Journal of Information Technology and Applications

(BANJA LUKA)

Exchange of Information and Knowledge in Research

VOLUME 2 NUMBER 1 BANJA LUKA, JUNE 2012 (1-52) ISSN 2232-9625 (Print) UDC 004

APEIRON

THE AIM AND SCOPE

The aim and scope of the Journal of Information Technology and Applications (JITA) is:

- to provide international dissemination of contributions in field of Information Technology,
- to promote exchange of information and knowledge in research work and
- to explore the new developments and inventions related to the use of Information Technology towards the structuring of an Information Society.

JITA provides a medium for exchanging research results and achievements accomplished by the scientific community from academia and industry.

By the decision of the Ministry of Education and Culture of the Republic of Srpska, no.: 07.030-053-160-4/10 from 3/3/2010, the journal „Journal of Information Technology and Applications“ Banja Luka is registered in the Registry of public organs under the number 591.
Printed by Markos, Banja Luka in 300 copies two times a year.

Indexed in: LICENSE AGREEMENT, 3.22.12. **EBSCO** Publishing Inc., Current Abstracts

Full-text available free of charge at http://www.jita-au.com

UDC 004

ISSN 2232-9625 (Print)
ISSN 2233-0194 (Online)

JITA

VOLUME 2 NUMBER 1 BANJA LUKA, JUNE 2012 (1-52)

Contents

Editorial

EDITORS:

GORDANA RADIĆ, PhD
Editor-in-Chief

ZORAN AVRAMOVIĆ, PhD

DUŠAN STARČEVIĆ, PhD

Dear Readers,
Welcome to the third issue of the JITA Journal!

Within the framework of publishing at the PanEuropean University APEIRON in Banjaluka, JITA – Journal of Information Technology and Applications has been published since June 2011.

We feel proud and privileged that we had the opportunity and were entrusted with preparation and organization of this journal, in which, modern scientific achievements and prospect scientific activities in Information Technology are being presented.

We are very pleased that the previous issues of this journal have aroused significant interest in professional and scientific public, both at home and abroad. This interest is, undoubtedly, the result of quality of papers presented by members of prestigious scientific associations, university professors and renowned scientists and researchers.

Besides renowned authors, whose names regularly appear in scientific journals, JITA also invites young and yet unestablished authors to submit papers in which they will present the results of their scientific research. Since JITA is based on strict selection of papers in the process of review, high-quality papers will be accepted for publication.

Papers in this issue are in the area of Intrusion Detection Systems in the Smart Grid networks, Extreme Programming, Computer Based Testing, Relational Data Bases, Data Mining and e-commerce.

We consider it our duty to thank the authors who made a tremendous effort to prepare papers in a high-quality manner, members of the Review Board on quality work and time spent which they dedicated to journal preparation and organization, unselfish engagement – in addition to their numerous daily duties.

Editors recommend intensive communication between authors and readers. In that way, we believe, the utility value of each presented paper will increase.

To the authors, we wish for their published papers to be well-received by scientific and professional public, and to the readers, to draw concrete and valuable scientific truths from those papers.

We will endeavor to ensure the reputation and quality of journal with each issue.

On Intrusion Detection in a Neighbourhood Area Network in the Smart Grid

Nasim Beigi-Mohammadi[1], Hamzeh Khazaei[2], Jelena Mišić[1], Vojislav B. Mišić[1]

[1]Ryerson University, Toronto, ON, Canada

[2]University of Manitoba, Winnipeg, MB, Canada

Contribution to the State of the Art

DOI: **10.7251/JIT1201007M** UDC **504.06+502.17**

Abstract: Smart grid, which is an upgrade of power electric system, mainly relies on powerful communication networks to provide a secure, reliable and efficient information delivery. Updating a system as complex as the electrical power grid with a large number of components has the potential of introducing new security vulnerabilities into the system. Hence, security mechanisms should be deployed to protect the smart grid as a first wall of defence against malicious attacks. As a second wall of defence, there should be intrusion detection systems in place to protect the smart grid against any security breaches. In this work, we describe an anomaly-based intrusion detection system (IDS) for neighbourhood area network whose security is of critical importance in smart grid.

Keywords: Intrusion Detection System, smart grid, Neighbourhood Area Network

Introduction

Advanced Metering Infrastructure (AMI) is a communication infrastructure that enables meters and utilities to exchange information such as power consumption, firmware updates, remote disconnects or outage awareness [1]. An AMI includes several communication networks that can be generally classified into: Home Area Network (HAN), Neighbourhood Area Network (NAN) and Wide Area Network (WAN) [18]. HAN is the network of sensors that communicate with smart meters in residential or industrial area while NAN is a network of neighbouring smart meters that communicate with collecting nodes, namely, collectors. WAN serves as a communication link between utility center and data collectors. An overview model of the AMI is shown in Figure 1. AMI introduces new security challenges since it consists of billions of low-cost commodity devices being placed in physically insecure locations. The equipment is under the control of the often disinterested, unsophisticated, or sometimes malicious users. The author in [3] discusses the security requirements and related threats of the four main components of an AMI: smart meters, the customer gateway, the communication network, and the head end. The fact that encryption and authentication alone are not sufficient to protect the infrastructure is emphasized. In AMI, availability and integrity of data take precedence over confidentiality [11][17]. Attacks targeting AMI can be classified into three categories including network compromise, system compromise and denial of service [2].

Traffic modification, false data injection and replay attacks try to compromise the network [10] while compromised node and spoofing of metering devices are examples of attacks which target the systems. Flaws or misuses of routing, configuration, and name resolution are considered as denial of service attacks. While threats discussed in [3] are required to be highly taken into account when designing security mechanisms, AMI lacks a reliable monitoring solution. One approach for designing an IDS for AMI is to leverage the existing IDS techniques that have been used in other types of networks. However, there

are AMI-specific challenges that need to be aware of when designing an IDS for AMI. The IDS should be highly accurate since at the ultimate end it deals with availability which is considered to be the most critical aspect of smart grid [14]. Moreover, it should have a low communication and computation overhead on the network due to resource constraint devices in AMI. Traditional IDS mechanism including a number of lightweight agents reporting to a central management server is not applicable in such system. For instance, AMI networks may contain millions of nodes that with a central approach for monitoring and intrusion detection, the traffic load, required storage and computational capabilities at the central server could be overwhelming. Therefore, a distributed approach should be considered. In a distributed IDS, data processing is distributed among intermediate nodes and only high level data is sent to the central server [5]. In this work, an anomaly-based IDS for NAN is proposed which utilizes several rules to detect anomalies in the network.

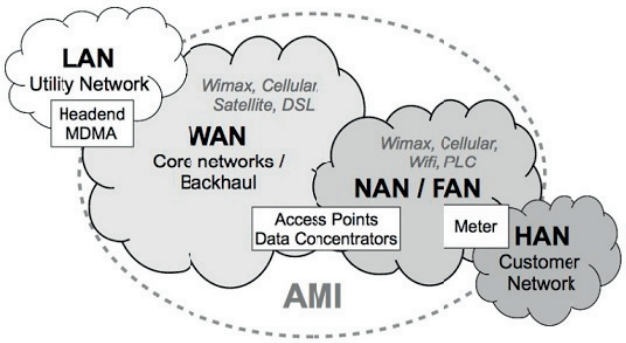


Figure 1. Overview of AMI networks (after [2]).

The rest of this paper is organized as follows: in Section 2, we briefly survey some related work. A realistic intrusion scenario is described in Section 3. Section 4 elaborates the IDS solution, while Section 6 presents the rules for the IDS. Finally, Section 5 concludes the paper and outlines some promising directions for future work.

