and the variables previously defined. Networks that can be extracted by WS-NEXT are figured by a tree starting from the root, going through each variable and ending by an underlined leaf.

Dependency Taxonomy

Fig. 6 shows the dependency networks taxonomy. Two dependency networks can be extracted with WS-NEXT, one syntactic with equal matching and one semantic with exact matching.

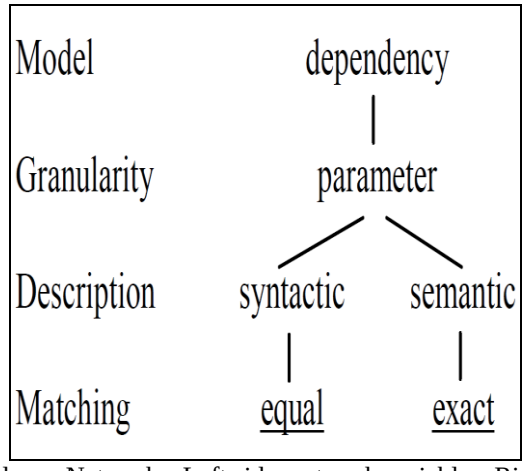


Figure 6: Dependency Networks. Left side: network variables. Right side: networks

Interaction Taxonomy

The interaction networks taxonomy is depicted by Fig. 7 and Fig. 8. Eighteen full interaction networks and eighteen partial interaction networks can be extracted with WS-NEXT.

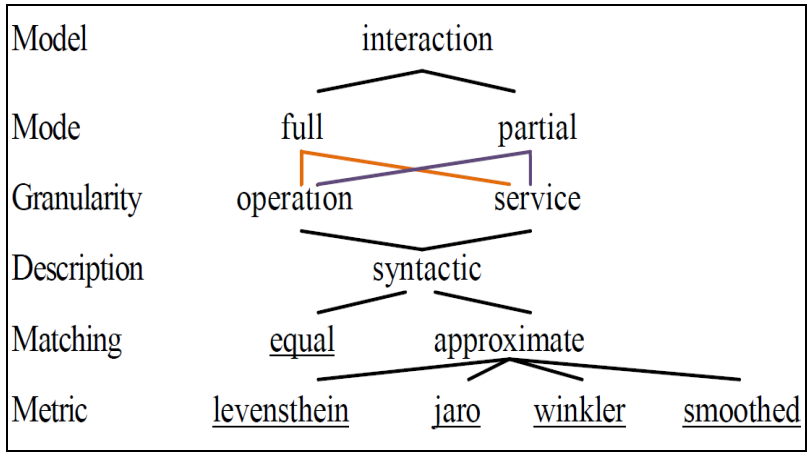


Figure 7: Syntactic Interaction Networks. Left side: network variables. Right side: networks.

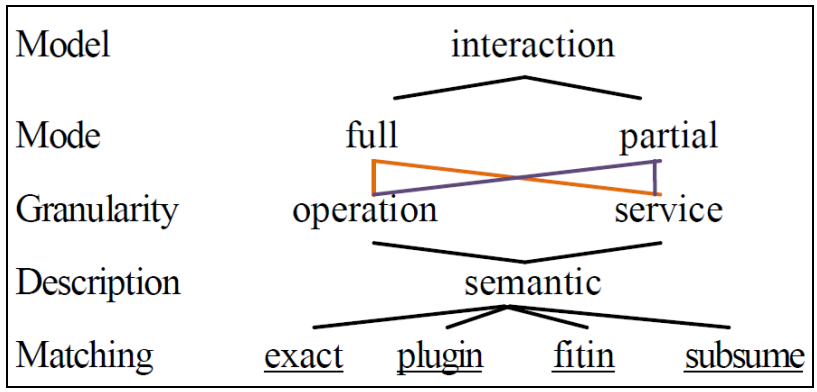


Figure 8: Semantic Interaction Networks. Left side: network variables. Right side: networks.

Similarity Taxonomy

The taxonomy of similarity networks is represented on Fig. 9. We can extract eight similarity networks with **WS-NEXT.**

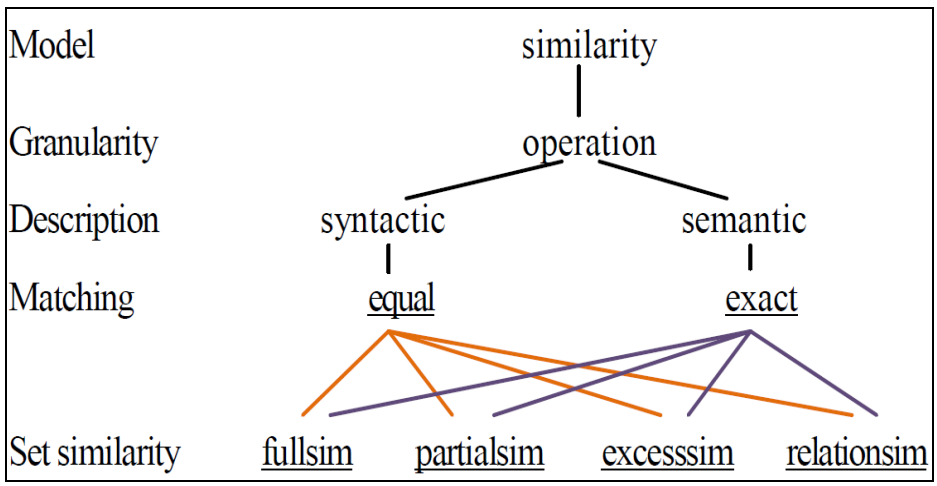


Figure 9: Similarity Networks Taxonomy. Left side: network variables. Right side: networks.

Networks Extraction and Analysis

From a collection of Web services descriptions, we extracted a set of ten networks as follows. Two dependency networks: 1 syntactic (equal) and 1 semantic (exact); four interaction networks: 1 syntactic (equal) and 3 semantic (1 exact, 1 plugin, 1 subsume); four syntactic similarity networks: 1 FullSim, 1 PartialSim, 1 ExcessSim, 1 RelationSim. We then investigated the structural properties of the networks.

In these experiments, we follow two main objectives. First, we want to study the influence of semantics on the composition process by comparing the structure of the syntactic and semantic composition networks. Second, we want to investigate the relation between the network structure and the application for which it is devised. Usually, networked systems exhibit a component organization. A component is a maximal connected sub-graph disconnected from the rest of the network. Either a network exhibit a giant component with small other ones, or all the components have an equivalent size. By tracking the component organization, the components size and links number, we can interpret the analysis results in terms of Web services composition and substitution.

The networks have been extracted from the SAWSDL-TC1 [22] Web services descriptions collection. Indeed, in this work, we want to simulate real-world conditions and to compare syntactic and semantic composition Web services networks. Hence, one need to have a collection of a large number of real-world Web services described both syntactically and semantically. SAWSDL-TC1 provides 894 Web services descriptions written in SAWSDL. Each description contains only one operation. The collection contains 2136 parameter instances. Parameters are syntactically described by their name and semantically described by their ontological concept. The descriptions are distributed over 7 thematic domains (education, medical care, food, travel, communication, economy and weapon). The collection originates in the OWLS-TC2.2 collection, which contains a part of real-world Web services descriptions retrieved from public IBM UDDI registries, and semi-automatically transformed from WSDL to OWL-S.

Structure of Dependency Networks

We extracted the syntactic and the semantic dependency networks with WS-NEXT, according to the dependency network definition. As matching functions gather similar parameters, there is a significant difference between the number of instances in the collection and the number of parameter nodes in the networks. The 2136 parameters instances of the collection are represented by 385 nodes in the syntactic network and by 357 nodes in the semantic one. As we used different matching functions to build the syntactic and semantic networks, the sets of similar parameters are not the same in the syntactic network and in the semantic network. The number of nodes is smaller in the semantic network. This indicates semantic matching allows associating more parameter instances. This result highlights the presence of false negatives

in the syntactic network. False negatives are instances associated to different nodes in the dependency network. They are actually conveying the same information and should be represented by the same node. These false negatives are usually syntactically different because they have different names. But they have the same meaning; hence they are associated to the same ontological concept. For example parameter instances _AUTHOR; _AUTHOR1 and _AUTHOR2 are represented by three distinct nodes in the syntactic network. In the semantic network, they are associated to a unique node as they all are associated to the same #author concept. The semantic matching also allows eliminating some false positives. False positives correspond to instances represented by the same node whereas they do not represent the same information. For example, many instances are simply called PARAMETER but are associated to very different concepts. The syntactic matching will improperly associate them to a common node, whereas the semantic matching will not.

Both networks exhibit the same structure: a giant component along with several small components and isolated nodes. Nevertheless, the distribution between these three types of entities is slightly different. The proportion of isolated nodes is 4.67% in the syntactic network and 4.2% in the semantic network. While this value is smaller for the semantic network, the number of isolated nodes remains quite small in both networks. Isolated nodes are parameters belonging to Web services having only input parameters or to Web services having only output parameters, and they are exclusively either input or output. The giant component in the syntactic network contains 73% of the remaining nodes and 86% of the remaining links while in the semantic networks it contains 78% of the nodes and 88% of the links. The syntactic network exhibits 17 other smaller components with a size ranging from 2 to 30 nodes. This is to compare to 15 small components for the semantic network (2 to 14 nodes). Fig. 10 shows the trimmed semantic network (isolated nodes have been discarded). The giant component stands in the middle surrounded by the small components.

The semantic network presents less isolated nodes and less small components than the syntactic one. These properties are more effective in terms of composition ability. Recall that if many distinct components exist this reflects that the collection is made of several non-interacting groups of parameters. Furthermore the semantic network has a larger giant component than the syntactic one both in terms of nodes and links. It shows that the numbers of dependencies in which several operations

are implied is higher. These results demonstrate that a larger proportion of Web services can interact if one uses the semantic network.

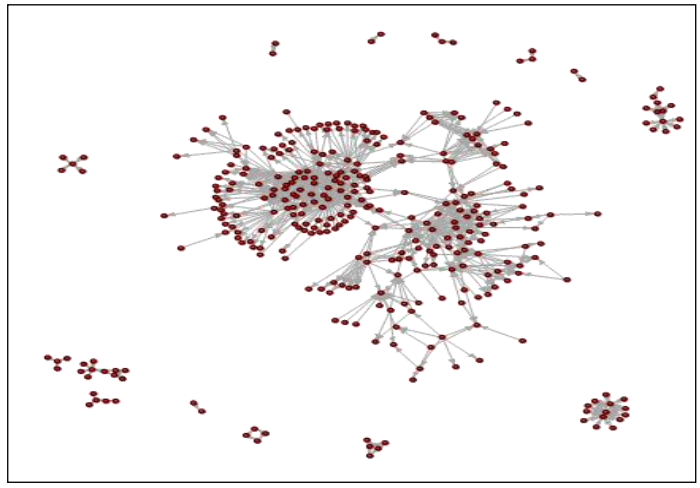


Figure 10: Trimmed exact semantic dependency network. The giant component is located in the middle surrounded by small components.

Structure of Interaction Networks

According to the interaction network definition, we extracted 4 networks with WS-NEXT from SAWSDL-TC1 collection, one syntactic and three semantic. The syntactic network is the full equal network. In some previous work [23] we performed a comparative study on the metrics performance by studying the topological properties of syntactic approximate networks. It appears that the use of the approximate metrics to build interaction networks is not very satisfying. For this reason, we concentrate on the equal network. The semantics networks are the full exact, the full plug-in network and the full subsume network. In this study, we discarded the fitin network to keep and compare only strict subsumption relationships. We restrict our investigations to the full mode. Indeed, we want to put ahead eventual differences between syntactic and semantic network structures not to compare intra-model variations.

The number of nodes and links is globally higher in the syntactic network than in the semantic networks. The syntactic network contains 395 nodes and 3666 links. The exact network contains 341 nodes and 3426 links. The plug-in network contains 369 nodes and 2446 links. The subsume network contains 329 nodes and 3864 links. This result is the consequence of the presence of some false positives in the syntactic

network.

The same structure is shared by all the networks. We remark the presence of isolated nodes, a giant component and small components much smaller than the giant one. The four networks contain many isolated nodes. They represent 44% of the total nodes in the syntactic network. This proportion is approximately 49% in the semantic networks. There is less isolated nodes in the syntactic network because some nodes are inappropriately linked to others and cannot participate in a composition. In an interaction network, isolated nodes represent Web services that do not interact with others. None of their output parameter can serve as input and none of their input parameter is provided by other Web services. Hence, they only can be invoked as atomic Web services. In the syntactic network, the giant component contains 90% of the remaining nodes and 99% of the remaining links. The giant component of the exact network contains 85% of the nodes and 98% of the links in the trimmed network.

The plugin and subsume semantic networks present very similar proportions. Once again these results highlight the presence of false positives in the syntactic network. The syntactic network exhibit 5 small components ranging from 2 to 22 nodes. The exact network has 7 small components whose sizes range from 2 to 28. Plugin and subsume networks exhibit 5 small components respectively ranging from 3 to 10 and from 5 to 90. Fig. 11 shows the trimmed exact semantic network separated in 8 components. The small components are less numerous and smaller in the syntactic network because of the presence of false positives that have been integrated in the giant component. In the plugin and in the subsume networks, the constraints on the interactions are relaxed comparing to the exact network. Hence, nodes are gathered within fewer components. The exact semantic network presents more isolated nodes, more small components and a smaller giant component than the syntactic one. These properties seem to be less effective in terms of composition ability. Nevertheless, the interconnection structure is more accurate in a semantic network. It should consequently results in a more efficient composition discovery process. One may consider the plugin and the subsume networks as additional solutions for this task. In this case, the resulting semantic search space becomes larger than the syntactic one.

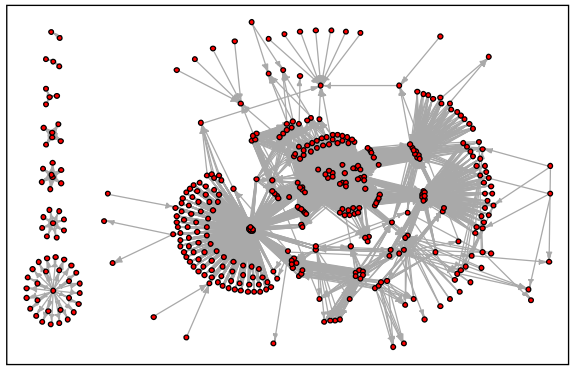


Figure 11: Trimmed exact semantic interaction network. The giant component is located on the right side. The small components stand in the right side.

Structure of Similarity Networks

According to the definitions, four syntactic similarity networks have been extracted with WS-NEXT from SAWSDL-TC1 collection. We choose to study only one description type to concentrate on similarity functions comparison. The four networks contain 785 nodes, corresponding to the 785 operations of the collection. Table 1 summarizes the values of the networks properties. Except for the first row, all the others properties are computed on the trimmed networks, i.e. without any isolated nodes. For all similarity networks under study no giant component is emerging, but numerous small ones, along isolated nodes. This reflects the decomposition of the collection into a reasonable number of communities. This is a good thing, because having only isolated nodes or a giant component would lead to useless communities. Indeed, in the former case, each community would contain only one operation, and in the latter all operations would be considered as similar to the all others. Both cases would have been surprising considering we processed a real-world collection. To illustrate the structure of the similarity network, a typical component from the partial network is presented in Fig. 12.

Operations get_DESTINATION_HOTEL, get_SPORT_HOTEL, get_ACTIVITY_HOTEL are linked with get_HOTEL. Indeed get_HOTEL operation provides only the HOTEL output parameter while the three others provide the HOTEL output parameter and an additional specific one. A get_HOTEL operation can partially satisfy a destination/hotel request, an activity/hotel request or a sport/hotel request. From full similarity to relation similarity according to table 1 order, the number of isolated nodes globally decreases while the number of links

and components increases. Indeed, as constraints on outputs become less strict, more links are created leading to new components or to the increase of the population of the existing ones. The number of nodes, the number of links and the number of components are the highest in the relation similarity network. In this collection, a lot of operations produce identical outputs with completely different inputs.

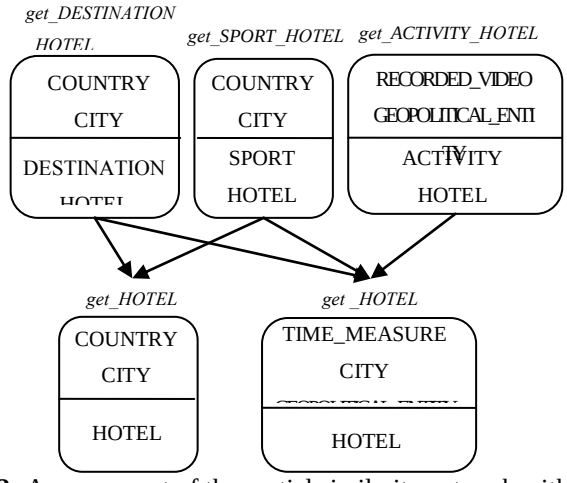


Figure 12: A component of the partial similarity network with 5 nodes.