Related Work

While many efforts have been made to investigate the security of AMI, there are a few works that focus on proposing and designing reliable and efficient IDS for AMI. Berthier, Sanders, and Khurana [2] discuss the requirements and practical needs for monitoring and intrusion detection in AMI. Kush *et al*. [7] have surveyed the gap analysis of intrusion detection in smart grid. They identify and present the key functional requirements of the IDS for smart grid environment. Jokar et al. [5] present a layered specification-based IDS for HAN. Their IDS is designed for ZigBee technology which is deployed in HAN communication. They specifically address the physical and medium access control (MAC) layers. Their work, however, is a partial solution since it only takes the two lower layers of ZigBee technology into account (i.e., considering only 802.15.4) as their feature space.

In [1], a specification-based IDS for AMI has been proposed. While the solution in [1] relies on protocol specifications, security requirements and security policies to detect security violations, it would be expensive to deploy such IDS since it uses a separate sensor network to monitor the AMI.

Roosta *et al*. [13] propose a model-based IDS working on top of the WirelessHART protocol, which is an open wireless communication standard designed to address the industrial plant application, to monitor and protect wireless process control systems. The hybrid architecture consists of a central component that collects information periodically from distributed field sensors. Their IDS monitors physical, data link and network layer in order to detect malicious behaviour. While authors provide a detailed explanation of their work, their IDS solution cannot be completely applied to NAN IDS because it is protocol-specific.

Authors in [11] investigate a technique for evaluating the security of the myriad of devices being deployed into the AMI. They show that they can leverage focused penetration efforts in one vendor to others, and explore where such evaluations must focus on the unique artefacts of a system under test. This work provides a comprehensive but high-level classification of attacks targeting AMI.

As a result, to the best of our knowledge, there is no published research that particularly addresses IDS for the NAN. In this work, IDS is proposed for the NAN which can be considered as the core

part of AMI. The proposed IDS is an anomaly-based solution which considers the constraints and requirements of NAN. The IDS captures the communication overhead constraints as well as the lack of a central point to install an IDS on it by proposing a distributed IDS that is run on some nodes which are powerful in terms of memory, computation and the degree of connectivity.

Realistic Intrusion Scenarios

One of main incentive to attack smart grid is energy fraud in which attackers try to tamper with metering infrastructure so that they are not billed for the energy they consume. The attempt to disable metering-related functions falls into the denial of service (DoS) category of attacks. One of the important DoS attacks that occurs in NAN prevents meters from acting on commands such as usage queries, firmware updates and remote disconnects. Figure 2 shows a typical DoS attack on meter command execution. A realistic example for this type of an attack is when a smart meter is failed to respond to a usage query and a malicious customer takes advantage of not being billed for some amount of time. The adversary has two choices to do so; either prevents the command from execution or prevents the command from reaching to the target smart meter. In former, adversary can either exhaust the system resource e.g., allocating and maintaining the maximum allowed number of open connections or by leveraging a firmware bug causing a system hang [11]. Another situation is when the adversary tampers with the forwarding of packets away from the meter by dropping traffic destined for that meter that can happen at link and routing layer at the back haul network (WAN) and NAN. An adversary can also prevent the packets from reaching his home smart meter by malfunctioning a middle smart meter which is one of the next hops of his own meter toward utility center.

The main focus of this work is on DoS attacks that occur in NAN as a result of the en route meter nodes that may malfunction and interfere with the proper forwarding of packets (e.g., by delaying, altering, misrouting and dropping.) Such smart meters are either spoofed or under attack. DoS attacks can be launched against physical layer by using radio jamming (e.g., a source of strong noise) which may interfere with the physical channels and hinder the availability of the network. Examples of such an attack include trivial jamming, periodic jamming and reactive jamming. At the MAC layer, a compromised node may not follow the agreed-upon frequency-hopping which will result in a large number of collisions. Unprompted CTS (Clear To Send) and reactive RTS (Request To Send) jamming attacks are examples of DoS attacks that occur at the MAC layer [15]. At the network layer, black hole, grey hole and wormhole attack can be performed by a malicious node. Such attacks will cause the packets to be dropped or misrouted. Another attack that may occur in the NAN, is when the attacker transmits a flood of packets toward a target node or congests the network and reduces its performance.

Table 1 shows some of the possible threats.

Table 1: Description of possible threats.

Threat	Threat description
1	Signal jamming at the PHY level
2	Packet collision at the MAC level
3	Misrouting and packet dropping attacks (e.g., black hole, wormhole, grey hole, …)
4	Packet flooding

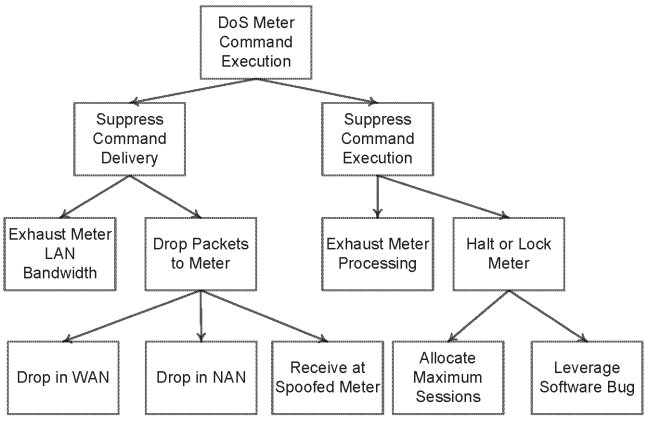


Figure 2: DoS Meter Command Execution Tree, adapted from [11].

An IDS which acts as a second wall of defence is necessary for protecting smart grid if security mechanisms such as encryption/decryption, authentication and etc. are broken. Generally, techniques for intrusion detection are classified into three main categories:

Signature- or pattern-based, which rely on a predefined set of the so-called attack patterns or signatures to identify attacks. Such techniques are often summarized as: what is bad, is known – what is not bad, must be good.

Anomaly-based, which rely on statistical knowledge and perhaps also particular models of correct node behaviours and mark nodes that deviate from these models as malicious. Such techniques are often summarized as: what is usual, is good – what is unusual, must be bad.

Specification-based, which rely on predefined behavior (often using a set of constraints and monitor the execution of programs/protocols with respect to these constraints. Such techniques are often summarized as: what is good, is known – what is not known, must be bad.

Out of these categories, anomaly detection performs best when there is a potential for unknown attacks to occur [6]. As noted above, anomaly detection uses statistical knowledge of correct node behaviour and flags behaviour that deviates from normal system use. A typical anomaly detection system takes in audit data for analysis. The audit data is transformed to a format statistically comparable to the profile of a user. Initially, the user's profile is generated dynamically by the system and it is subsequently updated based on the user's usage. Thresholds are always associated to all the profiles. If any comparison between the audit data and the user's profile results in a deviation that crosses a set of threshold, an intrusion alarm is set [9, 16]. The fundamental reason for choosing an anomaly-based IDS for NAN is because of the existence of many unknown attacks that target the NAN and the number of such attacks will most likely increase as the smart grid becomes more widespread. Therefore, the IDS should be capable of detecting not only existing attacks but also new attacks.

Proposed IDS for NAN

Figure 3 shows a typical NAN in which smart meters are connected in an adaptive wireless mesh network and all of them can perform routing. Each node maintains a list of parents so that in case of a failure of one parent, it can switch to the next available parent. Hence, redundant paths make the network more reliable. A fully redundant routing requires both spatial and temporal diversity; spatial diversity refers to enabling each smart meter to discover multiple possible parents and then establish link to two or more. Temporal diversity refers to fail-over and retry mechanisms [13]. RPL and geographical routing protocol are two popular candidates that can be used in such a RF mesh network [4, 8].

The proposed solution is a distributed and hierarchical anomaly-based IDS. The reason for proposing a distributed IDS is that the metering network is resource constraint; smart meters have limited computation power and they cannot spend more energy monitoring their neighbours. Moreover, if all smart meters are to run IDS, they are always busy sending monitoring messages to their supervisor nodes and this is not possible in the low bandwidth network that exists between smart meters. That is why the proposed solution requires only a subset of nodes to run IDS.