Table 1: Properties of full, partial, excess and relational similarity networks.

Property	**Full Sim**	**Partial Sim**	**Excess Sim**	**RelationSim**
Isolated nodes	604	447	486	227
Nodes in trimmed network	181	338	299	548
Components	38	61	67	123
Links	310	412	307	2254

Let's inspect the number of components containing 90% of nodes all together. We need 17 components in the full similarity network, 30 in the partial similarity network, 40 in the excess similarity network and 32 in the relation similarity network. Those results show that at least half of the components contain very few nodes while the other half contains at least 90% of nodes. These small components are not very interesting; they do not offer many opportunities in terms of substitution.

Discussion

From the comparison between syntactic and semantic networks, for both dependency and interaction models, it appears that the semantics in the Web services descriptions leads to more accurate interconnection structures. Indeed, we demonstrated that the inappropriate dependencies and interactions relationships that appear in the syntactic networks are discarded in the semantic networks due to the use of ontologies and semantic matching. One can expect though, a more efficient composition process using the semantic description. A large body of work exists in the domain of semantic descriptions and automatic Web services composition. Nevertheless, production Web services still widely rely on syntactic descriptions. To take advantage of the great potential of a semantic Web services pool, one should be able to annotate legacy Web services descriptions. Manual annotation is a complex and costly task hence there is a need to appropriate annotation tools. Few researchers have proposed solutions for this task [24] [25]. At this point there is no satisfying solution that can perform an efficient fully automatic annotation. Bridging the gap between a syntactic and a semantic notation is a difficult problem. We suggest devising semi automatic annotation tools as a first step towards this goal.

The giant component in the composition networks structure reflects the presence of a huge number of interconnected Web services. In these networks, the presence of a giant component is of great importance. It represents the largest fraction of the network within which compositions are possible. It is a guaranty for a composition process to be successful. In the similarity networks, no giant component emerges. They are rather divided into numerous small components. This structure reflects the decomposition of the networks into many Web services communities and, as a consequence, of substitutable operations. A composition process could take advantage of these two complementary structures. We can combine the two structures to obtain two-level architecture. We suggest an upper level containing an interaction network. Each node of this network could be an abstract operation gathering similar concrete operations. Hence, the composition search space would be reduced. The lower level is then represented by the similar networks. The abstract operations of a composition could be instantiated by concrete operations of the lower level with the possibility of substitution.

Conclusion

In this paper, we proposed three network models to structure a set of Web services. The models aim at organizing the Web services according two different directions. The first one takes into account the composition relationship between Web services. The second one is based on their similarity relationships. We provide a topological analysis of the networks. This analysis shows that the structure of the semantic description for composition networks is more accurate. Networks formed from the composition models exhibit a giant component in which a large number of Web services are interacting. Networks formed from the similarity model are composed by many small components which gather a pool of similar operations. Our future work will focus on algorithms based on the composition and similarity networks for composition discovery and substitution purpose.

References

[1] R. Hull, J. Su: Tools for Composite Web Services: A Short Overview, SIGMOD Vol. 34, pp. 86-95 (2005).

[2] B. Medjahed, A. Bouguettya: A dynamic Foundational Architecture for Semantic Web Services, Distributed and Parallel Databases, Vol. 17(2) (2005).

[3] Y. Taher, D. Benslimane, M. Fauvet, Z. Maamar: Towards an Approach for Web services Substitution, IDEAS (2006).

[4] B. Benatallah, M. Dumas, Q. Z. Sheng, A. H. H. Ngu: Declarative Composition and Peer-to-Peer Provisioning of Dynamic Web Services, ICDE (2002).

[5] L. d. F. Costa, O. N. Oliveira, G. Travieso, r. A. Rodrigues, P. R. V. Boas, L. Antiqueira, M. P. Viana, L. E. C. d. Rocha: Analyzing and Modeling Real-World Phenomena with Complex Networks: A Survey of Applications, arXiv 0711.3199 (2008).

[6] E. Christensen, F. Curbera, G. Meredith, et al.: Web Services Description Language (WSDL) 1.1, http://www.w3.org/TR/wsdl (2001).

[7] R. Akkiraju, J. Farrell, L. L. J. Miller, et al.: Web Service Semantics - WSDL-S, http://www.w3.org/Submission/WSDL-S/ (2005).

[8] J. Farrell, H. Lausen: Semantic Annotations for WSDL and XML Schema, http://www.w3.org/TR/sawsdl/ (2007).

[9] D. Martin, M. Burstein, J. Hobbs, et al.: OWL-S: Semantic Markup for Web Services, http://www.w3.org/Submission/OWL-S/ (2004).

[10] R. Nayak, B. Lee: Web Service Discovery with additional Semantics and Clustering, ICWI (2007).

[11] A. Konduri, C. Chan: Clustering of Web Services Based on WordNet Semantic Similarity, In: NCS, Vol. 3, pp. 868-877, (2007).

[12] H. Kil, S.C. Oh, E. Elmacioglu, W. Nam, D. Lee: Graph Theoretic Topological Analysis of Web Service Networks, WWW Vol. 12, No. 3, pp. 321-343 (2009).

[13] J. Gekas, M. Fasli: Employing Graph Network Analysis for Web Service

Composition, Agent Technologies and Web Engineering, IGI Global (2008).
[14] H. N. Talantikite, D. Aissani, N. Boudjlida: Semantic annotations for web services discovery and composition, CSI, Vol. 31(6), pp. 1108-1117 (2009).
[15] S. V. Hashemian, F. Mavaddat: A Graph-Based Approach to Web Services Composition, SAINT, pp. 183 - 189 (2005).
[16] L. Dekar, H. Kheddouci: A Graph b-Coloring Based Method for Composition-Oriented Web Services Classification, In: ISMIS, pp.599-604, (2008).
[17] C. Cherifi, Y. Rivierre, J.F. Santucci: WS-NEXT, a Web Services Network Extractor Toolkit, In: ICIT, (2001).
[18] M. Paolucci, T. Kawamura, T.R. Payne, K.P. Sycara: Semantic Matching of Web Services Capabilities, In: ICWS, pp. 333-347 (2002).
[19] C. Cherifi, V. Labatut, J.F. Santucci: Web Services Dependency Networks Analysis, In: ICNMI, pp. 115-120(2010).
[20] C. Cherifi, V. Labatut, J.F.Santucci: Benefits of Semantics on Web Service Composition from a Complex Network Perspective, F. Zavoral et al. (Eds), NDT 2010, CCIS Vol. 88, pp 80-90, Czech Republic (2010).
[21] C. Cherifi, V. Labatut, J.F. Santucci: Topological Properties of Web Services Similarity Networks, M. Sarrafzadeh, P. Petratos, (Eds.), Strategic Advantage of Computing Information Systems in Enterprise Management, pp. 105-117 (2010) .
[22] InfoEther, BBN Technologies: Semwebcentral.Org, http://wwwprojects.semwebcentral.org/
[23] C. Cherifi, V. Labatut, J.F. Santucci: On Flexible Web Services Composition Networks, CCIS Vol. 166, pp. 45-59 (2011).
[24] C. Aksoy, V. Labatut, C. Cherifi, J.F. Santucci: MATAWS: A Multimodal Approach for Automatic WS Semantic Annotation, S. Fong et al. (Eds), NDT 2011, CCIS Vol. 136, pp.319-333 (2011).
[25] A. Hess, E. Johnston, N. Kushmerick: Assam: A Tool for Semi-Automatically Annotating Semantic Web Services, In: ISWC, (2004).

Chapter 5

UTILIZATION OF STUDENTS' LEARNING STYLES FOR E-LEARNING EFFICIENCY IMPROVEMENT

Marek Woda, Konrad Kubacki-Gorwecki

Summary

Authors try to attract attention to important factor which often cripples e-learning efficiency and is the lack of individual approach to a student. In order to improve knowledge absorption students have be treated accordingly to their specific and preferred learning style. The paper discusses the results of studies in the literature for modeling learning styles. It presents the basic aspects of the problem, and selected learning models. The paper ends with a detailed description of the selected model suitable to be applied in e-learning system - including the manner of its recognition and how to store it in a computer system.

Introduction

The learning process in order to be effective needs to ensure a student certain conditions: quiet and comfortable environment for knowledge acquisition and aptly prepared teaching material. It is commonly believed that the ideal learning environment is characterized mainly by: a quiet room, a large desk with comfortable chair, a decent lighting, and books written in an accessible way appropriate to the age of the pupil. Such learning environment is being met by every young person starting his education in a primary school and continuing it through a high school, and finally at university. It is common knowledge that the ideal conditions for learning do not always meet the expectations of all learners. It is not always the case for everyone - no matter age, gender and psycho-emotional state at the moment, that the best timing to assimilate the knowledge is time between 8 am and 3 pm, or in a group 20-30 people, sitting in at a school desk. Over time, each person finds their own optimal learning environment, allowing maintain optimal concentration levels and to provide excitatory stimuli for the own characteristic.

However there are aspects that in a traditional teaching seem to natural, and which ought to be introduced into the e-learning, in order increase student friendliness and increase learning effectiveness. Experts [14, 16] suggest that the effectiveness of the on-line courses is impacted by given below factors, such as:

- ***Acquaintance of the other course participants*:** It is a typical aspect of the effectiveness of each learning environment, and it results directly from the natural human need to communicate and stay together. In the distant learning, communication between participants is substituted by the following means: forum, chat, instant messaging or e-mail. In recent years, the relationships between Internet users are being built around social networking sites (e.g. Facebook) [11], which may also serve as a communication platform – to introduce online training participants one another.
- ***Clear objectives of a course***: It is imperative, not only related online systems that courses should be created on the basis of the framework plans and have defined minimum standards to comprehend. Each student participating in a course should be familiar with the presented at the beginning, its requirements and objectives. A very important aspect is also a simple and easy to grasp assessment system - and agreed a minimal score to pass [13].
- ***Cooperation and competition***: A very important aspect in teaching is to provide participants possibility to work in groups, as well as the opportunity for discussion, criticism and self-assessment of their skills against the group. In the case of traditional teaching these things are natural and easy to implement. Students willing to work in groups create their internal structures and divide their responsibilities. Each group has a teacher that supervises their work - acting as the moderator and consultant - he can keep track and assess the progress of work of each group. At the same time, each participant during the lesson can evaluate the progress against own group or/and against the whole class. Cooperation and competition provide many incentives to deepen the knowledge and for gaining new skills.

Working in groups, in the case of e-learning is much more difficult because of communication problems, distributed responsibility and work synchronization. Nonetheless it is not impossible - but it is necessary to deliver adequate tools to facilitate the exchange of information, communication and rapid assessment of the progress within work group. The rivalry element within the entire group is relatively simple to achieve. Each student can compare the quality of his expressions, work

and test results against the performance of other participants, if only course supervisor decide to publish such, even aggregated, results or rankings.

- ***Technical Support***. The course participants on distant learning platforms are people with different levels of expertise and computer skills. Therefore, it is an essential element for online courses to provide continuous technical support for both students and teachers (who should not perform any administrative functions related to learning platform) and act only as knowledge experts or trainers within their domain of competence.
- ***Flexible learning environment***. A remote learning environment should adapt to the individual needs of students, their interests and priorities, while standing guard over the assumptions and objectives for the course taken. It is very difficult to achieve, in particular, if we think about the implementation of an automatic approach, without any involvement of a third party. What nowadays platforms provide, is the ability to build an individualized learning and repetition schedule of teaching material acquisition - but at a rate and within a time frame specified by a supervisor. A desirable solution would be systems capable of identify areas of interest of individual students, adapting the course material in such a way as to make it more interesting from the viewpoint of each participant.
- ***An extensive knowledge base***. Every online course should take advantage of benefits that come from the medium that links current material with external links to additional content e.g. in the Internet. With this solution, students, may quickly (materials are refined, categorized, pre-assessed) deepen even more their knowledge in particular topic related to the subject of training.

The above-mentioned elements of an effective learning environment should be treated as a base, a kind of skeleton for building online learning platforms [19]. Unfortunately, most commonly used tools available today do not implement all of these requirements, which mean that online teaching is not as effective as it could be [18]. The main reason for failures of modern tools may be found in relatively high, the potential gains from sharing the platform, cost of their preparation and maintenance. Generally speaking profits are primarily being generated by courses that are often in use and repeated – a specific training, due to a smaller audience (target group) is associated with far higher costs of teaching materials creation and technological platform maintenance.

In addition to the technical conditions mentioned above, the effectiveness of the learning in both classroom and distant approach is affected by factors such as the complexity of the learning material, quantity and diversity of material in a lesson, lesson duration, time of knowledge acquisition and the initial knowledge level of a student. In the given context, a platform of distance education is much more flexible than the equivalent of classroom learning. It allows a student for independent decision making about how long and when will he learn. It does not require the presence of the teacher when the student begins to learn - which allows the use it around the clock. However, traditional learning, has a huge advantage over e-learning, a teacher can observe not only the performance of the student (as in the case of e-course), but also to supervise and oversee the progress of knowledge acquisition and if necessary - to intervene, correct and explain. All these features are missing in on-line trainings.

Student's Evaluation

Any process, including teaching requires the tools to assess the correctness of its course. The main evaluative element in the learning process is to assess the acquisition of knowledge from the subject of training by participating students. At each stage of education, we are subjected to various forms of verification of our knowledge, and then receive grades, which in a discreet manner determine our expertise level of the subject. The emergence of new forms of learning such as distance learning, has forced the emergence of new measurement tools and evaluation of educational progress, relevant to the technical capabilities and taking into account the specificities of place and time.

Assessment in Distant Learning

The subject of e-learning is very interesting to the professionals from various branches including psychologists, educators, theorists and practitioners in methods of teaching [5, 14]. Many of them claim that e-learning methodology requires changes in both teaching methodologies and approach to its assessment [1]. However, a typical solution is to remodel the traditional assessment and adopt its principles into new situation. The main differences in assessment by a teacher and e-instructor result from replacement of direct contacts between a student and a teacher by indirect one through Internet medium. One can define two basic differences in traditional approach to assessment versus and remote one:

- The teacher cannot observe the student's progress in learning.
- The teacher assesses only the finished '*product'* that is submitted for assessment.

The first point is arguable. Student's actions can be tracked through the activity records, recorded automatically by e- platform in the logs. This information, however, does not provide us answers to the question whether the student actually devote time to analyze the content of the course, or just opened a browser window on the website of the course, and then took up other activities. Moreover, a teacher besides simple activity records is not being given the real picture what is the actual behavior of a student. This information is "an added value" to evaluate educational process - resolved task, ended up test in traditional teaching is used in evaluation and tracking of student's learning progress and accomplishments gained so far.

Another problem associated with the assessment of learning, is to verify if the task was performed by the student on his own, or with somebody's help. The absence of a teacher, while solving problems / writing tests becomes an extra incentive for use of additional materials, or somebody's help or even to commit plagiarism (conscious or unconscious). In the distant learning, the teacher must adapt the forms of assessment to features of given technological platform. In e-learning it is almost impossible to use such traditional methods like oral response, unannounced test, or student presentations for the group. The assessment model in the e-learning is usually very limited to tools such as:

- ***Student's Self-assessment*** – it requires form a student a lot of self-discipline, and a fair approach to the problem (thus the evaluation requires the proper pattern before it starts).
- ***Automatic Evaluation*** - such assessment may be subjected only 'closed' tasks that previously given answers. This method may prove to be ineffective if the student will give substantially correct answer which was not envisaged by the author and thus it cannot be validated.
- ***Group Assessment*** - a sense of anonymity on the Internet and the belief that we will never meet the other members of the course promotes this type of valuation. Using this approach, students can assess the involvement in the work of other learners, and reviewing the substantial value of their statements in the discussion forum. This method is not without a flaw, because it leaves room for abuse - which can lead to excessively high assessment of selected students, or the reverse situation - to the detriment of the persons concerned.

- ***Teacher's Evaluation*** - often used in traditional teaching, in e-learning makes many of aforementioned problems and it is limited by the availability of a teacher. E-learning courses are characterized by a large number of participants, which requires the teacher to devote much time to this type of assessment.

Due to the scale of the problem, *Automatic evaluation* rules in teaching via Internet. It is the least absorbent for the teacher (who turns out to be the narrowest bottleneck of e-learning), and the most objective of all mentioned above. Preparation of multiple-choice test in today's tools is an easy task. At the same time to check and evaluate it - is done automatically, and a student receives almost immediate feedback. In addition, the test results easy to compare with each other and provide clear information which lesson units caused the greatest problems for course participants. The main drawback *automatic evaluation* is the potential possibility to pass exam without mastering the course content - by simply guessing the answers, or filling up by a random selection resulting in a correct sequence of responses. However, assuming the students honesty, this method can be considered good, enough for traditional and e-learning.

Learning Styles

Learning is one of the fundamental concepts of psychology. By a term 'learning', one can understand the concept of a process of acquiring knowledge, skills, leading to permanent changes in the behavior of the learner. Whether there is a learning process, we conclude on the basis of observed changes. The effectiveness of learning depends on many factors [10, 19], both external - discussed earlier, as internal and individual, such as short and long term memory, attention, motivation, interests and abilities.

People learn in different ways. In this process, we use all the senses, but over time some of them specialize better than others, which makes it much easier to absorb a new material with the use of those senses. Theoretically, at each stage of education, we should use our best sense and style of learning. In a practice, it appears that it is impossible to adjust the content of didactic forms of communication to the preferences of a group of students at the same time - if we have only one room and one teacher. Only thanks to Internet, and e-learning materials, the postulate of adaptation to individual may come true. The starting point for its implementation is a correct diagnosis of learning styles. This is essential to e-learning become effective. In the [5, 9] one can find some

classifications of learning styles that are based on different aspects of learning process, among others:

- *Allinson and Hayes' Cognitive Styles Index (CSI),*
- *Apter's Motivational Style Profile (MSP),*
- *Dunn and Dunn model and instruments of learning styles,*
- *Herrmann's Brain Dominance Instrument (HBDI),*
- *Honey and Mumford's Learning Styles,*
- *Index of Learning Styles (ILS) - Felder, Silverman, Solomon,*
- *Kolb's Learning Style Inventory (LSI),*
- *Memletics Learning Styles, (MLS),*
- *Myers-Briggs Type Indicator (MBTI),*
- *Paragon Learning Style Inventory (PLSI)*

Kolb's Learning Style Inventory

One of well-known and the most frequently quoted classifications is David Kolb's learning styles Inventory (LSI). David Kolb, a professor at Case Western Reserve University, developed a theory of learning, and suggested four styles of learning. All four styles are based on the basic activities in the learning process: sensing, observation, thinking (reasoning) and acting. Their dependencies and relationships are shown in the figure below (Figure 1).

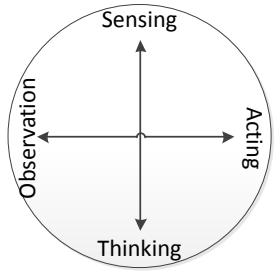


Figure 1: Activities and dependencies

Learners are being categorized depending on which phase is dominant during learning process in the following manner:

- ***(Diverging) sensing and observation - 15 - 20% of the population***. This group includes people who like to ask questions '*why*'. Their strengths are a rich imagination, the ability to look at a problem from different angles. They easily create new ideas, have wide cultural interests.

- ***(Assimilating) observation and thinking - 35 - 40% of the population***. This group includes people who like ask questions 'what' type. (What's the cause?). They are interested in abstract ideas, have the ability to create theoretical models, excel in inductive reasoning.
- ***(Converging) action and thinking - 30% of the population.*** This group includes people who like to ask questions 'how'. They are able practically use their ideas and focus on deductive reasoning. People in this group have rather narrow scope of interests.
- ***(Accommodating) action and perception - 15 - 20% of the population***. This group includes people who like to ask questions '*if*' type? They are ready to take risks, respond quickly to unexpected circumstances, in an intuitive approach in solving problems. The strongest characteristic is their specific approach to action.

According to [8], recognition of learning style is done by filling out the test, containing twelve sentences, prepared answers that allow classifying a learner. Unfortunately, the prepared test is copyrighted, and must be purchased to be used later on.

VAK Model

Another very popular and widely used model of learning style is the VAK model (Visual, Audio, and Kinesthetic). This classification, proposed by Neil Fleming, divides the population into three classes [1]:

- ***Visual learners*** - prefer quiet and order around them other case they have difficulty maintaining concentration. They well remember the colors, drawings, and faces and the position of objects in space as well. They have a problem with remembering names, and titles. Visual learners remember best what they see in the form of text, video, graphics, and plots. They like to make handwritten notes, prefer the visual arts. During learning are not bothered by surrounding sounds.
- ***Auditory learners*** like to talk, sing, and whistle. They learn by listening to lectures, reading aloud, and leading discussions. They remember well: music and the conversations however may have problems with reading the graphic forms, such as maps, geometry. They prefer to speak about the action rather than watching it. They require silence to learn, music and noise do not allow focus them. They have a great ease of language learning.
- ***Kinesthetic learner*** feels best in motion. They become tired sitting at the desk, listening to a lecture. During the speech they often

gesticulate. They require a break between learning sessions. They like to work in a group. Any movement around them may be disturbing for them.

The incidence of individual learning styles in the population is shown in figure above (Figure 2) [7]. One can notice that the dominant style is Visual learner, and a negligible share of the population have Kinesthetic. This is information quite important and the same time good from the perspective of e-education, because Kinesthetic is the most demanding kind to satisfy in online education.

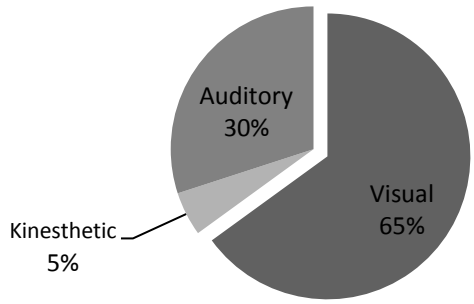


Figure 2: Learning style distribution in learners' population

Because of its simplicity, the VAK model is widely recognized in education. Virtually everyone has a general intuition about yourself to qualify for one of the categories proposed by this model. These characteristics have decided that further consideration will be used to define student's model in online educational system.

Student's Model in a E-System

Prior enrollment in the e-learning system every student should be obliged to solve a psychological test, which allows determining his dominant learning style. The multiple choice test has been developed by authors on the basis of [3, 4]. Here is a sample test:

Instructions

In answering questions 1 -13 select the answer that best reflect your preference. You can select more than one answer, if this does not suit your perception. Leave the blank if it does not apply to your person.

1. *You are about to give guidance to the person standing next to you. It stopped at the hotel in the city and wants to visit your house later. He has a rented car. Do you?*

(a) Draw a map on paper
(b) Tell him how to get to you
(c) Write the instructions (no maps)
(d) Give him a lift with your car

2. *You are not sure whether you should write the word "dependant" or "dependen" Do you?*
   (a) Check the dictionary
   (b) You can see the word in your mind and choose the correct one
   (c) It sounds in your mind
   (d) You write both versions on paper and choose one
3. *You just have received information about your participation in a trip around the world. Your friend is also interested in it. Do you?*
   (e) To call him immediately and tell about it
   (f) Send him a copy of the information you received
   (g) Show off him this tour on the world map
   (h) Share with him your plans for each site you visit.
4. *Would you like to cook something special as a treat for your family. Do you?*
   (a) Cook something familiar without the need for provision
   (b) Flip fingers cookbook looking for ideas with drawings
   (c) The right to resort to a cookbook where you described are good rules
5. *A group of tourists has been assigned to you to find information on nature reserves or parks. Do you?*
   (a) Drive them to a reserve or a park
   (b) Show them slides and photographs
   (c) Would give them a brochure or a book on reserves.
   (d) Would talk with them about the reserves or parks
6. *You're going to buy a new CD player. What, besides money, would have the greatest influence on your decisions?*
   (a) Salesman telling you all about what you want to know
   (b) Reading the specifications of the cd
   (c) Check up the features and listening to music
   (d) Whether the cd looks really smart and trendy
7. *Remember the time in your life when you learned how to play, for example, a new table game. How do you learn best? By:*
   (a) Visual traces - drawings, diagrams, graphs
   (b) Reading the instructions
   (c) Listening to someone
   (d) Doing it or trying to
8. *Having trouble with your vision. Do you prefer to ophthalmologist:*
   (a) Told you what is wrong
   (b) Shown a diagram informing what it is not right
   (c) Used the model to show you what is not right

9. *You want to learn how to use a new computer program. Do you:*
   (a) Sitting in front of the keyboard and start exploring the possibilities of the experiment
   (b) Read the instructions that tells how to use a computer program
   (c) You are calling a friend and ask him questions
10. *You're in a hotel and you have a rented car. Would you like to visit friends, whose address / location you do not know. Do you need your friends:*
   (a) Draw a map on paper
   (b) Hear how you go
   (c) Write you how to get there (without a map)
   (d) Pick you up from the hotel
11. *Whatever the price, which would have the greatest impact on your decision to buy a particular textbook?*
   (a) The earlier use of this manual
   (b) Talk about it with a friend
   (c) Quickly reading parts
   (d) Your impression when you watch it
12. *New touring cinema came to town. What would be the biggest influence on your decision to go (or not)?*
   (a) Hear the radio audition before
   (b) Read the review about it
   (c) See information about it
13. *Do you prefer a lecturer or teacher who likes to use?*
   (a) Book, writing, reading
   (b) Graphs, charts, transparencies
   (c) Laboratory classes, practical classes,
   (d) Speaking, discussions

Each answer in the test, was assigned a learning style from VAK model (Matching types assigned to answers are given in Table 1). Assessment, which style is dominant, for a student, requires counting the number of responses. Prevailing number of answers points out student's dominant learning style. The prepared system may be useful, but for the most accurate classification – calculation of answers belonging to other styles may be advisable. With this information, it will be possible to match student type and learning strategy that best fit to the student's learning style.

Table 1: Matching VAK model types and answers

Question	Answer A	Answer B	Answer C	Answer D
1	V	A	V	K
2	V	V	A	K
3	A	V	V	K
4	K	V	V	-
5	K	V	V	A
6	A	V	K	V
7	V	V	A	K
8	A	V	K	-
9	K	V	A	-
10	V	A	V	K
11	K	A	V	V
12	A	V	V	-
13	V	V	K	A

Student's model in the e-learning system may be represented by a vector: [preferred style of learning, student proficiency, the initial test result, profile of an individual learning style, the recent results] - Fig. 3. Preferred learning style, and the individual learning style profile – these can be derived from aforementioned VAK test. Other vectored components can be determined during e-system operation.

Student proficiency is determined empirically, it is the level of intellectual abilities of the student. On this basis, an algorithm can route a student on a minimal, average, or advanced educational learning path of a course [18].

This parameter is assigned and modified during the tests carried out after each portion of material that has been acquired by a student. It depends on the weighted average of its actual value and the result achieved in the tests (1).

$$W_z = 0.8 \cdot W_z + 0.2 \cdot T_i \qquad (1)$$

W_z – actual value of learning progress

T_i – result i-th test

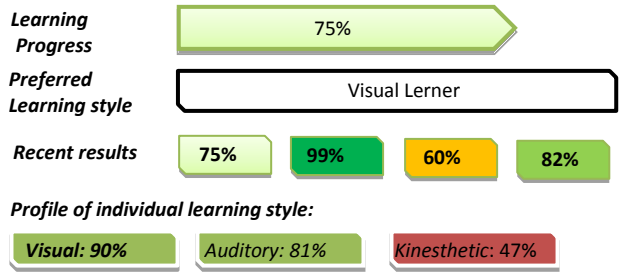


Figure 3: Student's profile - vector

Initial test result is defined by the percentage rating the student's knowledge of course content, prior to it, allowing to choose the path of learning appropriate to their knowledge of the subject - and best use of time student. Recent student results shape his results vector. It allows in time for detection a trend of grades received by a student, and for judgment if learning strategy adopted is right for him.

Research Results

As a method for testing purposes - simulation was chosen. Test results were obtained during the experiment in which two approaches of teaching were compared:

- Standard, implementing a single, universal learning path
- Adaptive, using recognized students' learning styles.

There has been generated an example of the course, consisting of 4 teaching units, each at three levels of difficulty and three styles of presenting the material. In addition, as a part of the course one quiz was included, two control tests and a final exam. Each lesson contains a description of the content in the form of meta-data. Student behaviour was simulated by using a user agent User Agent [20]. The test population consisted of 1000 randomly generated profiles of students.These profiles were generated with the following assumptions:

- Level of proficiency and initial test results were generated randomly according to Gaussian distribution ($\mu = 63$, $\sigma 2 = 19$). The standard deviation 1, the expected value (in this case symbolizing the percentage of test result) was 63%. This distribution is largely similar to the real distribution of test ratings.
- Preferred learning style of students was drawn from the proportions set out in Chapter 2.2.2: 65% *visuals*, 30% *auditory*, 5% *kinestics*. After the main learning style was determined, three components were generated – in order to match VAK, so that these fit into the selected

student's learning style (the main style got a value between 70 - 100%, the other two were in the range 20 - 70%).

- Student Test result was generated by rules that described the model (results above 100% were normalized):
  $result[\%]=100*Z_{SL}* P_z+ R$ (2)
  Z_{SL} - Compatibility of the student's learning style and presentation style lessons
  P_z – Student level of proficiency
  R – Random factor [-10; 10]

The experiment was carried out on two courses. The first represents a classic Internet course, is being represented by a single learning path. Teaching materials are described in the meta-data were: 80% text, 10% images, 10% video and 10% audio. The second course is compliant with the assumptions of serving different content to students represented by a three VAL styles – it consists of 3 proficiency levels, and lessons were tailored to be suitable for VAK styles. The results received after first course (using standard approach) will be called the reference level.

Important information from the viewpoint of each educational system is a course pass rate. Assumptions about the adaptive system say that students' grades are not so important in comparison to knowledge absorption itself, and ability to allow as many students as possible to end the course with a positive result. But it is clear that this objective is achieved through the same methods as a result of raising grades. Both these aspects are leading to one thing - deepening of knowledge by a student. The chart below shows the results of the adaptive system against the conventional educational strategies. Threshold pass rate was fixed at 50%. The so-called "underperforming" student group consist of students are not able - in the test environment - to achieve results above 40%. This group consists of 26 persons, representing 2.6% of the total population.

Improvement of course pass rate, after the application of learning styles has increased by nearly 9 percentage points, which seems to be very good result. This difference will increase more if the test population will not reflect the real life distribution of preferred style in a society.

Test results shown in Figure 5 takes into account only *kinestics* (standard course material meets their expectations minimally). All *kinestics* represent only 5% of the population (50 students). You will notice that the lack of content adaptation for *kinestics* in the course, rendered course pass rate low - after the exclusion from underperforming students group - course pass rate reaches only 70%, in comparison to 98% percent using the adaptive approach.

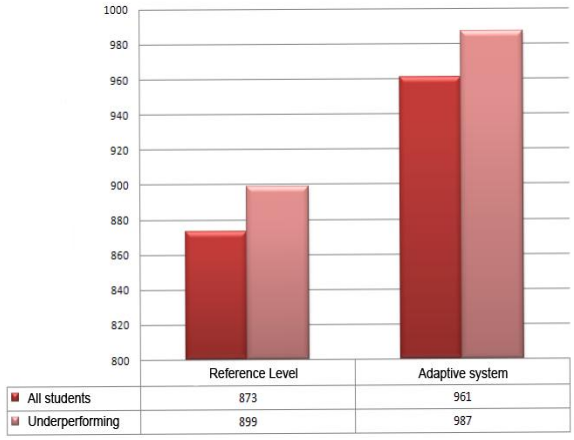


Figure 4: Pass rate chart (1000 students)

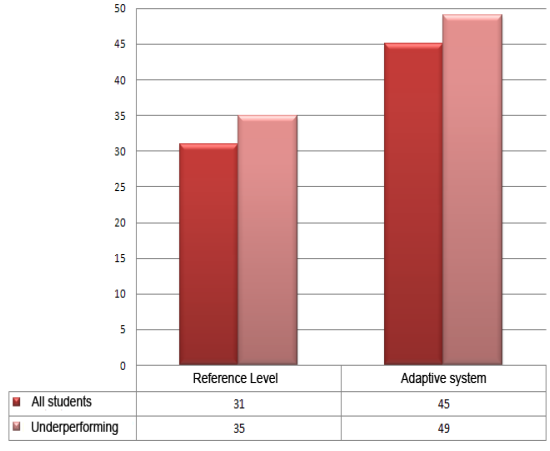


Figure 5: Pass rate chart – *kinestics* (50 students)

Conclusions

Problems related to the effectiveness of teaching in distant learning are well known and often described in the literature [7, 15]. Authors presented a number of principles and rules which were supposed to make e-learning environment more student-friendly and above all more efficient. The main focus of these considerations was put on the adaptability of the course content to suit needs and preferences of the individual student. Psychologists and educators have developed many models, illustrating the techniques and styles of learning, which

familiarity with, promotes effectiveness of the knowledge acquisition. One of models, which is the most frequently used in conventional teaching, and in authors opinion can greatly contribute in increasing of teaching in e-leaning systems is VAK model. Research on learning style showed that the majority of the population constitutes Visual learners (65%) and then Auditory (30%). So it is recommended to focus on these two styles and try to prepare appropriate e-content that matches learning style of each group.