The proposed IDS embraces three different IDS nodes, namely, field IDS, WAN IDS and central IDS. Field IDSs are run on the collectors as well as some smart meters whose connectivity degree is between certain thresholds. Such smart meters should also have extra memory compared to ordinary smart meters so that they can be capable of monitoring their neighbours in addition to normal functions. Note that, each smart meter should be directly connected to at least one IDS node. Field IDS nodes should be tamper resistant as nowadays most smart meters are. Field IDSs are responsible for passively monitoring the communication of the neighbour smart meters to collect trace data. They provide reports of detected attacks to central IDS in utility center. Another option is that field IDSs send detection messages to base stations residing in the WAN. The WAN base stations that act as bridges between NAN and WAN are assumed to have sufficient computational power and memory, so that they can run WAN IDSs. WAN IDSs are responsible for the incoming and outgoing traffic from and to collectors and, in case of intrusion detection, they report the malicious collectors to the central IDS. Central IDS resides in the utility cen-

ter which is responsible for making global decisions based on alarms and notifications coming from the WAN and field IDSs.

The proposed IDS has three phases; data gathering, compliance check, inference that are explained in the followings. A feature set is selected from the intrinsic and observable characteristics of communications to distinguish normality from anomaly.

Phase 1 (data collection phase): in this phase, field IDSs listen on the communication of neighbour nodes and check them to see if there is any abnormal behaviour in their communication. WAN IDSs also check the communication coming from the collectors seeking for unusual activities. Central IDSs also check the communication of WAN access points and make sure about the healthiness of their communication. The communication information about each neighbour can include, but not limit to, number of transmission attempts, number of ACKs received, number of received packets and etc.

Phase 2 (compliance check phase): IDS nodes extract the data from phase 1 and perform compliance check with the normal behaviour.

Phase 3 (inference phase): After finishing phase 2, the results are sent to an inference part to derive the final decision in order to see whether the detected anomaly is a malicious attack or it is just a transient failure. To make accurate decisions in this phase, the IDS node must keep the history of the monitored nodes to distinguish between occasional network failures from real attacks.

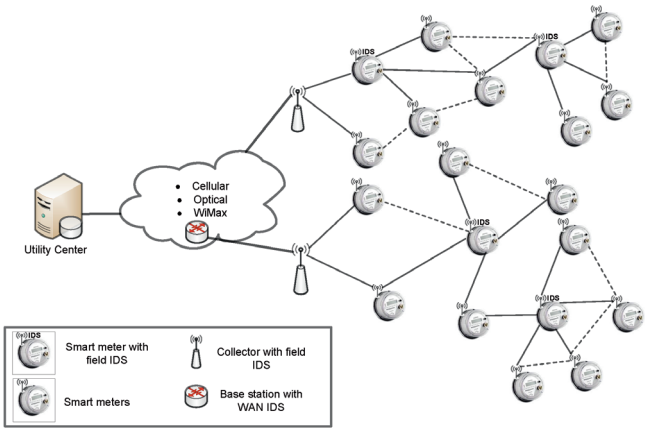


Figure 3: Neighbourhood Area Network IDS.

When an intrusion is detected, the system should take appropriate actions in response to an attack. Passive response is typical in the IDS in which the information is logged of and there is also a real-time notification. However, since NAN comprises wireless networks and the devices are located in insecure places, there should be an active response in place. If detected threat reaches a certain confidence level, required counter measures should be taken. For instance, in case of jamming attack in MAC layer, central office in substation will send a control message to the target meter to change its transmission channel.

Note that it is assumed that the communication between nodes is secure and IDS nodes are authenticated with each other using digital signatures. It is also assumed that there is an Access Control List (ACL) that all nodes have unique link keys associated with their unique IDs.

Policy Rules

In this section, we discuss in detail the policy rules which are used to detect anomalies in the system.

The IDS node should monitor the number of packets its neighbours transmit in number of bytes. Since the number of communication message types (e.g., firmware updates, usage queries and responses, offers and etc.) between smart meters and utility center is not infinite, therefore, the size of exchanged data between smart meters and utility center can be determined. Any size of data beyond the maximum value can be tagged as a suspicious message. If the number of such messages exceeds a certain threshold, IDS node should raise a flag indicating a potential threat. An example of such an attack is flooding attack. Such a rule can be implemented at field IDSs, WAN IDSs and central IDS.

Transmission power level is another parameter that can be used to detect a signal jamming attack at physical layer since the level of power for transmission is a pre-configurable parameter for deployed nodes. The IDS node can monitor its neighbour to detect any deviation from the accepted levels. Such a rule can be implemented at field IDS since the central IDS cannot monitor such a feature.

Field IDS nodes monitor the frequency/channel hopping sequence. Two nodes should agree on the frequency hopping sequence for the time slots in which they want to communicate with each other. As a result, if there is any sequence-nonconforming nodes, the node IDS should raise a flag indicating a DoS attack at MAC layer. Field IDS should apply this rule.

MAC delay transmission is another characteristic that can be monitored by the field IDS. When the utility center issues a command to a smart meter, the smart meter should respond by a certain delay. If the smart meter does not respond in the expected time frame, the field IDS should tag the smart meter node as a suspicious one and watch for more such anomalies. An example is when the smart meter is under jamming attacks and cannot transmit the data by the expected time out.

The central IDS should look for normal behaviour of smart meter applications for sending ACKs. Only the central IDS can check this feature, since the application data is encrypted in the transmission layer (e.g., using SSL) and it can be decrypted only at the utility center. Therefore, if there is a large number of missing ACKs and retransmissions, the central IDS should tag the smart meter as a suspicious one. Next, the central IDS launches an investigation to identify the source of malicious activity using lower level IDS nodes. By probing the nodes along the path to the suspicious node, the source of problem will be detected. An adversarial case is where one of the next hops of the smart meter is intentionally dropping the packets destined for that meter.

Field IDS should monitor the layer at which nodes are communicating. Since smart meters are supposed to communicate with each other only at the network layer, any smart meter's attempt to communicate with its neighbour at a different layer should raise a flag. An example of such an attack is warm hole attack in which the malicious node tries to send the traffic to some illegitimate destinations.

WAN and field IDS should monitor the request/reply pattern that is coming from the central office and smart meters. Requests must only arrive from the central office and responses must be directed to central office. If a request is coming from another source or the smart meter is trying to send the packets somewhere different from the central office, IDS nodes should alarm and notify the central office.

Table 2 links the threats listed in Table 1 with the applicable IDS detection rules outlined above.

Table 2: Threats and Corresponding Rules.

Threat	Rule 1	Rule 2	Rule 3	Rule 4	Rule 5	Rule 6	Rule 7
1	✓	✓					
2			ü	ü			
3				✓	✓	✓	✓
4	✓			ü			

Conclusion

The development of practical and efficient IDS for smart grid is highly crucial. While reasonable amount of research has been done in designing and implementing of IDS for different parts of smart grid, there is a critical need for designing an IDS for the NAN part of the smart grid. The insecure places where metering devices are located increase the potentials for intrusions within the grid. This work focuses on designing an IDS for NAN by taking the constraint of NAN into account. The proposed IDS scheme is a distributed anomaly-based solution which looks for anomalies at different layer of network stack by applying a set of rules. In case of detecting an attack, the IDS will raise an alarm highlighting the malicious activity.

In order to measure the detailed performance of the proposed IDS, such as false positive and false negative rates, detection time and the ability to differentiate between transient failures and malicious behaviours, we need a more detailed analysis of the IDS solution. Furthermore, we plan to expand the threat model to capture more adversarial cases and examine the proposed IDS using a suitable simulator such as OPNET [12].

References:

[1] Berthier, R. and Sanders, W.H. (2011) Specification-based intrusion detection for advanced metering infrastructures. *Proc. IEEE 17th Pacific Rim Int. Symp. Dependable Computing (PRDC'2011)*, pages 184–193, Dec. 2011.