The recognition of a student's learning style is not a trivial task [17]. The problem lies in that in human population, there are no two other persons alike, everyone differs – but there are similarities among individuals. None of the publications, neither from pedagogy nor from psychology of teaching, state unequivocally that a student must represent only one teaching style. It is quite the contrary, the variety of learning styles models [10], only confirms this fact. They are based not only on the method of learning material presentation, as it is in the case of VAK model, but also take into account the process of deduction, emotional states, the impact of environment, attitude to learning - and all the combinations between them. This means learning style even properly recognized could not be good enough to encourage s student to learn. None the less it may at least help to better acquire the knowledge provided by us.

Despite of difficulties with recognition learning style of a student and matching it with teaching strategy, one problem still remains - what to do next with successfully detected learning style. Although we can determine, better or worse, the student's preferred learning style, there are still not present widely available educational systems that implement the adaptability of the course material to the student type. Main cause of this situation primarily are high costs associated with both preparation of appropriate tools (e-learning platforms), and teaching materials (e-content) that would meet the requirements of the model. Taking into account simple VAK model, and taking assumption that the material is delivered on three levels for every lesson unit one has to prepare as many as nine different versions of the content. Hypothetical solution for the problem of the large amount of lesson variants to prepare, can be shared lesson objects, such as LOM, SCORM, which allowing reusing lesson units (called Learning Objects) for various courses on different platforms [2]. Thanks to them, there exists a sort of confidence that it will be easier preparing diversified teaching content suitable for an individual.

Chapter 5

Analysis of simulation results received after learning style application to students, gave more than satisfactory results. As expected, when the content matches student's style, the results are far better than for a standard approach (where all students are served with the same content). The experiment was based only on simulation activities. The nature of this approach is high predictability of getting the same results, even despite the introduction of some random factors. Human behavior is difficult to write in an algorithmic way, and therefore presented the results of simulation will not necessarily coincide with the possible results of a similar experiment involving real students. Characteristic of man is unpredictability. Theoretically, you can simulate human behavior in a computer system - but such a large dose of randomness even in controlled environment still might distort the experiment image.

References

[1] Al-Dahoud A., Walkowiak T., Woda M. "*Dependability aspects of e-learning systems*". Proceedings of International Conference on Dependability of Computer Systems, DepCoS - RELCOMEX 2008, Szklarska Poręba, Poland, 26-28 June, 2008 IEEE Computer Society [Press], p. 73-79.

[2] Di Bitono, "*Multi-agent Architecture for Retrieving and Tailor Los in SCORM Compliant Distance Learning Environment*", University of Bari

[3] Flemming N. VARK Questionnaire (accessed Dec 2010) http://www.vark-learn.com/english/page.asp?p=questionnaire

[4] Fleming N. D., "*Teaching and Learning Styles: VARK Strategies*". ISBN 0-473- 07956-9, Honolulu Community College, 2001.

[5] Graf S., "*Adaptavity in Learning Management Systems Focusing on Learning Styles*", Ph.D. Thesis, Technische Universitat Wien, 2007

[6] Kanninen E. "*Learning Styles and E-Learning*",Master of Science Thesis, Tampere University Of Technology, 2009

[7] Kar-tin L.: "*E-Learning: The Quest for Effectiveness*", Malaysian Online Journal of Instructional Technology Vol. 2, No.2, p. 61-71, 2005

[8] Kolb A., Kolb D. "*The Kolb Learning Style Inventory*", Case Western Reserve University, 2005

[9] Kinshuk, Taiyu Lin, Improving mobile learning environments by applying mobile agents technology, Massey University, Palmerston North, Third Pan-Commonwealth Forum on Open Learning, 2004, New Zealand

[10] Manochehr N.-N. "*The Influence of Learning Styles on Learners in E-Learning Environments: An Empirical Study*" Information Systems Department, Qatar University, CHEER Virtual Edition Volume 18, 2006

[11] Mitrano T. "*Thoughts on Facebook*", Cornell University, 2006, www.cit.cornell.edu/policies/socialnetworking/facebook.cfm (accessed Dec 2010)

[12] Pei-Shan T., Dengchuan C., Yao-Jen F. "*A Study of Cognitive Style on Design Performance*", Proceedings of IASDR, Korea, 2009

[13] Rudak L.,"*Wybrane zagadnienie e-edukacji*", ISBN: 978-83-204-3354-4, WNT, 2009, in Polish

[14] Rudak L.,"Ocenianie w e-edukacji vs. Ocenianie tradycyjne", Warszawa 2008, (in Polish)

[15] Shivkumar S. "*Strategies for improving eLearning effectiveness*", International Workshop on e-Learning for Adult Continuing Education, Cochin University of Science and Technology, 2006

[16] Soomro S., "*E-learning, experiences and future*", Vukovar 2010

[17] Villaverde J. E., Godoy D., Amandi A. "*Learning styles' recognition in e-learning environments with feed-forward neural networks*",ISISTAN Research Institute, UNICEN University, Campus Universitario, Paraje Arroyo Seco, Tandil, 2006, Argentina

[18] Woda M. „*Zarządzanie procesem uczenia w komputerowych systemach wspomagających nauczanie*". New Media in Education, Wrocław. Oficyna Wydawnicza Politechniki Wrocławskiej, 2008. p. 147-155. (in Polish)

[19] Woda M. "*Effective knowledge acquisition by means of teaching strategies*". Computational intelligence and modern heuristics. Vukovar: In-Teh, 2010. p. 99-106.

[20] Woda M., Kubacki-Gorwecki K.: „*System e-nauczania w oparciu o system agencki*", in review for Nowe Media w Edukacji 2011 (in Polish)

[21] Zając M. „*E-learning "szyty na miarę, czyli o indywidualizacji w nauczaniu online*" e-mentor, Warsaw School of Economics (accessed Dec 2010) (in Polish) http://www.e-mentor.edu.pl/artykul/index/numer/17/id/362

Chapter 6

HUMAN ASPECTS OF DISCRETE TRANSPORT SYSTEMS DEPENDABILITY AND MANAGEMENT PROBLEMS

Jacek Mazurkiewicz, Tomasz Walkowiak

Summary

The chapter is focused on the human resource influence on dependability of discrete transportation systems (DTS). The human resource means the driver of the vehicle. We add him/her as a new element of the system description. The dependability means the combination of the reliability and functional parameters of the DTS. This way the analysis of the DTS behavior seems to be more sophisticated. The unified containers transported by trucks with the set of time-type assumptions are the essence of the system discussed. The proposed method is based on modeling and simulating of the system behavior. The income of containers is modeled by a stochastic process. Each container has a source and destination address. The central node is the destination address for all containers generated in the ordinary nodes. We also propose the heuristic management approach as well as the functional metric for DTS and we test the example system based on the real data.

Introduction

Administration of a large transport system is not a trivial task. The transport systems are characterized by a very complex structure. The performance of the system can be impaired by various types of faults related to the transport vehicles, communication infrastructure or even by traffic congestion [1]. It is hard for human (administrator, owner) to understand the system behaviour. To overcome this problem we propose a functional approach. The transport system is analysed from the functional point of view, focusing on business service realized by a system [16]. The analysis is following a classical [16]: modelling and simulation approach. It allows calculating different system measures which could be a base for decisions related to administration of the transport systems. The metric are calculated using Monte Carlo

techniques [4]. No restriction on the system structure and on a kind of distribution is the main advantage of the method. The proposed model allows forgetting about the classical reliability analysis based on Markov or Semi-Markov processes [4] - idealised and hard for reconciliation with practice.

The paper presents an analysis of transport system of the Polish Post regional centre of mail distribution (described in section 2). Base on which we have developed the discrete transport system model presented in section 3. The main service given by the post system is the delivery of mails. From the client point of view the quality of the system could be measured by the time of transporting the mail from the source to destination. A driver is a new element of the system description. The pointed the set of states to characterise the actual driver position including formal – law-origin aspects: number of hours he or she can work daily for example. We offer three approaches to system management: based on time-tables, heuristic and focused on soft-computing (section 3). In our opinion the heuristic one seems to be the most adequate to the level of detail discussed in the described work. The quality of the analysed system is measured by the availability defined as an ability to realize the transportation task at a required time (described in section 4).

The post system is very hard to be analysed by a formal model since it does not lie in the Markov process framework. Therefore, we have used a computer simulation [4] described in section 5. Next (section 6), we give an example of using presented model and simulator for the analysis of the Polish Post regional centre in Wroclaw transport system and discussed the performance of developed simulator.

Real Transport System

The analysed transport system is a simplified case of the Polish Post. The business service provided the Polish Post is the delivery of mails. The system consists of a set of nodes placed in different geographical locations. Two kinds of nodes could be distinguished: central nodes (*CN*) and ordinary nodes (*ON*). There are bidirectional routes between nodes. Mails are distributed among ordinary nodes by trucks, whereas between central nodes by trucks, railway or by plain. The mail distribution could be understood by tracing the delivery of some mail from point *A* to point *B* (Fig. 1.). At first the mail is transported to the nearest to *A* ordinary node. Different mails are collected in ordinary nodes, packed in larger units called containers and then transported by trucks scheduled

according to some time-table to the nearest central node. In central node containers are repacked and delivered to appropriate (according to delivery address of each mail) central node. In the Polish Post there are 14 central nodes and more than 300 ordinary nodes. There are more than one million mails going through one central node within 24 hours. It gives a very large system to be modelled and simulated. Therefore, we have decided to model only a part of the Polish Post transport system – one central node with a set of ordinary nodes.

Essential in any system modelling and simulation is to define the level of details of modelled system. Increasing the details causes the simulation becoming useless due to the computational complexity and a large number of required parameter values to be given. On the other hand a high level of modelling could not allow recording required data for system measure calculation. Therefore, the crucial think in the definition of the system level details is to know what kind of measures will be calculated by the simulator. Since the business service given by the post system is the delivery of mails on time.

Therefore, we have to calculate the time of transporting mails by the system. Since the number of mails presented in the modelled system is very large and all mails are transported in larger amounts containers, we have decided to use containers as the smallest observable element of the system. Therefore, the main observable value calculated by the simulator will be the time of container transporting from the source to the destination node. The income of mails to the system, or rather containers of mails as it was discussed above, is modelled by a stochastic process. Each container has a source and destination address. The central node is the destination address for all containers generated in the ordinary nodes. Where containers addressed to any ordinary nodes are generated in the central node. The generation of containers is described by some random process. In case of central node, there are separate processes for each ordinary node. Whereas, for ordinary nodes there is one process, since commodities are transported from ordinary nodes to the central node or in the opposite direction.

The containers are transported by vehicles. Each vehicle has a given capacity – maximum number of containers it can haul. Central node is a base place for all vehicles. They start from the central node and the central node is the destination of their travel. The vehicle hauling a commodity is always fully loaded or taking the last part of the commodity if it is less than its capacity. Vehicles operate according to the time-table. The time-table consists of a set of routes (sequence of nodes

starting and ending in the central node, times of leaving each node in the route and the recommended size of a vehicle).

The number of used vehicle and the capacity of vehicles do not depend on temporary situation described by number of transportation tasks or by the task amount for example. It means that it is possible to realize the route by completely empty vehicle or the vehicle cannot load the available amount of commodity (the vehicle is to small). Time-table is a fixed element of the system in observable time horizon, but it is possible to use different time-tables for different seasons or months of the year. Summarising the movement of the containers in the system, a container is generated with destination address in some of node (source) at some random time. Next, the container waits in the node for a vehicle to be transported to the destination node. Each day a given time-table is realized, it means that at a time given by the time table a vehicle, selected from vehicles available in the central node, starts from central node and is loaded with containers addressed to each ordinary nodes included in a given route.

This is done in a proportional way. When a vehicle approaches the ordinary node it is waiting in an input queue if there is any other vehicle being loaded/unload at the same time. There is only one handling point in each ordinary node. The time of loading/unloading vehicle is described by a random distribution. The containers addressed to given node are unloaded and empty space in the vehicle is filled by containers addressed to a central node. Next, the vehicle waits till the time of leaving the node (set in the time-table) is left and starts its journey to the next node. The operation is repeated in each node on the route and finally the vehicle is approaching the central node when it is fully unloaded and after it is available for the next route.

The process of vehicle operation could be stopped at any moment due to a failure (described by a random process). After the failure, the vehicle waits for a maintenance crew (if there are no available due to repairing other vehicles), is being repaired (random time) and after it continues its journey. The vehicle hauling a commodity is always fully loaded or taking the last part of the commodity if it is less than its capacity.

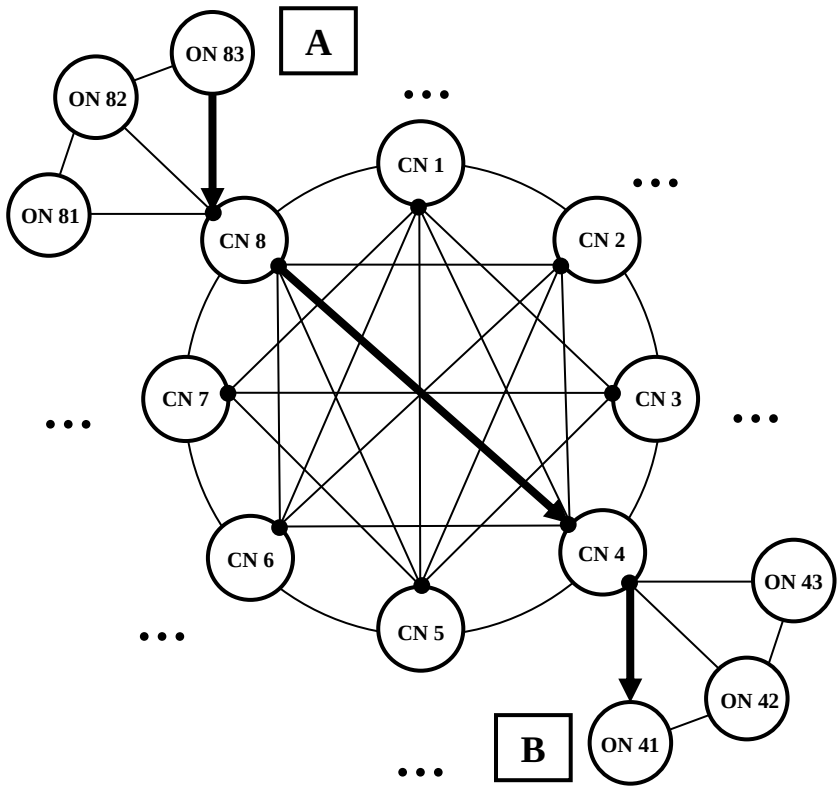


Figure 1: The structure of the Polish Post transportation system

Transport System Model

Overview

The described in the previous section regional part of the Polish Post transport system with one central node and several ordinary nodes was a base for a definition of a formal model of a discrete transport system (*DTS*). Generally speaking users of the transport system are generating tasks which are being realized by the system. The task to be realized requires some services presented in the system. A realization of the system service needs a defined set of technical resources. Moreover, the operating of vehicles transporting mails between system nodes is done according to some rules – some management system. Therefore, we can model discrete transport system as a 5-tuple:

$$DTS = \langle Client, Driver, BS, TI, MS \rangle \qquad (1)$$

where: *Client* – client model, *Driver* – driver model, *BS* – business service, *TI*

– technical infrastructure, *MS* – management system.

Infrastructure

During modelling of technical infrastructure we have to take into consideration functional and reliability aspects of the post transport system. Therefore, the technical infrastructure of *DTS* could be described by three elements:

$$TI = \langle No, V, MM \rangle, \qquad (2)$$

Where: *No* set of nodes, *V* set of vehicles,
MM: maintenance model.