[2] Berthier, R., Sanders W.H. and Khurana, H. (2010). Intrusion detection for advanced metering infrastructures: Requirements and architectural directions. *Proc. IEEE SmartGridComm*, pages 350–355, Oct. 2010.

[3] Cleveland, F.M.(2008). Cyber security issues for advanced metering infrasttructure (ami). In Power and Energy Society General Meeting -Conversion and Delivery of Electrical Energy in the 21st Century, 2008 IEEE, pages 1–5, Jul. 2008.

[4] Iyer, G., Agrawal, Monnerie, P. E. and Cardozo, R.S. (2011). Performance analysis of wireless mesh routing protocols for smart utility networks. *IEEE SmartGridCom*m, pages 114–119, Oct. 2011.

[5] Jokar, P., Nicanfar, H. and Leung, V.C.M. (2011). Specification-based intrusion detection for home area networks in smart grids. *IEEE SmartGridComm*, pages 208–213, Oct. 2011.

[6] Komninos, N. and Douligeris, Ch.(2009). LIDF: Layered intrusion detection framework for ad-hoc networks. *Ad Hoc Networks* **7**(1):171–182, 2009.

[7] Kush, N., Foo, E., Ahmed, E., Ahmed, I. and Clark A. (2011). Gap analysis of intrusion detection in smart grids. In Craig Valli, editor, 2nd International Cyber Resilience Conference, pages 38–46. secau-Security Research Centre, Aug. 2011.

[8] Lichtensteiger, B., Bjelajac, B., Muandller, C. and Wietfeld, C. (2010) RF mesh systems for smart metering. *IEEE SmartGridComm*, pages 379–384, Oct. 2010.

[9] Liu, Y., Li, Y. and Man, H.(2005). Mac layer anomaly detection in ad hoc networks. In Information Assurance Workshop IAW'05. pages 402–409, Jun. 2005.

[10] Lu, Z., Lu, X., Wang, W and Wang, C. (2010) Review and evaluation of security threats on the communication networks in the smart grid. *Proc. MILCOM 2010*, pages 1830–1835, Nov. 2010.

[11] McLaughlin, S., Podkuiko, D., Miadzvezhanka, S., Delozier, A. and McDaniel, P.(2010). Multi-vendor penetration testing in the advanced metering infrastructure. Proc. 26th Ann. Comp. Security Applications Conf., pages 107–116, 2010.

[12] OPNET Technologies, Inc. Opnet modeler 17.1. Website, Mar. 2011. http://www. opnet.com.

[13] Roosta, T., Nilsson, D.K., Lindqvist, U. and Valdes, A. (2008). An intrusion detection system for wireless process control systems. Proc. 5th IEEE Int. Conf. Mobile Ad Hoc Sensor Syst. MASS'2008, pages 866–872, Oct. 2008.

[14] Schuba, C.L., Krsul, I.V., Kuhn, M.G., Spafford, E.H., Sundaram, A.and Zamboni. D. (1997). Analysis of a denial of service attack on TCP. Proc. IEEE Symp. Security and Privacy, pages 208–223, May 1997.

[15] Seth, S. and Gankotiya, A.(2010). Denial of service attacks and detection methods in wireless mesh networks. In Int. Conf. Recent Trends in Information, Telecommunication and Computing (ITC), pages 238–240, Mar. 2010.

[16] Xu, R., Li, J., Zhang, F. and Levy, R. (2006). Model selection for anomaly detection in wireless ad hoc networks. 12th International Conference on Parallel and Distributed Systems ICPADS'06, (CD-ROM), 2006.

[17] Yan, Y., Qian, Y. and Sharif, H.(2011). A secure and reliable in-network collaborative communication scheme for advanced metering infrastructure in smart grid. Proc. IEEE Wireless Comm. Networking Conf. (WCNC'2011), pages 909–914, Mar. 2011.

[18] Zhang, Y., Wang, L., Sun, We., Green, R. C. and Alam, M. (2011). Distributed intrusion detection system in a multi-layer network architecture of smart grids., *IEEE Transactions on Smart Grid* **2**(4):796–808, Dec. 2011.

Submitted: April16, 2011
Accepted: June 1, 2011

Analyzing the Cost and Benefit of Pair Programming Revisited

Lev Faynshteyn, Vojislav B. Mišić, Jelena Mišić
Ryerson University, Toronto, Canada

Contribution to the State of the Art
DOI: **10.7251/JIT1201014F** UDC: **502.171:551.782**

Abstract: Pair programming has received a lot of attention from both industry and academia, but most paper focus on its technical aspects, while its business value has received much less attention. In this paper, we focus on the business aspects of pair programming, by using a number of software development related met rics, such as pair speed advantage, module breakdown structure of the software and project value discount rate, and augmenting them by taking into account the cost of change after the initial product release and inherent non-linearity of the discount rate curves. The proposed model allows for a more realistic estimation of the final project value, and the results of System Dynamics simulations demonstrate some useful insights for software development management.

Keywords: Pair Programming, Extreme Programming (XP), System Dynamics, Waterfall, Cost of Change

Introduction

Pair Programming (PP) is one of the key paradigms in Extreme Programming (XP). It stipulates that any coding task should always be performed by a pair of programmers working at the same computer using only one set of input devices. This approach has been demonstrated [7][11][21][22] to provide tangible benefits in such areas as design and code quality (fewer bugs per line of code), problem solving (two heads are better than one) and general satisfaction with a job well done (people like to share responsibility, which in turn makes them feel more confident and comfortable). At the same time, benefits related to the improved productivity have not been fully corroborated [10], and experience has shown that to leverage the full potential of Pair Programming, that is to keep the original level of task parallelism in any given company while taking advantage of all the benefits, the number of developers has to be doubled, which in most cases also means doubling the personnel costs. Naturally the question arises when and under what conditions the additional expenses are justifiable. To answer this question Padberg and Müller created a mathematical model [18] which is based on three categories of metrics used as input to the model:

- *Process Metrics:* productivity of a single developer, pair speed advantage (PSA), defect density of code, pair defect advantage (PDA) and defect removal time.
- *Product Metrics:* product size and module breakdown structure of the software.
- *Project Context Metrics:* project value discount rate, initial asset value, number of single developers, number of programmer pairs, developer and project leader salaries, monthly working hours.

By analyzing these metrics and studying their relationships they came up with a mathematical expression for Net Present Value (NPV) of a software project, which in a nutshell repre sents the initial monetary value of a project (AssetValue) discounted at a certain rate (DiscountRate) minus the development expenses (DevCost) throughout the entire duration of the development process (DevTime):

$$NPV = \frac{AssetValue}{(1+DiscountRate)^{DevTime}} DevCost$$

By adjusting the model's input parameters (mostly PSA, PDA and MP) they collected results representing various configurations for both projects developed under traditional software development practices (e.g. a Waterfall process) and projects that utilized Pair Programming. Their conclusion was that for projects where PSA is moderate and MP is not very high, conventional development methods will produce better financial returns. In fact, even in cases where discount rate figures are very high (e.g. 75% per year), PSA is large (such as 1.8 times of a single developer productivity) and PDA is quite significant (15% or more less bugs in the code), Pair Programming would just break even with conventional practices. However, successful real life applications [5][11][12][20] of XP practices offer ample evidence that Pair Programming does work and is certainly economically feasible. Naturally a question arises whether the original model is missing on some aspects of XP in general and Pair Programming in particular that might change the balance in favor of the latter. The following sections will try to address this question.

Motivation

According to the original model, Pair Programming will be economically feasible only in extreme cases, where time to market is absolutely critical (i.e. project value discount rate is extraordinarily high). However this, as the examples in the previous section show, is often not true and XP, and Pair Programming as an inherent part of it, are used for projects of all scales and time durations, with many showing positive results in terms of both productivity and profitability. While looking for an explanation of this discrepancy we came to realize two things:

First of all, discount rate in the original paper is always a constant value. This seems unrealistic since discount rate itself is subject to many factors. For instance it would be reasonable to assume that for a brand new product not only would it be very high, but after some point it would accelerate at a much higher rate than initially due to the fact that market rivals would have released or would be drawing ever so closer to releasing a competing product. At the same time for a well established product the acceleration would be very slow at first since the established user base would be unwilling to upgrade too often and conversely they would be willing to wait for quite a long time for an update for a product that has already proved itself. However after a certain moment in time it would also start to accelerate at a faster rate, since going beyond a certain point in time without a new version would test customers' patience. These ideas are in fact confirmed by the real world data [14] and thus this change will be a good candidate for an improved model.

A second and probably more important observation was that the original model did not consider the cost of change (CoC) of the code after the initial release date. In 1981 Barry W. Boehm did a study [3] of the cost of change ratio between implementing a feature or fixing a bug in production vs. requirements stage and found it to grow exponentially with time. Even for projects of moderate size it could be very high (up to 100 times and more). His much more recent book [4] confirms these numbers. This makes sense for traditional approaches where the requirements and features are for the most part determined once at the beginning of a project and stay the same throughout the whole development cycle until the software is released. Any new feature requests are being deferred until after the release, thus making their implementation potentially very difficult and labor inten-sive.

On the other hand, in XP the development process starts with only a general idea about how the final product will look like or function, and is constantly refined by means of customer feedback. It thus allows in a way to defer the cost of making big and costly decisions early on and to have the best chance that once these big decisions are made they would be the right ones. This is the premise on which Kent Beck in [2] based his argument that for XP the curve of cost of change is way more shallow than for the traditional methods, and the actual costs of changes in production vs. requirements phase can be as low as five and would stay close to these low values for extended periods of time (in fact this might be the very reason why XP is economically feasible).

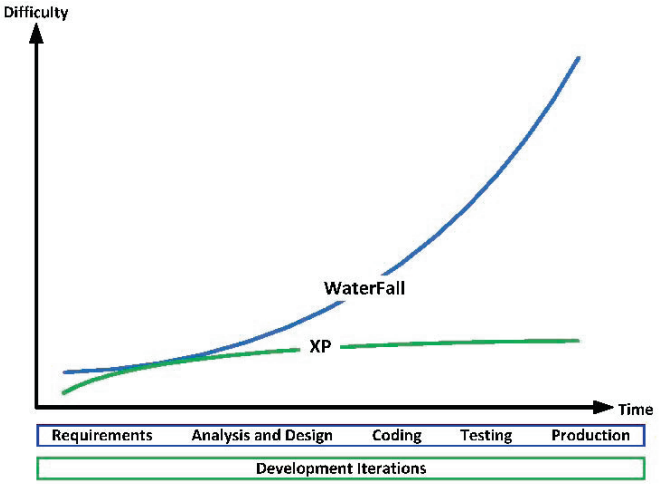


Figure 1. Cost of Change in cases of a Waterfall and an XP development approaches (after[2])

Fig. 1 shows a qualitative comparison of the curve of CoC for a hypothetical software project done using a traditional (e.g. Waterfall) and an agile (e.g. XP) approach.

An example in [4] of a project (TRW CCPDS-R) that employed an innovative hybrid approach to the development process, in which both the traditional and agile practices were employed, seems to confirm this supposition, as the design, implementation and maintenance changes throughout the lifecycle of the project remained at a very low level.

Considering all of the above it seems reasonable that a new version of the model should include a CoC metric since it seems quite probable that it might significantly affect the results of simulations.

To implement a modified version of the model a System Dynamics approach was chosen. System Dynamics is a proven technique that allows flexible and efficient exploration and analysis of the behavior of complex systems over time by describing them in terms of interconnected elements that continually interact with each other and the outside world to form a unified whole [15]. Due to the inherent dynamic nature of software development processes and their often complex interrelations, Systems Dynamics has been long recognized as a very potent approach to modeling of the former, often resulting in exposure of surprising non-linearities in models of even modest dimensions.

Related Work

To date, many areas of agile development practices have been analyzed in numerous publications, spanning all the way from studies in the area of psychology and pair compatibility [9] to more generic evaluation of the effects of the learning phase in the context of XP on productivity [17] to the attempts to model the entire XP development process [18]. However, since the scope of this paper mainly deals with the economic benefits of Pair Programming, we will focus our attention on the most relevant publications.

In [19] Padberg and Müller extended their model for NPV to include the effects of the learning phase (inherent to the Pair Programming) on the final value of the project. Their results showed that due to the fact that the learning phase typically incurs a onetime cost and the learning process itself does not take long, the overall effect of the former is typically minimal and amounts only to a few percent of the total project cost: the learning overhead did not exceed 10% even in cases with very high staff turnover. Thus, the estimates and conclusions in the original paper remained largely unchanged.

In [8] the authors propose a metrics-oriented evaluation model that allowed them to assess a chosen development model based on the project's predicted NPV value. The proposed NPV formula takes into account such variables as development time and cost, asset value, operation cost, flexibility value and product risk. However, this model deals with the high-level representation of the underlying development process and as such does not reflect the intricacies of any particular approach (whether it be traditional or agile method).

In [13] the attempt is made to test the validity of the supposition that the cost of change curve in case of the agile development practices is indeed much flatter than that of the traditional approaches. The authors employ a System Dynamics approach to build a fairly involved model in which the main criterion of effi ciency is the number of requested vs. the number of implemented user stories. However, the results and conclusion sections are very scarce and superficial and fail to elaborate on the actual out-

come of the simulation runs, thus leaving the question unanswered.

Various forums and software development websites (e.g. [1][6]) have discussions related to the nature of the cost of change and its effects on the development costs in particular and feasibility of different project development strategies as a whole. However, as of today, the opinions vary wildly and conclusions seem to be based on mostly anecdotal evidence and common sense. Thus, the authors of this paper understand that the proposed enchantments are merely an educated guess and reflect their subjective opinion on the matter.

The Extended System Dynamics Model

The implementation was done in GoldSim, a simulation package by GoldSim Technology Group, which is a quite powerful and feature rich Monte Carlo [16] simulation suite. To account for both traditional and agile approaches (later referred to Waterfall and XP respectively) two separate models were created. To keep the results of the simulations consistent both models share the same set of input metrics, which in their turn, to make results comparable to the ones in the original paper, were kept the same (see the original paper [18] for details and the rationale behind selecting the particular values) and are presented in Table 1.

Both models contain elements that correspond to a defect generation process (defects are produced at a defect density rate depending on the volume of the written code at any given moment). These bugs cause additional workload for the developers (their approximated number of LOC is added to the initial prod uct size) and thus, the defect removal time metric of the original model is implicitly expressed through a dynamic feedback loop.

In addition, both models implement a concept of a code backlog. The idea behind it is that as time goes by customers will be asking to introduce new features into the system that is currently being developed. In case of a Waterfall process all of these features will be delayed until after the initial release, thus creating a code backlog, which basically consists of a sum of all approximated numbers of LOC needed to implement all of the features at the time when they are requested. Depending on a chosen market pressure curve and a product release date, this aggregate number will be multiplied by the CoC value and the work will continue, marking a new development period with additional expenses for the company. In case of Pair Programming the backlog will be much smaller since user requirements will be, for the most part, implemented and integrated into the system during its development stage. Note that not all of the fea-

Table 1. Input metrics and their values (after [18]).