Set of nodes (*No*) consists of single central node (*CN*), a given number of ordinary nodes (ON_i). The distance between each two nodes is defined by the function:

$$distance: No \times No \to R_+. \qquad (3)$$

Each node has one functional parameter the mean (modelled by normal distribution) time of loading a vehicle:

$$loading: No \to R_+. \qquad (4)$$

Moreover, the central node (*CN*) has additional functional parameter: number of service points (in each ordinary node there is only one service point):

$$servicepoints: CN \to N_+. \qquad (5)$$

Each vehicle is described by following functional and reliability parameters:

- Mean speed of a journey
  $$meanspeed: V \to R_+, \qquad (6)$$
- Capacity – number of loaded containers
  $$capacity: V \to R_+, \qquad (7)$$
- Mean time to failure
  $$MTTF: V \to R_+, \qquad (8)$$

Time when failure occurs is given by exponential distribution with mean equal to a value of *MTTF* function,

- Mean repair time
  $$MRT: V \to R_+. \qquad (9)$$

The traffic is modelled by a random value of vehicle speed and therefore the time of vehicle (v) going from one node (n_1) to the other (n_2) is given by a formula

$$time(v, n_1, n_2) = \frac{distance(n_1, n_2)}{Normal(meanspeed(v), 0.1 \cdot meanspeed(v))} \qquad (10)$$

Where Normal denotes a random value with the Gaussian distribution.

Maintains model (*MM)* consists of a set of maintenance crews which are identical and unrecognized. The crews are not combined to any node, are not combined to any route, they operate in the whole system and are described only by the number of them.

The time when a vehicle is repaired is equal to the time of waiting for a free maintains crew (if all crews involved into maintenance procedures) and the time of a vehicle repair which is a random value with the Gaussian distribution ($Normal(MRT(v), 0.1 \cdot MRT(v))$.

Business Service

Business service (*BS*) is a set of services based on business logic that can be loaded and repeatedly used for concrete business handling process. Business service can be seen as a set of service components and tasks that are used to provide service in accordance with business logic for this process. Therefore, *BS* is modelled a set of business service components (*sc*):

$$BS = \{sc_1, ..., sc_n\}, n = length(BS) > 0, \quad (11)$$

The function *length*(*X*) denotes the size of any set or any sequence *X*. Each service component in *DTS* consists of a task of delivering a container from a source node to the destination one.

Client Description

The service realised by the clients of the transport system are sending mails from a source node to a destination one. Client model consist of a set of clients (*C*). Each client is allocated in one of nodes of the transport system:

$$allocation: C \to No. \quad (12)$$

A client allocated in an ordinary node is generating containers (since, we have decided to monitor containers not separate mails during simulation) according to the Poisson process with destination address set to ordinary nodes. In the central node, there is a set of clients, one for each ordinary node. Each client generates containers by a separate Poisson process and is described by intensity of container generation:

$$intensity: C \to R_+. \quad (13)$$

The central node is the destination address for all containers generated in ordinary nodes.

Driver Description

Each driver is describing by actual state of him/her (s_d): rest (not at work), unavailable (illness, vacation, etc.), available (at work – ready to start driving), break (during driving), driving. The number of driver working hours is limited by the labour law. The regulation is rather complicated – depends on a lot of parameters and the number of drivers fixed to the truck for example, but for transport system we analyse we can say that the daily limit for each driver equals to 8 or 9 hours and single driver operates with one truck. This assumption is legal in European Union. So the problem of working hours (w_h) we can solve as follow:

if w_h > *limit* then s_d = rest & w_h = 0,
where *limit* = 8 hours or *limit* = 9 hours,
limit – is the time period of a single change in work.

The single shift can be distinguished as: morning or afternoon. So at 6am for each driver:
if *shift* == morning & s_d == rest then s_d = available.

So at 1pm for each driver:
if *shift* == afternoon & s_d == rest then s_d = available,

The next problem ought to be modelled is the driver's illness state. We propose the following approach:
for every driver at 4am: if s_d == rest,
if *rand()* < d_i then during *x* days (according to the given distribution) the driver is in s_d = unavailable
Where d_i – driver's illness parameter.

Moreover we propose to categorise the driver's illnesses as follow: short sick: 1 to 3 days, typical illness: 7 to 10 days, long-term illness: 10 to 300 days.

We store the daily driver's record. The algorithm to fix the driver to the vehicle is the last part of the driver model:
if no driver – the vehicle does not start,
driver can be chosen if: s_d = available
and w_h + *estimated time of journey* < *limit* * 1.1,
The driver is chosen randomly or by least value:
abs (limit – w_h - *estimated time of journey*).

Legacy Management Solution

The management system (*MS*) of the *DTS* controls the operation of vehicle. It consists of a sequence of routes:

$$MS = \langle r_1, r_2, ..., r_{nr} \rangle . \qquad (14)$$

Each route is a sequence of nodes starting and ending in the central node, times of leaving each node in the route (t_i) and the recommended size of a vehicle (*size*):

$$r = \langle CN, t_0, n_1, t_1, ..., n_m, t_m, CN, size \rangle$$
$$v_i \in No - \{CN\} \quad 0 \le t_0 < t_1 < ... < t_m < 24h \tag{15}$$

The routes are defined for one day and are repeated each day. The management system selects vehicles to realise each route in random way, first of all vehicles (among vehicles available in central node) with capacity equal to recommended size are taken into consideration.

If there is no such vehicle, vehicles with larger capacity are taken into consideration. If still there is no vehicle fulfilling requirements vehicle of smaller size is randomly selected. If there is no available vehicle a given route is not realised. The pros and cons of legacy approach to management problem were discussed in our previous papers. The results are available in [19]. In our opinion the solution does not fit to the level of detail if we have the driver as a part of the system.

Heuristic Management Solution

As it was mentioned in the introduction, we proposed a replacement of legacy management system by a heuristic decision algorithm [19]. The decisions (send a truck to a given destination node) are taken in moments when a container arrives to the central node. The truck is send to a trip if:

- The number of containers waiting in for delivery in the central node of the same destination address as that just arrived is larger than a given number.
- There is at least one available vehicle.
- The simulated time is between 6 am and 22 pm minus the average time of going to and returning from the destination node.

The truck is send to a node defined by destination address of just arrived container. If there is more than one vehicle available in the central node, the vehicle with size that a fits the best to the number of available containers is selected, i.e. the largest vehicle that could be fully loaded. If there are several trucks with the same capacity available the selection is done randomly. On the other hand we observe in the same way the vehicles available in the ordinary nodes. The only difference is the greater level of threshold to initialise the vehicle journey. The restriction for the time of truck scheduling (the last point in the above algorithm) are set to model the fact that drivers are working on two 8 hours shifts.

Soft Computing Management Solution

As it was mentioned in the introduction we also proposed the other replacement of the legacy management system based on a neural network

based [21]. The system consists of a multilayer perceptron to decide if and where to send trucks. The input to the neural network consists of:

$$in = \langle pkc_1, pkc_2, ..., pkc_{npk}, cnc_1, cnc_2, ..., cnc_{npk}, nfv \rangle \qquad (16)$$

Where: *npk* – number of ordinary nodes,
pkc_i – number of containers waiting for delivery in the central node with destination address set to *i*-th ordinary node, *nfv* – number of free vehicles in the central node,
Each output of the network corresponds to each ordinary node:

$$nnout = \langle out_1, out_2, ..., out_{npk} \rangle, \qquad (17)$$

The output of the network is interpreted as follows (for sigmoid function used in output layer):

$$j = \arg\max_{i=1...npk} \{out_i\} \qquad (18)$$

If out_j is greater than 0.5 send a vehicle to node *j* else do nothing. If there are more vehicles available in the central node, the largest vehicle that could be fully loaded is selected. If there are available several trucks with the same capacity selection is done randomly. The neural network decision (send a truck or not and where the truck should be sent) are taken in given moments in time. These moments are defined by following states of the system:

- The vehicle comes back to the central node and is ready for the next trip,
- If in central node there is at least one available vehicle and the number of containers of the same destination address is larger than the size of the smallest available vehicle.

The neural network used in the management system requires a learning process that will set up the values of its weights. The most typical learning in the case of multilayer perception is the back propagation algorithm. However, it cannot be used here since it is impossible to state what should be the proper output values of the neural network. Since it is hard to reconcile what are the results of a single decision made by the management system. Important are results of the set of decisions.

Since the business service realised by transport system is to move commodities without delays, the neural network should take such decisions that allows reducing delays as much as possible. To train neural network to perform such task we propose to use genetic algorithm [18, 21]. Similar approach to training neural network is applied in case of computer games. The most important in case of genetic algorithm is a definition of the fitness function. To follow business service requirements

of transport system we propose following definition of the fitness function calculated for a given neural network after some time (T) (therefore after a set of decisions taken by neural network):

$$fitness(T) = \frac{N_{ontime}(0,T) + N_{ontimeinsystem}(T)}{N_{delivered}(0,T) + N_{insystem}(T)}. \qquad (19)$$

It is a ratio of on-time containers (delivered with 24h and being in the system but not longer then 24h) to all containers (that already delivered $N_{delivered}(0,T)$ and still being presented in the system $N_{insystem}(T)$). The solution described above is very complicated and the first necessary step – learning phase – takes a lot of time and requires a lot of data to create the proper weights and other initial parameters. This is the reason why – taking into account quite good results of the soft computing approach to management [21] – we does not decide to use it the contemporary work.

Dependability Metric

Introduction

The formal model described previously was designed to allow developing a simulator which allows observing the time of transporting each container. Based on this observation several metrics could be defined. As it was mentioned in the introduction we focus here on the service oriented approach [17]. Therefore we propose that the availability to be a key parameter for the evaluation of the quality of the *DTS*.

One can define the availability in different ways, but always the value of availability can be easy transformed into economic or functional parameters perfectly understood by owner of the system. The availability in mostly understood as a probability that a system is up, and is defined as a ratio of the expected value of the uptime of a system to the observation time. It is simple definition but requires defining what does it mean that transport system is working. The similar metric is the acceptance ratio defined in information since as a number of accepted requests to the total number of requests.

Acceptance Ratio

Let introduce the following notation:

- T – a time measured from the moment when the container was introduced to the system to the moment when the container was transferred to the destination (random value),
- T_g – a guaranteed time of delivery, if exceeded the container is delayed.

In [16] we have proposed performance metric – acceptance ratio. It is defined as a ratio of on-time containers (containers for which $T < T_g$) to all containers within a given time of observation $(0, \tau)$.

Within the time period a given number of containers are delivered ($N_{delivered}(\tau)$), a part of them or all delivered on time ($N_{ontime}(\tau)$), but at the end of analysed period time there could be some containers not yet delivered (waiting in the source node or being transported) ($N_{insystem}(\tau)$) and all or part of them being not late yet ($N_{ontimeinsystem}(\tau)$). Taking into consideration introduced symbols the availability could be calculated as the expected value (Monte-Carlo approach) of ratio of on-time containers to all containers:

$$AR_{\tau} = E\left(\frac{N_{ontime}(\tau) + N_{ontimeinsystem}(\tau)}{N_{delivered}(\tau) + N_{insystem}(\tau)} \right) \qquad (20)$$

DTS Simulation

Event-Driven Approach

Discrete transport system described in the section 3 is very hard to be analysed by formal methods. It does not lie in the Markov process framework [2]. A common way of analysing that kind of systems is a computer simulation. To analyse the system we must first build a simulation model, which was done based on the formal model presented in the previous section, and then operate the model. The system model needed for simulation has to encourage the system elements behaviour and interaction between elements. Once a model has been developed, it is executed on a computer. It is done by a computer program which steps through time. One way of doing it is so called event-driven simulation. Which is based on an idea of event, which could is described by time of event occurring and type of an event.

The simulation is done by analysing a queue of event (sorted by time of event occurring) while updating the states of system elements according to rules related to a proper type of event. Due to a presence of randomness in the *DTS* model the analysis of it has to be done based on Monte-Carlo approach [4] what requires a large number of repeated simulations. Summarising, the event-driven simulator repeats *N*-times the following loop:

- Beginning state of a *DTS* initialization,
- Event state initialisation, set time $t = 0$,
- Repeat until $t < \tau$:

- Take first event from event list,
- Set time equals time of event,
- Realize the event – change state of the *DTS* according to rules related to proper type of event: change objects attributes describing system state, generate new events and put them into event list, write data into output file.

DTS Simulation Details

In case of *DTS* following events (mainly connected with vehicles) have been defined: vehicle failure, vehicle starts repair, vehicle repaired, vehicle reached the node, vehicle starts from the node, and vehicle is ready for the next route, time-table (starting the route in the central node). The processing of events of done in objects representing *DTS* elements. The objects are working in parallel. Following types of system elements were distinguished: vehicle, ordinary node, central node, time table. The life cycle of each object consists of waiting for an event directed to this object and then execution of tasks required to perform the event.

These tasks includes the changes of internal state of the object (for example when vehicle approaches the node it is unloaded, i.e. the number of hauled containers decreases) and sometimes creating a new even (for example the event vehicle starts from the node generates new event vehicle reached the node – next node in the trip). The random number generator is used to deal with random events, i.e. failures. It is worth to notice that the current analysed event not only generates a new event but also could change time of some future events (i.e. time of approaching the node is changed when failure happens before).

The time of a new event is defined by the sum of current time (moment of execution of the current event) and the duration of a given task (for example vehicle repair). Only times of starting a given route (event vehicle starts from the central node) are predefined (according to the time table). Duration of all other tasks are defined by system elements states: time when vehicle waits in the queue for loading/unloading, time when vehicle waits in the queue for maintains crew, or are given by random processes, time of vehicle going between two nodes, time of loading/unloading, time to failure, repair time.

Moreover each object representing a node has additional process (working in parallel) which is responsible for generating containers. The life cycle of this process is very simple: waiting a random time, generating a container with a given destination address (central node for all ordinary nodes, and each ordinary nodes for process in the central

node) and storing a container in the store house (implemented as a queue) of a given node.

Implementation

The event-simulation program could be written in a general purpose programming language (like C++), in a fast prototyping environment (like *Matlab*) or a special purpose discrete-event simulation kernel. One of such kernels, is the Scalable Simulation Framework (*SSF*) [16] which is a used for *SSFNet* [16, 17] computer network simulator. *SSF* is an object-oriented API - a collection of class interfaces with prototype implementations. It is available in C++ and Java. *SSFAPI* defines just five base classes: *Entity, inChannel, outChannel, Process*, and *Event*. The communication between entities and delivery of events is done by channels (channel mappings connects entities). For the purpose of simulating *DTS* we have used Parallel Real-time Immersive Modelling Environment (*PRIME*) [17] implementation of *SSF* due to a much better documentation then available for the original *SSF*. We have developed a generic class derived from *SSF* Entity which is a base of classes modelling.

DTS objects which models the behaviour of presented in section 2 and 3 discrete transport system. As it was mentioned a presence of randomness in the *DTS* model, the Monte-Carlo approach is used. The original *SSF* was not designed for this purpose so some changes in *SSF* core were done to allow to restart the simulation from time zero several times within one run of simulation programme. The statistical analysis of the system behaviour requires a very large number of simulation repetition, therefore the time performance of developed simulator is very important.

Case Study Analysis

Exemplar DTS

We propose for the case study analysis an exemplar *DTS* based on Polish Post regional centre in Wroclaw. We have modelled a system consisting of one central node (Wroclaw regional centre) and twenty two other nodes - cities where there are local post distribution points in Dolny Slask Province [18]. We have fixed the most important reliability and functional parameters of the key elements of the system. The length of roads was set according to real road distances between cities used in the analysed case study.

The intensity of generation of containers for all destinations was set to 4.16 per hour in each direction giving in average 4400 containers to be

transported each day. The vehicles speed was modelled by Gaussian distribution with 50 km/h of mean value and 5 km/h of standard deviation. The average loading time was equal to 5 minutes. There was single type of vehicle with capacity of 10 containers. The MTF of each vehicle was set to 20000. The average repair time was set to 5h (Gaussian distribution). We also have tried to model the drivers' availability parameters. We have fulfilled this challenge by using the following probability of a given type of sickness: short sick: 0.003, typical illness: 0.001, long-term illness: 0.00025. We hope that the proposed approach can properly model the real problems with driver availability at transportation enterprises.

Results and Discussion

We tried to realise the transportation tasks defined above using: 41, 44, 46 and 48 trucks. The number of drivers to operate the vehicle fleet we changed from 75 to 100. As we can see in Fig. 2. 75 drivers are absolutely inefficient for the analysed system – the acceptance ratio is almost zero and it does not depend on the number of used trucks. The number of vehicles equal to 44, 46, and 48 can be noticed as correct to make the *DTS* operative. The acceptance ratio grows up quickly and reaches the value of 0.9 – when we can say that system works "safety-far" from the border of inacceptable state. It is interesting that we need (at the level of acceptance ratio equal to 0.9) only 82 drivers to operate with 48 trucks, 85 drivers if we have 46 trucks and 90 drivers for 41 trucks.