Input Metric Name	Value/Range	Comments
Productivity of a Single Developer	350	Lines of code (LOC) per month
Pair Speed Advantage (PSA)	1.4-1.8	A speed advantage of a pair of programmers as compared to a single programmer
Product Size	16800-100000	Initially estimated size of the project in LOC
Defect Density	0.03	Defects per LOC
Pair Defect Advantage (PDA)	5% - 15%	A quality advantage a pair of programmers has as compared to a single programmer
Module Breakdown Value	8	Maximum developer/pair parallelism in the scope of the project
Discount Rate	0 - 1	Various linear and non-linear values as functions of time (see below for details)
Asset Value	1000000 – 3000000	Initial amount agreed upon for the development of a project in $USD
Number of Single Developers	Equal or less than a Module Breakdown Value	
Number of Programmer Pairs	Equal or less than a Module Breakdown Value	
Developer Salary	50000	in $USD
Project Leader Salary	60000	in $USD
Monthly Working Hours	135	

Table 2. Newly introduced input metrics and their values approaches.

Input Metric Name	Value/Range	Comments
User Stories	Random	A new user story is randomly generated (uniform distribution) approximately once every 1.5 month
Feature Size	Random	A size of a new feature in LOC is generated randomly (exponential distribution) with big features progressively less likely to occur vs. small features
Feature to Backlog Ratio	0.2 – 0.3 for Waterfall 0.05 for XP (PP)	Represents percentage of the Code Backlog that will actually be implemented after the initial release
Cost of Change Curve	X^3 growth for Waterfall ln(X) growth for XP (PP)	Changes with time according to the relevant equation. Please see the models for the actual values.

tures in the code backlog will be implemented (some of them will be covered by other features, others will be dropped, etc.). This fact is represented by another input metric called Feature to Backlog Ratio. Table 2 lists the newly introduced metrics that are used in the CoC related part of the model.

One more important difference between the new and the original model lies in the fact that for the new Pair Programming model a concept of refactoring has also been implemented. Refactoring happens whenever a new bug is reported, a new feature is added or when the number of both bugs and new features introduced into the system exceeds a certain value [2] (this in XP circles is sometimes referred to as "when the code start to smell"; in this particular case the values are 5 for bugs and 10 for features).

The basic idea behind both models is the same: the initial estimated size of the project gets chipped away at a development rate that depends on the productivity and team size. As the code is being generated, bugs start to appear according to the predefined defect density and features are requested according to the predefined random distribution. The weights of the bugs and features (that is how many LOC each of them will take to fix/implement) are also randomly determined according to separate random distributions. The resulting values are added to the total pull of work (for XP features, for the most part, are added right away, for Waterfall they go into the code backlog). When the size of the project goes down to zero (that is there is no more work to be done) for the first time, we reach a stage of the first release. At this point, a second part of the model activates that determines how many LOC it would take to clear up the code backlog considering the current value of the CoC. The project size depository gets refilled with the newly calculated value for the LOC and the work resumes in the same way as earlier, except that new features are no longer accepted. Note that it is also possible to run multiple realizations for each of the models by specifying the number of Monte Carlo stages. This allows us to see how such random input variables as features and bugs affect the results of the simulations.

Results

Since the newly created Waterfall model is basically identical to the model in the original paper [18] we can use it as a gauging device to see if the results produced by it are comparable to the reference results in the original publication. Using the original model values (see Table 1.) for a product with 16800 LOC, an asset value of 1000000 dollars and

Table 3. Results of the simulations

PSA	PDA	System Size	Asset Value	Discount Rate	NPV_{WFR} in Days	NPV_{WFB} in Days	NPV_{PPR} in Days	NPV_{PPB} in Days
1.4	5%	16800	1000000	Constant 10% per year	**710016** in **192**	**520663** in **318**	**355379** in **247**	**292323** in **271**
1.4	15%	16800	1000000	Constant 10% per year	**710016** in **192**	**520663** in **318**	**360379** in **245**	**297820** in **268**
1.8	15%	16800	1000000	Constant 10% per year	**710016** in **192**	**520663** in **318**	**704761** in **113**	**704085** in **114**
1.4	5%	30000	3000000	Constant 10% per year	**2085645** in **455**	**1186641** in **903**	**1650857** in **435**	**1566864** in **462**
1.8	15%	30000	3000000	Constant 10% per year	**2085645** in **455**	**1186641** in **903**	**2050841** in **306**	**2013761** in **318**
1.4	15%	30000	3000000	**New** product (very steep drop after about 200 days)	**369769** in **455**	**-834246** in **903**	**48425** in **428**	**- 120590** in **454**
1.4	15%	30000	3000000	**Mature** product (slow acceleration of pressure)	**2254082** in **455**	**1279895** in **903**	**1833987** in **428**	**1759799** in **454**

a constant discount rate of 10% per year, the reference NPV for conventional development process was estimated at 723,463 dollars. The modified Waterfall model produced an average NPV value (for the release date) of 710016 dollars, which considering the random nature of the simulation is close enough. Having established that reference point, let us now see how different discount rate, project size, PSA and PDA values affect the results of the simulations.

To compare the results of the new model simulations to the results presented in the original paper [18] a number of runs (each consisting of 100 realizations) with different input values were executed. The results are given in Tables 3 (all of the parameters, except the ones listed in the table, were kept the same throughout all of the runs), where NPVWFR, NPVWFB, NPVPPR, NPVPPB are NPV values for Waterfall Release, Waterfall Backlog, Pair Programming Release and Pair Programming Backlog milestones respectively.

The first batch of experiments was run at a constant yearly discount rate of 10%. As can be seen from the table, for a project of a relatively small size of 16800 LOC, conventional development methods prove superior when PSA is kept at a reasonable level of 1.4 (PDA variations have very limited effect on the results, thus they are largely disregarded in the discussion). So far this is in line with the results of the original paper, though one interesting point to note is that even at this low level of PSA, the Pair Programming model has finished processing its code backlog considerably earlier than the Waterfall model. Increasing PSA to 1.8 (the same highest value that was used in the original paper) changes the picture quite a bit: now Pair Programming basically breaks even with Waterfall model for the release date in terms of money, and considerably outperforms it in terms of simulation durations in both release and backlog cases.

Increasing the project's size by roughly three times and performing the same tests shows us that for bigger projects (i.e. those that will take longer to deliver) even with modest levels of PSA, Pair Programming often breaks even (a little less money at the release date, but on the other hand release is done some-