If we require the acceptance ration at the level of 1 – 85 drivers are necessary for 48 trucks, 90 drivers for 46 vehicles and 95 drivers for fleet of 44 trucks. It is easy to notice that 41 trucks are not acceptable size of vehicle fleet to make the example *DTS* operative. There is no chance to substitute the shortage of trucks by the number of drivers. The acceptance ratio cannot reach the value of 0.6 even if we use 200 drivers (Fig. 3). The example shows how we can tune the size of *DTS* if we know the possible tasks definitions. The *DTS* owner's decisions ought to be taken in multidimensional environment. Our approach to *DTS* modelling and simulation can make the decision easier – we can observe immediately the results of possible solutions. The final decision ought to be the best – because it always generates financial consequences.

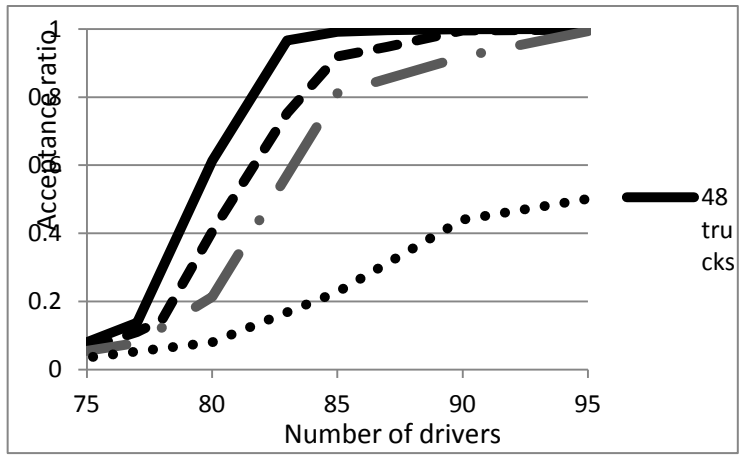


Figure 2: Acceptance ratio in a function of number of drivers for fixed number of vehicles

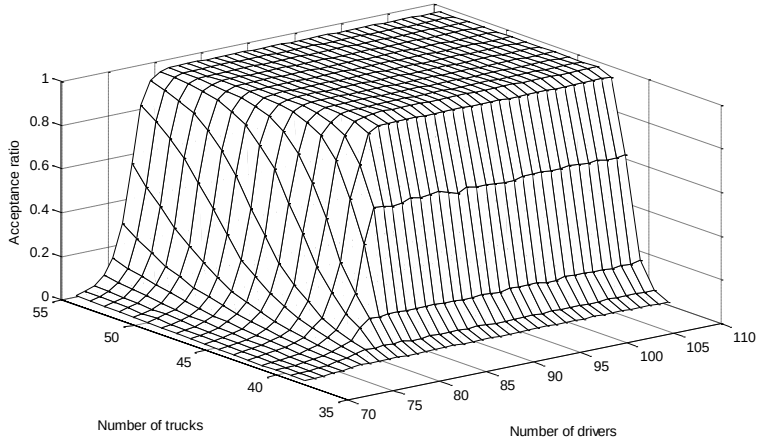


Figure 3: Acceptance ratio in a function of number of trucks and number of drivers for tested *DTS*

Conclusion

We have presented a formal model of discrete transport system (*DTS*). The *DTS* model is based on Polish Post regional transport system. The proposed approach allows performing dependability analysis of the *DTS*, for example:

- Determine what will cause a ”local” change in the system,
- Make experiments in case of increasing number of containers per day incoming to system,

- Identify weak point of the system by comparing few its configuration,
- Better understand how the system behaves,
- Foresee changes caused by human resource influence.

Based on the results of simulation it is possible to create different metrics to analyse the system in case of reliability, functional and economic case. The acceptance ratio of the system was introduced - defined in a functional way as an average of a ratio of on-time containers to all containers. The metric could be analysed as a function of different essential functional and reliability parameters of *DTS*. Also the system could be analyse in case of some critical situation (like for example a few day tie-up [16]). The chapter includes some exemplar systems, based on real Polish Post Wroclaw area, and calculated metric. The developed *DTS* simulator [17] makes it a practical tool for defining an organization of vehicle maintenance and transport system logistics.

Acknowledgement

Work reported in this paper was sponsored by a grant No. N N509 496238, (years: 2010-2013) from the Polish Ministry of Science and Higher Education.

References

[1] J. Barcelo, E. Codina, J. Casas, J.L. Ferrer and D. Garcia: Microscopic Traffic Simulation: a Tool for the Design, Analysis And Evaluation Of Intelligent Transport Systems, Journal of Intelligent and Robotic Systems: Theory and Applications, Vol. 41, pp. 173-203 (2005)

[2] R. Barlow and F. Proschan: Mathematical Theory of Reliability. Philadelphia: Society for Industrial and Applied Mathematics, (1996)

[3] M. Ben-Akiva, D. Cuneo, M. Hasan, M. Jha and Q. Yang: Evaluation of Freeway Control Using a Microscopic Simulation Laboratory, Transportation Research, Part C (Emerging Technologies), Vol. 11C, pp. 29-50 (2003)

[4] L. Birta and G. Arbez: Modelling and Simulation: Exploring Dynamic System Behaviour. London: Springer, (2007)

[5] C.N. Burt and L. Caccetta: Match Factor for Heterogeneous Truck and Loader Fleets, International Journal of Mining, Reclamation and Environment, Vol. 21, pp. 262-270 (2007)

[6] M.B. Duinkerken, R. Dekker, S.T.G.L. Kurstjens, J.A. Ottjes and N.P. Dellaert: Comparing Transportation Systems for Inter-Terminal Transport at the Maasvlakte Container Terminals, OR Spectrum, Vol. 28, pp. 469-493 (2006)

[7] G. Fishman: Monte Carlo: Concepts, Algorithms, and Applications. Springer-Verlag, (1996)

[8] N.Gartner,C.J.Messer and A.K. Rathi: Traffic Flow Theory and Characteristics. In: T.R. Board (Ed.). Texas: University of Texas at Austin, (1998)

[9] N. Gold, C. Knight, A. Mohan and M. Munro: Understanding service-oriented

software, IEEE Software, Vol. 21, pp. 71-77 (2004)
[10] P.A. Ioannou: Intelligent Freight Transportation. Carolina: Taylor and Francis Group, (2008)
[11] J. Krzyzanowska: The Impact of Mixed Fleet Hauling on Mining Operations at Venetia Mine, Journal of The South African Institute of Mining and Metallurgy, Vol. 107, pp. 215-224 (2007)
[12] H. Liu, L. Chu and W. Recker: Performance Evaluation of ITS Strategies Using Microscopic Simulation. In: Proceedings of the 7th International IEEE Conference on Intelligent Transportation Systems, pp. 255-270 (2004)
[13] B. Sanso and L. Milot: Performability of a Congested Urban-Transportation Network when Accident Information is Available, Transportation Science, Vol. 33, No 1, pp. 10-21 (1999)
[14] M.A.P. Taylor, J.E. Woolley and R. Zito: Integration of the Global Positioning System and Geographical Information Systems for Traffic Congestion Studies, Transportation Research, Part C (Emerging Technologies), Vol. 8C, pp. 257-285 (2000)
[15] I.F.A. Vis: Survey of Research in the Design and Control of Automated Guided Vehicle Systems, European Journal of Operational Research, Vol. 170, pp. 677-709 (2006)
[16] T. Walkowiak and J. Mazurkiewicz: Analysis of Critical Situations in Discrete Transport Systems. In : Proceedings of International Conference on Dependability of Computer Systems, Brunow, Poland, Los Alamitos: IEEE Computer Society Press, pp. 364-371 (2009)
[17] T. Walkowiak and J. Mazurkiewicz: Availability of Discrete Transport System Simulated by SSF Tool. In: Proceedings of International Conference on Dependability of Computer Systems, Szklarska Poreba, Poland,. Los Alamitos: IEEE Computer Society Press, pp. 430-437 (2008)
[18] T. Walkowiak and J. Mazurkiewicz: Functional Availability Analysis of Discrete Transport System Realized by SSF Simulator. In: Computational Science – ICCS 2008, 8th International Conference, Krakow, Poland,. Springer-Verlag, LNCS 5101, Part I, pp. 671-678 (2008)
[19] T. Walkowiak and J. Mazurkiewicz: Algorithmic Approach to Vehicle Dispatching in Discrete Transport Systems. In: Technical approach to dependability / Ed. by Jarosław Sugier, et al. Wroclaw: Oficyna Wydawnicza Politechniki Wroclawskiej, pp. 173-188 (2010)
[20] T. Walkowiak and J. Mazurkiewicz: Functional Availability Analysis of Discrete Transport System Simulated by SSF Tool, International Journal of Critical Computer-Based Systems, Vol. 1, No 1-3, pp. 255-266 (2010)
[21] T. Walkowiak and J. Mazurkiewicz: Soft Computing Approach to Discrete Transport System Management. In: Lecture Notes in Computer Science. Lecture Notes in Artificial Intelligence. Springer-Verlag, Vol. 6114, pp. 675-682 (2010)
[22] W. Zamojski, T. Walkowiak: An Approach to Service Net Modelling. The 4th International Conference on Information Technology, ICIT 2009, Amman, Jordan, Al-Zaytoonah University of Jordan, pp. 112-120 (2009)
[23] W. Zamojski, T. Walkowiak: Services Net Modeling for Dependability Analysis. Computational intelligence and modern heuristics, Eds. by Al-Dahoud Ali. Vukovar: In-Teh, pp. 1-15 (2010)

Chapter 7

ENERGY AWARENESS SPATIAL ROUTING IN WIRELESS SENSORS NETWORK

Jan Nikodem, Marek Woda, Maciej Nikodem and Ryszard Klempous

Summary

This Chapter presents how to realize in WSN networks energy awareness spatial routing employing relations. Considering the routing issue in WSN networks we looked for such a solution that will allow defining the strategy for a global network operation without giving detailed, explicit orders that can interfere with nodes local activities. As the result of research carried out, we proposed novel approach using the relationships. This approach allows for the routing area sizing and implementation costs allocation for a number of nodes in a network.

Introduction

The main task of WSN is to collect information from the area in which sensors are deployed. To perform this task it is necessary to retransmit information towards base station. Hence the two main interesting aspects of the network activity is efficient implementation of routing and efficient distribution of energy consumption resulting from the operation of the network. WSN is a distributed system not only territorially but also functionally and its individual components (nodes) can execute only tasks much simpler than these destined for the whole network. How efficiently manage such a large set of elements? How to ensure the appropriate cooperation, resulting in a global desired effect? Indeed, WSN is a system of large complexity and there is only one tool which can master intellectually this complexity. It is an abstraction. Therefore we are looking for new solutions in this domain. We obtained promising results by applying the new approach based on the set theory and relation defined on it. Such approach is really native one for WSN. Actually everything that happens in the network is based on collections (sets of nodes, sets of routing paths, sets of neighbors i.e. neighborhoods) and relations ("is a member of network", "is a cluster head", "belongs to routing path"). Traditionally we

met in the literature, methods based on functions; this is the result of the simplifications that have taken the authors of the different approaches. Simplifications which, make the issue simpler, often more transparent, allow us to grasp a certain part of the WSN network complexity but at the same time they narrow our considerations to some specific cases only. These limitations, inevitable on functional abstraction level may be omitted at a higher level of abstraction like sets and relations. This is because, proposed by us relations enable decomposition of a global task (the routing towards base station) for number of local tasks in the node's neighborhood. Another enticement of such approach stems from the fact that it leaves space for nodes interactivity.

Existing Approaches and Motivations for a New Method

In the literature on WSN networks [1]-[3], [5] authors devote most space for routing issues what is essential of the proposals in the literature of the routing seems to be:

a. Determination of optimal (according to a criterion) routing paths for messages.
b. The costs of routing activity in the network unequally burden nodes.
c. Cyclic procedure for routing paths determination, in order to provide routing adaptability (to change working conditions in each node of a network).

In brief, our study consisted of the search for such a method of routing, which would be devoid of the negative characteristics of action outlined in paragraphs A, B, C. In the literature, the most dealt with are the optimality related to; time-minimum time of packet delivery, space-minimum length path or costs-minimum cost (energy) of packet delivery. But no matter what optimality is taken into account, always when determining the routing paths global criterion is formulated and the calculation of optimal values requires information on the entire WSN network. This collected information is coming from a certain time interval (data acquisition needs time). Therefore, for the calculated solution (based on this data) is difficult to indicate the point in time when it was optimal. Moreover, evaluate exactly one solution; we do not take into account other, perhaps a bit worse. As a result we obtain a solution which is approximate rather than accurate; furthermore optimality of so obtained solutions is questionable. This becomes particularly evident in the case of algorithms that minimize energy consumption during the routing activity. Obtained routing paths are the most energy-efficient but

their determination absorbs some energy and the total energy balance is no longer so promising. Proposed method allows determining collections of efficient routing paths for sending information towards the BS.

With many alternative paths, we can make such choices, which guarantee the uniformity of the burden to pay for a routing of all network nodes. Leaving the determination of optimal routing paths we tend to implement routing activity as Pareto improvement process i.e. for a given set of routing paths we propose a way of evaluating them and as a result we obtain the Pareto frontier. This is a set (Pareto set) of routing paths that are Pareto efficient. To restrict our attention to the set of choices that are Pareto-efficient, we can make tradeoffs within this set, based on relations χ, π, ϑ. In future work we plan to extend the set of choices using less stringent Kaldor-Hicks efficiency criteria. Our proposition of ranking of alternatives from a set of efficient actions based on using a set of additive value surfaces spread over network area. These surfaces reflect the preference of information. The first surface is being determined only once in the beginning of network activity during self-organization process. The second surface has been determined at the same time but during the network lifespan it is modified many times locally.

Text Format Relational Attempt to WSN Simulation

WSN is a system of large complexity. There is only one tool which can master intellectually this complexity. It is an abstraction. Because of that, below we present how to describe WSN activity using a relational abstraction.

Relational Description of WSN Activity

Our approach utilizes three basic dyadic relations defined on the set of actions Act that describes communication activities in WSN. These three relations: collision χ, subordination π and tolerance ϑ were first introduced and described by Jaron [6] and later by Nikodem et al. [7], [8]. They are essential to describe variety of dependencies between real life objects (and also nodes of WSN). If we consider packet send and packet receive actions, then subordination relation means, that node x receives the packet whenever node y sends it.

$$x_R \, \pi \, y_S \qquad (1)$$

Subordination is transitive, so from x being subordinated to y, and z subordinated to x follows that z is also subordinated to y:

$$x_R \,\pi\, y_S \;\wedge\; z_R \,\pi\, x_S \;\Rightarrow\; z_R \,\pi\, y_S. \quad (2)$$

Subordination is not symmetric (it is asymmetric) which means that subordination of x to y implies that there is no subordination of y to x:

$$x_R \,\pi\, y_S \;\Rightarrow\; \neg(y_R \,\pi\, x_S). \quad (3)$$

Tolerance relation:

$$x_R \,\vartheta\, y_S, \quad (4)$$

Indicate that node x may receive packet sent to him from node y. Since node y decides to which node send the packet to, therefore, it is less likely that y sends data to nodes that are in tolerance relation with it – it is more likely that packets will be send to subordinated node. Tolerance, in contrast to subordination, is a symmetric relation, that is:

$$x_R \,\vartheta\, y_S \;\Rightarrow\; y_R \,\vartheta\, x_S. \quad (5)$$

It follows from the fact that x is in tolerance with y that y is in tolerance with x. Moreover, if x tolerates y and y is subordinated to z then also x is in tolerance relation to z:

$$x_R \,\vartheta\, y_S \;\wedge\; y_R \,\pi\, z_S \;\Rightarrow\; x_R \,\vartheta\, z_S. \quad (6)$$

It is a property of both relations that all nodes that are in subordination are also in tolerance relation. Collision is the last elementary relation that identifies nodes that will not exchange any packet with each other:

$$x_R \,\chi\, y_S. \quad (7)$$

Equation (7) means that node x will not receive any packet from y. Moreover, it is quite obvious that collision is symmetric so y will neither receive any message from x. Additionally if x is in collision to y and z is subordinated to x then z is also in collision with y:

$$x_R \,\chi\, y_S \;\wedge\; z_R \,\pi\, x_S \;\Rightarrow\; z_R \,\chi\, y_S. \quad (8)$$

Nodes that belong to collision relation cannot belong to tolerance at the same time. Therefore, nodes that may be in collision can be neither in tolerance nor in subordination relation. The later comes from the fact that nodes that are subordinated constitute subset of nodes that are in tolerance relation. Using relational framework to model packet transmission in WSN networks, we also used a neighborhood relation η. This relation is defined both for a single node of the network and for a group of nodes. It determines the set of nodes that are neighbours of a particular node or any node from the set, respectively. It is worth to point out that the neighbourhood relation is of the great significance since whole activity of every node of WSN is determined by the state of a node and its neighbours. Ensuring operation of the network nodes draw on relations π, ϑ, χ in order to determine global strategy of operation.

Neighbourhood is then used to perform local activities and to choose the best tactics that will be implemented in practice. We assume that neighboring relation is symmetric, i.e.

$$x\eta y \Rightarrow y\eta x \qquad (9)$$

Which means that if x is in relation with y (i.e. x can communicate with y) then y is also in neighboring relation with x? Using this relation we can define set of neighbours for a particular node x:

$$N(x)=\{y \mid y \in \text{Nodes} \wedge y\,\eta\, x\} \qquad (10)$$

and for a group of nodes S:

$$N(S)=\{y \mid y \in \text{Nodes} \wedge (\exists_{x\in S} \mid y\,\eta\, x)\} \qquad (11)$$

Using a neighborhood abstraction we can try to decompose globally defined activities to locally performed identical task ascribed to each node of the network. It will not be an easy task to cast all global dependencies from network area to the neighbourhood one. It will be even more difficult because neighbourhood conditions for the network nodes might be, and usually are, quite dissimilar.

Spatial Communication via Relational Attempt

Based on relational attempt, in the paper [9] we described the method of modeling spatial communication activity in wireless sensor network (WSN). Focusing on χ, π, ϑ and neighborhood relations instead looking for routing path towards base station (BS) (fig.1.a) we consider all possible retransmission nodes (fig.1.b) within neighbourhood.

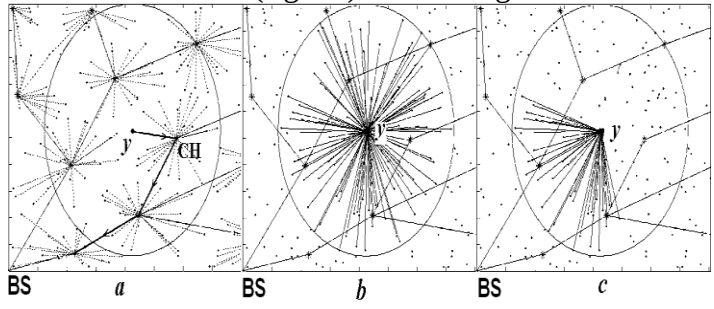


Figure 1: Clusters (a) vs. Neighbourhoods (b) and subordinated (towards BS) directions (c) in node y.

Our proposition for alternatives ranking from set of possible retransmission directions (fig.1.b) is based on using a set of additive value surfaces spread over network area (these surfaces will be discussed below). As a result of this process we obtain effective directions of

retransmission (fig.1.c). Paper [9] presents step by step how to use relations χ, π, ϑ in order to model spatial communication. Subordination π is responsible for multi-hop path generation. The increased intensity quotient of π results in extension of different multi-hop paths in communication space. π is responsible for a set of pontifixes (elements joining different paths). Tolerance ϑ is responsible for range of communication space. A bigger intensity quotient of ϑ widens communication space and extends possibility of parallel paths. However collision (χ) allows to form surface restrictions for the communication space.

Relational approach provides us with good tool for profile communication space. Using this tool, it is possible to design required properties of communication space. It is possible to profile communication space narrow or wide (ϑ), to obstruct selected area (χ) and prefer other point as especially recommended freeways for information flow (π). A local/global activity dilemma is a starting point of our consideration of modelling communication activity in WSN. We split all important aspects of communication activity into two classes. First class is composed of invariable aspects, while second class relates to aspects with local to global or local1 to local2 sensibility.

The network topology constitutes the first (invariable aspects) class. In contrast, node's energy states, cooperation and interference have been taking into account as the second (relative aspects) class. When modelling data flow from a network area towards base station we do this similarly to rainwater surface flow. Packets produced in WSN nodes flow like raindrops which streaming down in a direction determined by a slope of the modelled surface. During this process, drops merge with another (data aggregation), carve terrain or build it like lava tears (energy awareness). A resulted flow has been finally conditioned by the local neighbourhood conditions and environmental stimulus (cooperation and interference).

We model natural network topology features using digital surface model (DSM). It is a component of a topographic map (bare drainage surface), which gives a basic reference frame that ensures packets are send towards the BS. In a real WSN network nodes usually have no information about their Euclidean distance from the BS (fig.2). Therefore in the paper [9] we propose a measure of dis(k) (distance between BS and node k) based on the amount of hops (h) required to send packet from node k to the BS. We determine bare drainage surface (BDS) only once during network self-organization phase, so it is invariant in WSN

lifespan. Superposition of BDS and relational surface constitutes effective communication space for each node (fig.3).

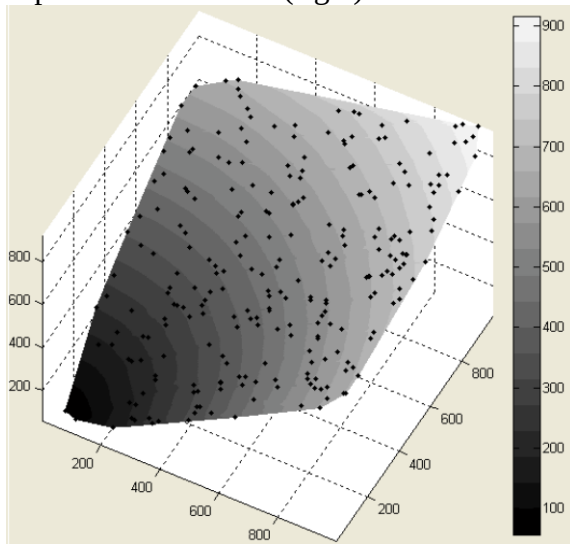


Figure 2: Drainage function based on node to BS Euclidean distance.

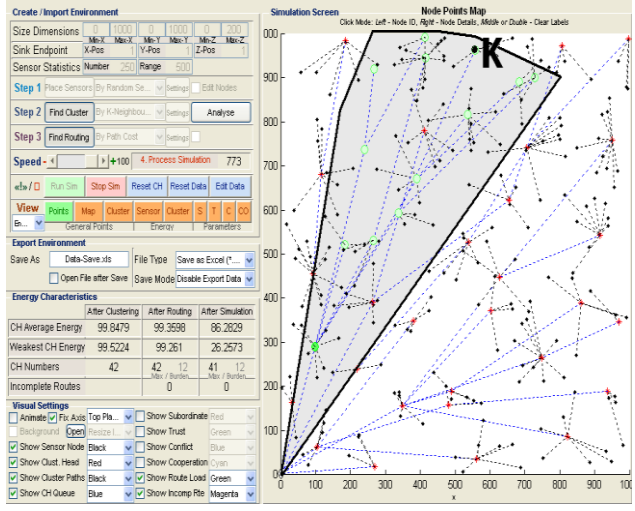


Figure 3: Effective communication space for node K.

Principles of Energy Awareness Routing

Based on relational attempt, for each node in the network we can determine a set of efficient routing directions for sending a packet towards the BS (fig.1.c). From many alternative nodes for packet

retransmission, we can choose one, which guarantee (within neighbourhood) the uniformity of the energetic burden to pay for sending packet. For this purpose (for ranking of alternatives from the set of efficient actions) we use, an energy remain in node, value surface stretched over neighbourhood of each node.

In our routing implementation each node sends packet to this of his neighbours which belongs to the set of efficient routing. However, a node will choose such of them which will have the largest energy resources (i.e. hills on fig.4). This simple rule ensures even distribution of energy consumption for each network node (right side on fig.5).

Such protection of the lowest energy node in the neighborhood is adaptive. Nodes that have least energy are protected more efficiently. Especially pontifixes, nodes strategically situated on the crossings of many routing paths. These nodes, in traditional algorithms, most often die thereby necessitating the calculation of a new routing tree. Moreover we obtain a very effective tool to governance the cooperation between nodes during routing activity.

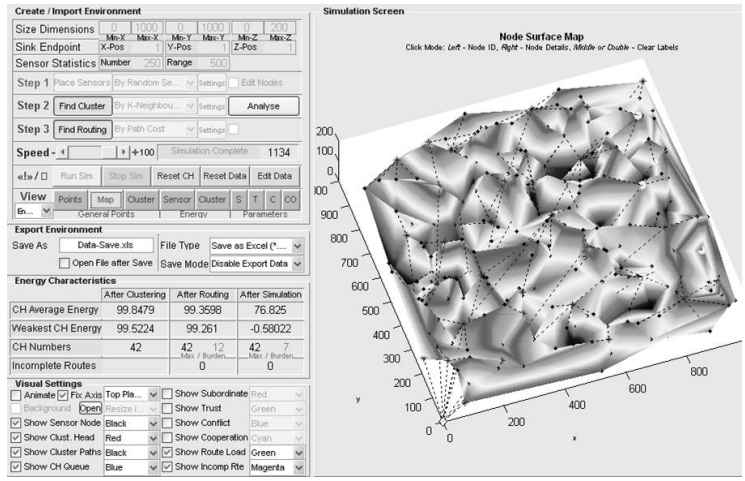
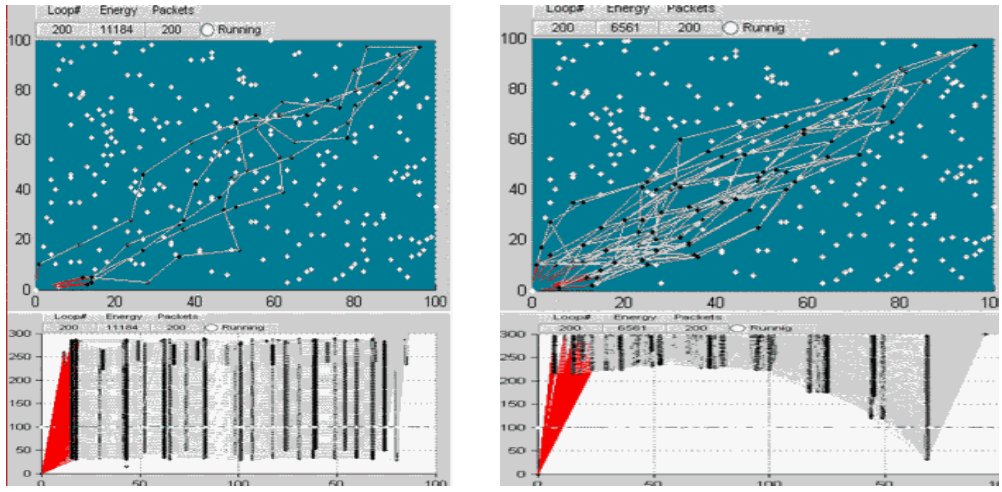


Figure 4: Energy remained surface - modelled in simulator.



(a) HEED Algorithm (b) . Energy Awareness Approach

Figure 5: Distribution of Energy Consumption

Realization of Energy Awareness Spatial Routing

This paragraph describes how we implemented a novel, relation based routing in WSN used Crossbow Technology Inc. motes programmed with Berkeley's TinyOS operating system.

Hardware and Software Environment

The used hardware was a selection of MICAz and Iris motes from Crossbow Technology Inc. (fig.6). The motes model used was MPR400CB along the CC1000 900MHz data radio which enables the mote to measure the battery voltage (used to power a mote). The base station interface unit, model MIB520CB, was USB based and it serves two main purposes. Firstly, it allows the user to reprogram any mote by plugging the mote directly into the base. Secondly, operates as part of the root node interface giving the PC a data conduit onto the radio based sensor network (all motes worked on the same frequency). PC was connected to BS and worked as operator's console with network address - ID=0x0000.

The mentioned above hardware worked under TinyOS operating system specifically designed for network embedded systems. TinyOS has component based architecture, simply event based concurrency model and split-phase operations.

Figure 6: Crossbow Technology Inc. MICAz mote and MIB520CB base station

All programs were written in NesC language which is extremely sensitive to hardware and software configurations.

Packet's Header Structure

Below, on fig.7 there is an outline of packet's header structure. We present only these fields in packet's header which are important from the self-organization activity point of view.

Dest Node ID	Message type	SrcNode ID	Sequence No.	Command Type	...	

Figure 7: The structure of packet's header

- DestNode ID, SrcNode ID are address 0x0000- 0xffff
  - 0x0000 means operator's console (PC),
  - 0xffff means broadcasting mode.
- Message type (byte) serves two purposes.
  - 0x0–message,
  - 0x1–command.
- Sequence No is useful during broadcast communication to avoid multiply retransmission of the same packet.
- Command type serves multipurpose. In self-organizing phase we use three of them
  - 0x3-prepare for routing command,
  - 0x4-start routing command,
  - 0x14-acknowledgement of routing command.

Self-Organization Phase

Immediately after the power is turned on each node reads its address (ID) and the level of its battery voltage. Then begins sniffing what is happening in its neighbourhood. Self-organization phase started when base station (BS) receives from PC (ID=0x0000) command 0x03 with DestNodeID=0xFFFF. This is a signal to prepare for the routing. Regarding to DestNodeID= 0xFFFF this command should be broadcasted

in the network, so base station re-transmits it further. Similarly make the other nodes, re-transmitting this command upwards. Each node retransmits this command only once because using the field SequenceNo node do not retransmit command with the same or less value in this field. To avoid sudden spikes in demand on the transmission channel, each node re-transmits command after a randomly determined time interval. The value of this delay is determined based on the difference between node IDs (current node ID and previous node ID). In this way, based on flooding technique the entire network learns to prepare for self-organization process.

DescNode ID	HopDist	Battery voltage	Transmission power	...

Figure 8: The structure of packet's data field

During the process of flooding the network nodes sniffing the network activity out and create a matrix of their neighbours updating the number of neighbours and the address (ID) of each of them. The final step of network self-organization is initialized by PC (ID=0x0000). From operator's console we send command 0x04 with DestNodeID=0xFFFF. This is a signal to construct spatial routing. Regarding to DestNodeID= 0xFFFF this command should be broadcasted in the network, so base station re-transmits it further. Similarly make the other nodes, re-transmitting this command upwards. Each node retransmits this command only once because using the field SequenceNo node do not retransmit command with the same or less value in this field. To avoid sudden spikes in demand on the transmission channel, each node re-transmits command after a randomly determined time interval. The value of this delay is determined based on the difference between node IDs (current node ID and previous node ID). In this way, based on flooding technique the entire network learns to prepare for self-organization process.

Deja vu! Yes, moreover this is not very sophisticated algorithm. For command 0x04 each node executes nearly the same procedure as for command 0x03. The slight difference we explain below on fig.8 where is an outline of packet's data field structure.

- DescNode ID is address 0x0000- 0xffff
    - If node received packet, DescNode ID points out the address of descender node downwards base station,
    - While node sends packet, DescNode ID points out its address to show potential ascenders the way towards base station.

- HopDist (byte) serves only one purpose. Helps receiver to determine bare drainage surface which slope allows packets streaming down (like raindrops) downwards base station. The base station puts into this field value 0x00 before broadcasting command 0x04 (start of routing process).
- Battery voltage (16-bits value) is useful for ranking of alternatives from the set of efficient directions of re-transmission. Receiver uses this value to calculate energy remained in preceding node (descender). Finally, receiver chooses that node which has maximal reserve of energy. It guarantees (within neighbourhood) the uniformity of the energetic burden to pay for re-transmitting a packet.
- Transmission power (16-bits value) is useful for determining the relative distance between receiver and preceding node (descender). Comparing this value with the strength of receiving signal, receiver can calculate relative distance to preceding node.

During the self-organization phase each node while received the command 0x04 in broadcast mode, determines:

- its distance (HopDist+1) from the base station, expressed in numbers of hops,
- stores the following data of its predecessor
    - DescNode ID,
    - Calculated reserve of its energy,
    - Calculated relative distance to it.

Next, node (which received the command 0x04) puts its ID to packet DescNode ID field, increases by 1 the value HopDist, and puts its value of Battery voltage and its value of Transmission power. Then, node sends that packet backwards (towards base station by its predecessor) as confirmation of ending a routing procedure. The confirmation packet has in the field "Command type "of header value 0x14 (command value + 0x10). Finally, node re-transmits command 0x04 in broadcast mode with modified data fields.

Determining the Node's Neighborhood

During the self-organization process each node receives in packet data field information about its neighbours. Based on this data nodes creates its neighbourhood matrix. The structure of this matrix is presented below in table I. Policy of this data acquisition and of completing the fields of this table is given in the preceding chapter.

Table 1: Structure of k-node's neighborhood matrix

No. of neighbours = n	**Neighbour ID(8)**	**Neighbour ID(12)**	**Neighbour ID(5)**	**Neighbour ID(7)n**
HopDist to BS	3	1	2	1
Battery voltage	2.945	2.876	2.432	2.109
Relative dist.	21.35	17.45	16.85	18.55

Row "Relative dist." (distance) allows us to optimize energetic burden to pay for re-transmitting packets towards base station. Row "Battery voltage" is fruitful while we realize an even distribution of energy consumption for each network node. It allows us to distribute energy costs of re-transmission within all neighbourhoods. Row "Hop distance to BS" allows us to create (basing on relations χ, π, ϑ) a bare drainage surface what we discuss in the next chapter.