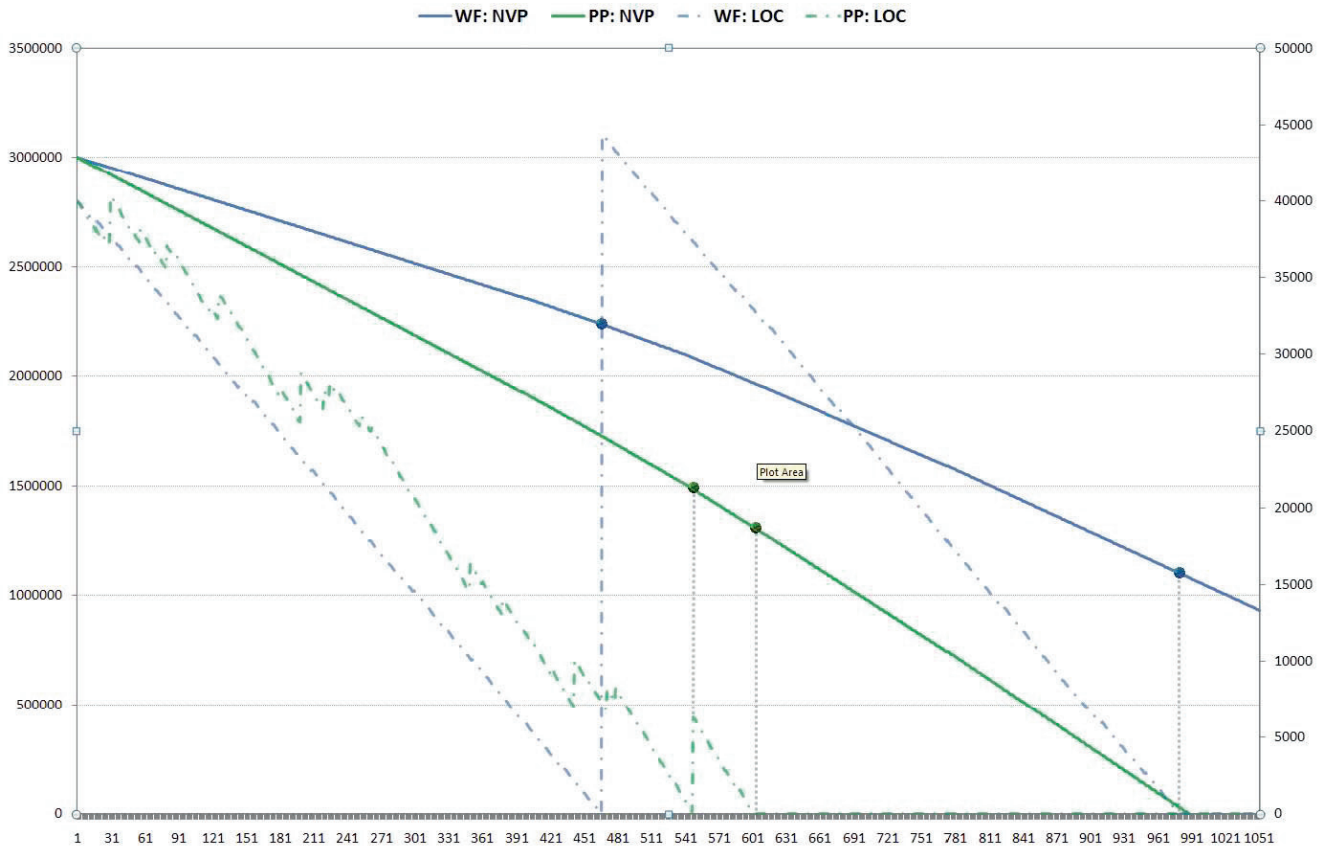


Figure 2. Examples of NPV values obtained using the Pair Programming (green) and Waterfall (blue) models

what earlier) or actually pulls ahead of a conventional model. Due to the ever accelerating CoC curve for the Waterfall model, this gap will be only growing. For example increasing the size of a project to 50000 LOC brings us to a situation where a team of 8 programmers is unable to clean up the Code Backlog within a maximum simulation period of 1050 days, whereas a Pair Programming team can finish it in less than 600 days with 1.2 million dollars in profits.

The simulations with non-linear discount rates confirm earlier observations. Namely, Waterfall might win some runs on the release day, but it will lose everything later due to a huge code backlog. Note that the Feature to Backlog ratio was actually kept at a low level of 0.2 (only 20% of the feature related code was implemented in the course of the backlog stage), thus actually favoring the Waterfall model.

Fig. 2 shows examples of two realizations from one of the runs in case of a mature product for both the Pair Programming and Waterfall models. The solid green and blue lines represent the NPV values expressed in $USD for Pair Programming and Waterfall model, respectively. The first solid dot on each line corresponds to the NPV value at the moment of the initial software release, and the second solid dot corresponds to the NPV value at the moment when the entire code backlog has been taken care of. Similarly the dash-and-dot green and blue lines represent the amount of coding that still remains to be done at any given moment in time expressed in LOC. The left vertical axis shows the amount of $USD, the right vertical axis is the LOC number, and the horizontal axis is the time of the simulation expressed in days.

Conclusions and Future Work

As the results of the previous section show, conventional development approaches such a Waterfall model can prove to be a better choice in cases of a smaller project with relatively low rate of new feature requests. However even with a low feature request rate used in the models (1 new feature every 1.5 months) and a low Feature to Backlog ratio value of 0.2, it is struggling to keep up with the Pair Programming model. Even for modest values of PSA (1.4 is actually a very realistic value [2][18] confirmed by several sources) Pair Programming proves to be a better approach: the initial release dates are close enough to the ones obtained using conventional methods, while the ability to quickly clean up the backlog will be a real boon for any company. Also note that even if Pair Programming losses on paper moneywise, it often delivers the product earlier (for example in Table 3 there are cases where the NPVPPR is less than NPVWFR, but "time to market" is shorter) and though it is not quantifiable in the scope of this model it has to be worth something in real life.

That having been said, the results of this simulation should be taken with a grain of salt. First, many values, especially those related to the new feature generation and code backlogging processes, are no more than educated guesses, which are mostly based on the authors' industrial experience. In real life, they are likely to vary considerably from project to project and company to company. However, the results are representative of qualitative trends.

Second, there is only limited evidence of what the actual CoC curves look like. In real life, too many variables, such as coding and managerial practices, technology and tools used, programmers' compatibility and expertise, etc. can affect their actual shape and values and definitely more research based on real life data is needed in this field.

Our future research will look into improvements of the model, for which there are quite a few possibilities. For instance, a feature generation rate can be made a function of the discount rate, thus reflecting the fact that customers usually want to see in the developed software the same or similar features to the ones competitors already have in theirs. At the same time this rate will have to be checked against some kind of a deadline/cutoff condition. Otherwise we might end up being swamped with features without hope of ever finishing the project. Finally, different parts of the model can also be broken down into smaller pieces to reflect the underlying processes with greater detail and accuracy. For instance, such an aspect of Pair Programming as pair switching and associated learning curve can be included in the model.

This task is simplified by the fact that the process of System Dynamics model conversion between different modeling suites is a pretty straightforward one, and thus any aspect of XP software development cycle implemented as a System Dynamics model to date can be readily converted into the necessary format and integrated with the current model with only minor investments in terms of both time and efforts.

References

[1] Ambler, S. W. (2003). Examining the Agile Cost of Change Curve, http://www.agilemodeling.com/essays/costOfChange.htm

[2] Beck, K. (1999). Extreme Programming Explained:Embrace Change. Addison Wesley.

[3] Boehm, B. (1981). Software Engineering Economics.Prentice-Hall.

[4] Boehm, B. and Turner, R. (2003). Balancing Agility and Discipline: A Guide for the Perplexed. Addison-Wesley Professional.

[5] Capiluppi, A., Fernandez-Ramil, J., Higman, J., Sharp, H. C. and Smith, N. (2007). An Empirical Study of the Evolution of an Agile-Developed Software System. 29th Int. Conf. on Software Engineering ICSE 2007, Minneapolis, MN.

[6] Cockburn, A. (2000). http://xprogramming.com/articles/cost_of_change/

[7] Cockburn, A. and Williams, L. (2000). The costs and benefits of pair programming. eXtreme Programming and Flexible Processes in Software Engineering XP 2000 .

[8] Erdogmus, H. (1999). Comparative evaluation of software development strategies based on Net Present Value. ICSE Workshop on Economics-Driven Software Engineering, Los Angeles, CA

[9] Hannay, J. E., Arisholm, E., Engvik, H. and Sjøberg, D. I. K. (2010). Effects of personality on pair programming. IEEE Transactions on Software Engineering 36(1):61-80.

[10] Hannay, J. E., Dybåa, T., Arisholm, E. and Sjøberg, D. I. K. (2009). The effectiveness of pair programming: a metaanalysis. Information and Software Technology 51(7):1110-1122.

[11] Hulkko, H., & Abrahamsson, P. (2005, 15-21 May 2005). A multiple case study on the impact of pair programming on product quality. Int. Conf. on Software Engineering ICSE'2005, St. Louis, MO.

[12] Khalaf, S. J. and Maria, K. A. (2009). An Empirical Study of XP: The Case of Jordan. Int. Conf. Information and Multimedia Technology ICIMT'09, Jeju Island, Korea.

[13] Kuppuswami, S., Vivekanandan, K. and Rodrigues, P. (2003). A system dynamics simulation model to find the effects of XP on cost of change curve. 4th Int. Conf. Extreme Programming and Agile Processes in Software Engineering XP 2003, Genova, Italy.