Building a Bare Drainage Surface within Neighborhood

The node's neighbourhood matrix established during the self-organization process, contains all information necessary to implement the spatial and energy distributed routing in the network. There are many various feasible strategies for determining an effective communication space. Here, to outline the principles of proposed method, we will present only one and relatively simple strategy. For this purpose, we use the relations χ, π, ϑ.

Considering as an example table I, the hop distance from considered node to base station is equal 3. There are other nodes within its neighbourhood. Some of them are located closer (hop distance=2), another further (hop distance=4), base station.

Table 2: k-Node's Bare Drainage Surface Matrix

No. of neighbours =n	**Neighbour ID(7)**	**Neighbour ID(12)**	**Neighbour ID(5)**	**Neighbour ID(8)n**
HopDist to BS	**1**	**1**	**2**	**3**
Battery voltage	2.109	2.876	2.432	2.945
Relative dist.	18.55	17.45	16.85	21.35
Relations	π	π	ϑ	χ

There are also nodes within neighbourhood with the same hop distance. During creation of bare drainage surface (fig.9) within node's neighbourhood we use a simple trivalent classifier (closer BS, at the same distance, further BS).

$$N<(k) = \{y \in N(k) \mid dis\,(y) < dis(k)\};$$

$$N=(k) = \{y \in N\,(k);\ y \neq k \mid dis(y) = dis\,(k)\}; \qquad (12)$$

$$N>(k) = \{y \in N\,(k) \mid dis(y) > dis\,(k)\};$$

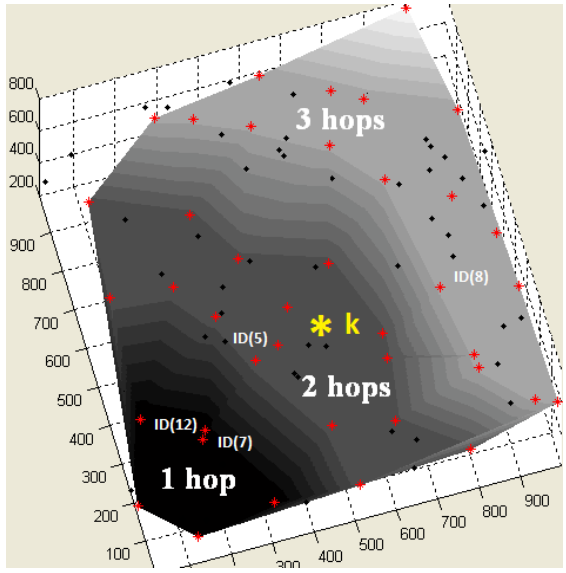


Figure 9: The bare drainage surface with k-node's neighbourhood 3D view

For these three classes we assign appropriately relations; subordination, tolerance and collision. Primary set designating an effective communication space within these neighbourhood consists of two elements {ID (7), ID (12)} with possible extension to {ID (7), ID (12)} + {ID5)} if we concern tolerance relation. Ordering node's neighbourhood matrix (table II) according to hop distance to base station we obtain very simple, but effective operating, bare drainage surface within neighbourhood (fig.10).

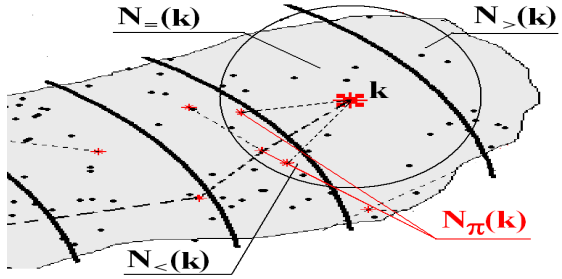


Figure 10: The bare drainage surface with k-node's neighbourhood 2D view

Conclusions

Written in NesC language software for MicaZ motes allows testing WSN activity related to packet re-transmission. Implemented algorithm was based on relations and set theory. Relational framework allows determining a strategy for a global network operation without giving detailed explicit orders that can interfere with nodes local activities. We can determine, using relations, the recommended global routing areas (spatial routing), giving re-transmitters possibility to explicitly choose a next path. The decision, which re-transmitter to select next, is completely local, and it takes into consideration the instant neighbourhood conditions. However, it is coordinated globally and thus fits into the operation strategies of entire network.

Writing the program we involve relations to create a drainage surface that determine direction of all routing paths towards base station. The drainage surface represents a natural network topology features helpful for routing and communication activity. The pivotal role and main goal of drainage function are - simplification of the next hop selection process during on the routing path determination as well as to guarantee that chosen direction of data-flow is always correct (i.e. BS-oriented).

We use the relational approach to attain spatial routing. There were three relations defined: subordination, tolerance, and collision and additionally we define their properties that guarantee proper order within neighbourhood of each WSN node. Relational approach allows enforcing the required global strategy scenario that is prepared for the whole network by sending triple of intensity quotients (subordination, tolerance, and collision). By a modification of these intensity quotients value, the drainage surface is being upheaval (communication is being blocked in that region) or lowered (communication activity is being intensified) (fig.11).

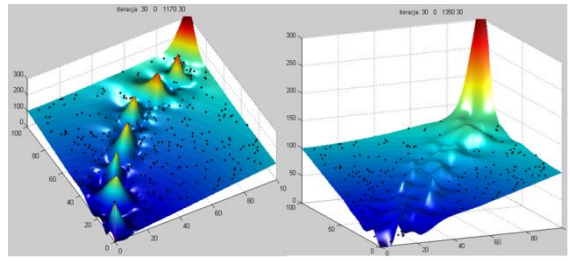


Figure 11: The energy consumption during; traditional (left) and energy awareness spatial routing (right)

Acknowledgement

This work was partially supported by E.U. Regional Development Fund and by Polish Government within the framework of the Operational Programme - Innovative Economy 2007-2013. The contract for refinancing No. POIG.01.03.01-02-002/08-00, detectors and sensors for measuring factors hazardous to environment - modelling and monitoring of threats.

References

[1] D. Braginsky and D. Estrin, "Rumor routing algorithm for sensor networks", in Proc. of the 1-st Workshop on Sensor Networks and Applications, Atlanta, GA, 2002.

[2] O. Fang, J. Gao, V.L. de Silva Guibas and L. Zhang., "GLIDER: gradient landmark-based distributed routing for sensor networks", in Proc. INFOCOM'05, 2005, vol.1, pp. 339-350.

[3] A. Manjeshwar and D.P. Agrawal, "TEEN: a routing protocol for enhanced efficiency in wireless sensor networks", in Proc. IPDPS'01, 2001.

[4] J. Nikodem, "Designing communication space in wireless sensors network based on relational attempt", ser. Lecture Notes in Computer Science, Berlin, Germany: Springer, 2009, vol.5717

[5] O. Younis and S. Fahmy, "HEED: A hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks", IEEE Transactions on Mobile Computing, vol.3, no.4, pp.366-379, October 2004.

[6] J Jaroń, "Systemic prolegomena to theoretical cybernetics", ser. Scient. Papers of Inst. of Techn. Cybernetics, Wroclaw, Poland: Wroclaw Univ. of Techn., 1978, vol.45.

[7] J. Nikodem, "Autonomy and cooperation as factors of dependability in wireless sensor network", in Proc. DepCoS – RELCOMEX'08, pp.406-413.

[8] J. Nikodem, R. Klempous and Z. Chaczko, "Modelling of immune functions in a wireless sensors network", in Proc. EMSS'08.

[9] J. Nikodem, R. Klempous, M. Nikodem, M. Woda and Z. Chaczko,"Modeling a spatial communication activity in wireless sensors network", in Proc. MAS 2009, pp.50-55.

APPENDIX

Chapter 1

A-1: A Program Sample for Risa/Asir(Circle ratio)

```
def
cpi(ID, Loop)
{
  N = length(ID);
  Loop1 = idiv(Loop, N);
  Start = 0;
  Width = 1.0 / Loop;
  for (I = ID; length(I) > 1;
      I = cdr(I)) {
    ox_rpc(car(I), "cpi_sub",
          Start, Start+Loop1,
          Width);
    Start += Loop1;
  }
  ox_rpc(car(I), "cpi_sub", Start,
        Loop, Width);

  Pi = 0;
  for (I = ID; I != []; I=cdr(I))
  {
    Pi += ox_pop_local(car(I));
  }
  return Pi / Loop;
}

def
cpi_sub(Start, End, Width)
{
  Pi = 0;
  for (I = Start; I < End; I++) {
    X = (I + 0.5) * Width;
    Pi += 4.0 / (1.0 + X ^ 2);
  }
  return Pi;
}

def
time_cpi(ID, Loop)
{
  T0 = time();
  print(cpi(ID, Loop));
  T1 = time();
  return T1[3]-T0[3];
}
def
rpc_cpi2_init(L)
{
  rpc_load(L, "cpi2.asir");
}

def
rpc_load(L, FN)
{
  for (I = L; I !=[]; I = cdr(I))
    ox_rpc(car(I), "load", FN);
}
end$
```

A-2: A C Program Sample for CELL B.E. with SPE Library (Circle ratio)

```
# PPU
P-GCC= gcc
P-CFLAGS= -lpthread -lspe2 -Wall -
Werror
P-SRC= pi-ppe.c
P-OUT= pi-ppe

# SPU
S-GCC= spu-gcc
S-CFLAGS= -Wall -Werror
S-SRC= pi-spe.c
S-OUT= $(SPU-SRC:.c=.elf)

all:
   $(P-GCC) $(P-SRC) \
          -o $(P-OUT) $(P-CFLAGS)
   $(S-GCC) $(S-SRC) \
          -o $(S-OUT) $(S-CFLAGS)
```

–***Makefile for a CELL-based C program*** –

```
typedef struct {
 unsigned long long  ea_in;
 unsigned long long  ea_out;
 unsigned int        size;
 int                 pad[3];
} pi_params_t; //8+8+4+4*3=32bytes

stypedef struct {
 int  start, end;
 int  pad[2];
 double  width, s;
} pi_data_t; //4*4+8*2=32bytes
```

–***pi.h*** *: a header for a CELL-based C programe* –

```
#include <spu_intrinsics.h>
#include <spu_mfcio.h>
#include "pi.h"

int
main(
  unsigned long long spe,
  unsigned long long argp)
{
  static
  pi_params_t pi_params
    __attribute__((aligned(16)));
  static
  pi_data_t pi_data
    __attribute__((aligned(16)));
  int i, tag = 1;
  double x;

  spu_mfcdma64(&pi_params,
            mfc_ea2h(argp),
            mfc_ea2l(argp),
            sizeof(pi_params_t),
            tag, MFC_GET_CMD);
  spu_writech(MFC_WrTagMask,
           1 << tag);

spu_mfcstat(MFC_TAG_UPDATE_AL
L);

  spu_mfcdma64(&pi_data,
     mfc_ea2h(pi_params.ea_in),
     mfc_ea2l(pi_params.ea_in),
     pi_params.size,
     tag, MFC_GET_CMD);
  spu_writech(MFC_WrTagMask,
           1 << tag);

spu_mfcstat(MFC_TAG_UPDATE_AL
L);

  pi_data.s = 0.0;
  for(i = pi_data.start;
      i < pi_data.end; i++) {
    x = (i + 0.5) * pi_data.width;
    pi_data.s
      += 4.0 / (1.0 + x * x);
  }

  spu_mfcdma64(&pi_data,
       mfc_ea2h(pi_params.ea_out),
       mfc_ea2l(pi_params.ea_out),
       pi_params.size,
       tag, MFC_PUT_CMD);
  spu_writech(MFC_WrTagMask,
           1 << tag);

spu_mfcstat(MFC_TAG_UPDATE_AL
L);

  return 0;
}
```

–***pi-spe.c****: a CELL-based C program for worker* –

```
#include <stdio.h>
#include <stdlib.h>
#include <libspe2.h>
#include <pthread.h>
#include "pi.h"
#define NUM_SPE 6

typedef struct {
  spe_context_ptr_t spe_ctx;
  pi_params_t       *pi_params;
} thread_arg_t;
```

```
void *run_spe_thread(void
*thread_arg)
{
  thread_arg_t *arg =
     (thread_arg_t *) thread_arg;
  unsigned int entry;
  spe_stop_info_t stop_info;
  entry = SPE_DEFAULT_ENTRY;
  spe_context_run(arg->spe_ctx,
                &entry, 0,
                arg->pi_params,
                NULL,
                &stop_info);
  return NULL;
}

int main(int argc, char *argv[])
{
  static
  pi_params_t pi_params[NUM_SPE]
  __attribute__((aligned(16)));
  static
  pi_data_t pi_data[NUM_SPE]
  __attribute__((aligned(16)));
  int i, ret, start,
     loop = 50000000, loop1;
  double width, pi;
  spe_program_handle_t *spe_prog;
  spe_context_ptr_t
    spe_ctx[NUM_SPE];
  pthread_t thread[NUM_SPE];
  thread_arg_t arg[NUM_SPE];

  spe_prog
   = spe_image_open("pi-spe.elf");
  for (i = 0; i < NUM_SPE; i++) {
    spe_ctx[i]
      = spe_context_create(0,
                    NULL);
    ret
     = spe_program_load(spe_ctx[i],
                  spe_prog);
  }
  loop1 = loop / NUM_SPE + 1;
  start = 0;
  width = 1.0 / loop;
  for (i = 0; i < NUM_SPE; i++) {
    pi_data[i].start = start;
    if (start + loop1 < loop)
      pi_data[i].end
        = start + loop1;
    else
      pi_data[i].end = loop;
    pi_data[i].width = width;
    pi_params[i].ea_in
      = (unsigned long)
         &pi_data[i];
    pi_params[i].ea_out
      = (unsigned long)
         &pi_data[i];
    pi_params[i].size
      = sizeof(pi_data_t);
    arg[i].spe_ctx = spe_ctx[i];
    arg[i].pi_params
      = &pi_params[i];
    ret
      = pthread_create(&thread[i],
         NULL, run_spe_thread,
         &arg[i]);
    start += loop1;
  }
  for (i = 0; i < NUM_SPE; i++) {
    pthread_join(thread[i], NULL);
    ret
      = spe_context_destroy
         (spe_ctx[i]);
  }
  ret = spe_image_close(spe_prog);
  pi = 0.0;
  for(i = 0; i < NUM_SPE; i++)
    pi += pi_data[i].s;
  pi /= loop;
  printf("[PPE] pi = %g\n", pi);

  return 0;
}
```

–***pi-ppe.c****: a CELL-based C program for master –*

A-3: A Program Sample for Risa/Asir(exp(x))

```
def
apx_exp(X, N) {
  EX = 0;
  NM = 1;  DN = 1;
  for(I = 0; I <= N;
```

```
    I++, NM *= X, DN *= I)
   EX += NM/DN;
  return EX;
}
def
apx_exp_para(ID, X, N)
{
  Length1
    = ceil((N+1) / length(ID));
  Start = 0;
  for (I = ID; length(I) > 1;
       I = cdr(I),
       Start += Length1)
    ox_rpc(car(I),
           "apx_exp_para_sub",
           X, Start,
           Start + Length1);
  ox_rpc(car(I),
         "apx_exp_para_sub",
         X, Start, N + 1);
  EX = 0;
  C = 1;
  for (I = ID; I != [];
       I = cdr(I)) {
    A = ox_pop_local(car(I));
    EX += C * car(A);
    C *= car(cdr(A));
  }
  return EX;
}
def
apx_exp_para_sub(X, Start, End)
{
  EX = 0;
  if (Start > 0) {
    NM = X;  DN = Start;
  } else {  /* Start == 0 */
    NM = 1;  DN = 1;
  }
  for(I = Start; I < End;
      I++, NM *= X, DN *= I) {
    Term = NM/DN;
    EX += Term;
  }
  return [EX, Term];
}
def
rpc_apx_exp_init(L)
{
  rpc_load(L, "apx_exp.asir");
}
def
rpc_load(L, FN)
{
  for (I = L; I !=[]; I = cdr(I))
    ox_rpc(car(I), "load", FN);
}
end$
```

ADVANCES IN INFORMATION TECHNOLOGY

From Artificial Intelligence To Virtual Reality

This book contains the extended version of the best published papers in *The fifth International Conference on Information Technology 2011*. It includes 7 chapters' covers the following areas include, but are not limited to:

- **Distributed systems**
- **Parallel computing**
- **Embedded artificial intelligence**
- **Web services networks**
- **E-learning,**
- **Discrete transport systems**
- **Routing In Wireless Sensors Network.**

UBICC Publisher

ISBN-13: 978-1468128994

ISBN-10: 146812899X