[14] Little,T. (2004). Value creation and capture: a model of the software development process. IEEE Software 21(3):48-53.

[15] Madachy, R. J. (2008). Software Process Dynamics. Wiley.

[16] Metropolis, N. and Ulam, S. (1949). The Monte Carlo Method. J. of the American Statistical Association 44(247):335–341.

[17] Mišić, V. B., Gevaert, H. and Rennie, M. (2004). Extreme Dynamics: Towards a System Dynamics Model of the Extreme Programming Software Development Process. ProSim'04 Workshop, Edingburgh, UK.

[18] Padberg, F. and Müller, M. M. (2003). Analyzing the cost and benefit of pair programming. IEEE METRICS pp. 166-179, Sydney, Australia.

[19] Padberg, F. and Müller, M. M. (2004). Modeling the impact of a learning phase on the business value of a pair programming project. 11th Asia-Pacific SoftwareEngineering Conf., Busan, Korea.

[20] [20] Shaochun, X. and Rajlich, V. (2006). Empirical Validation of Test-Driven Pair Programming in Game Development. IEEE/ACIS ICIS-COMSAR 2006, Honolulu, HI.

[21] Williams, L. (2000). The Collaborative Software Process. PhD dissertation, University of Utah, Salt Lake City, UT.

[22] Williams, L., Kessler, R. R., Cunningham, W. and Jeffries, R. (2000). Strengthening the case for pair rogramming. IEEE Software 17(4), 19-25.

[23] Yang, Y. and Bosheng, Z. (2009). Evaluating Extreme Programming Effect through System Dynamics Modeling. Int. Conf. on Computational Intelligence and Software Engineering (CiSE 2009), Wuhan, China.

Submitted: April 23, 2012
Accepted: June 7, 2012

Comparison of Examination Methods Based on Multiple-choice Questions Using Personal Computers and Paper-based Testing

[1]Sanja Maravić Čisar, [2]Robert Pinter, [3]Dragica Radosav, [4]Petar Čisar

[1]Subotica Tech-College of Applied Sciences, Subotica, sanjam@vts.su.ac.rs,

[2]Subotica Tech-College of Applied Sciences, Subotica, probi@vts.su.sc.rs

[3]Technical Faculty "Mihajlo Pupin" Zrenjanin, University of Novi Sad, radosav@tfzr.uns.ac.rs

[4]Telekom Srbija, Subotica, petar.cisar@gmail.com

Contribution to the State of the Art

DOI: **10.7251/JIT1201022C** UDC: **004.032.6**

Abstract: Computer-based testing, by facilitating the interaction between teaching and learning, can improve the quality of learning through improved formative feedback which is a key aspect of formative assessment. This study makes a contribution to the research on computer-based testing by examining the mode differences between the paper-and-pencil test and computer-based test. The previously conducted researches in this area dealt with the students of primary and secondary schools. In those researches the points of observation were the students' successes in mathematics, English and social sciences; no research was done in field of programming languages such as C++ with post-secondary students.

The main aim of this study was to find out whether there are differences in the achieved results in two ways of testing: computer-based testing and paper-and-pencil test. Also, the intention was to detect those characteristics of computer based test, which may have a negative effect on students' achievements. The participants were a representative sample of the population of all engineering students studying computer science at Subotica Tech. The findings of this study led the authors to reach the conclusion that there are no significant differences in scored results for the paper-and-pencil testing and the computer-based testing.

Keywords: computer-based test; paper-and-pencil test; assessment; testing; post-secondary education

Introduction

Traditional methods of assessment have limited capabilities in measuring the learning and progress of each student, especially in guiding the study process. These methods are particularly inappropriate today, when knowledge and the working environment change rapidly and complement each other, and the ability for independent lifelong learning is becoming more than necessary.

Modern technology offers many possibilities for improving the process of education and knowledge assessment. The history of using computers to perform the review process of knowledge begins with the 1970s [6]. However, the high price of computers at that time and their technical capabilities limited their application for testing. The progress of technology enabled the development and application of computers for testing in many areas, including the education process.

In the system of education, testing and evaluation of knowledge is of particular importance. Checking and evaluating knowledge enables teachers to deter-

mine the level to which students adopted the curricula and gained some knowledge and to get feedback about their work and applied teaching methods in order to improve it.

The marks are described as quantitative, numerical, qualitative, i.e. descriptive and by ranking or analytical. The criteria for evaluating the success of students are type, scope and level of approved knowledge, and skills in relation to what is prescribed by the curriculum of post-secondary institutions. In order to test whether evaluation has the proper effect, it is of great importance for the teacher's assessment of student knowledge to be accurate, objective and reliable.

The true strength of assessment is reflected in the feedback information to students. Improving the quality of the learning process involves not only the final determination of student knowledge at the end of the course, but more importantly the measurement of achieved knowledge during the course. Thereby students are more strongly motivated by their success in learning, they are taking more self-responsibility in the process of learning, they discover their "strong and weak points", and thus become active participants in learning.

The wide-spread popularity of computers resulted in directing attention to the possible use of computers in the process of knowledge evaluation. Advantages and benefits of this method of assessment and knowledge evaluation are various: the time needed to review the work of students is significantly reduced, there is the possibility of statistical analysis of questions, cost reduction in comparison with the validation of knowledge which includes printing tasks, the application of multimedia in setting questions, the possibility of measuring the time needed for response, and increasing the level of security.

However, all these advantages of computer-based testing become irrelevant, if it turns out that the test of knowledge with computers has side effects for individuals, i.e. it is not appropriate for all students.

Since there is an increasing number of schools in Serbia that have PC laboratory rooms, there is a growing interest in computer based assessments. However, there is also the ever-present question of the value and comparability of the results that are attained on computer tests and in the conventional way. The primary concern is whether the form of test delivery affects the results achieved by students on the test. For example, it is possible that the level of skills in computer use affects the final result of the test when compared with the result of the same test but in paper format.

The research that was done has an empirical - theoretical character. The problem into which the research was conducted was to investigate, whether the delivery of knowledge (computer or paper-and-pencil test) in the process of evaluation has a statistically significant impact on the results achieved and in increasing the quality of the teaching process. Following the research, students completed the questionnaire about their attitudes towards this kind of knowledge testing, in what way and whether the manner of presenting questions (one question or several questions simultaneously shown) had any impact on the achieved results.

Literature review

The use of computers in the process of testing began in the early 1970s. Initially, the technical capabilities of computers and their prices restricted the use of computerized tests. With the advantages that the new technology provides, this type of testing is beginning to develop, and consequently there are a number of researches that examine the role and application of computers in the process of knowledge evaluation.

According to the Guidelines for Computer-Based Tests and Interpretations from American Psychological Association (APA) [2], score comparability or equivalence between computer-based tests and paper-based tests is defined as follows: "Scores from conventional and computer administrations may be considered equivalent when (a) the rank orders of scores of individuals tested in alternative modes closely approximate each other, and (b) the means, dispersions and shapes of the score distributions

are approximately the same, or have been made approximately the same by rescaling the scores from the computer mode."

Lee and Hopkins [11] in their study found that the mean paper-and-pencil test score was significantly higher than the mean computerized test score. They also concluded that only software that allows the conveniences of paper-and pencil test, e.g., the ability to change answers and the ability to review past items, should be used in future applications.

The study of Shermis and Lombard [16] examined the degree to which computer and test anxiety had a predictive role in performance across three computer-administered placement tests (math, reading, written English). Results showed that age and test anxiety were both significant predictors for math performance, with lower values on the two variables associated with better performance. When reading was the outcome variable, age and computer anxiety were statistically significant performance predictors, with older readers faring better and less anxious individuals achieving higher scores. No predictors were statistically significant for the written English essay.

Nichols and Kirkpatrick [15] explored the impact of the mode of presenting the test for the Florida state assessment in high school reading and mathematics. They found that for both reading and mathematics, the mean raw score, mean scale scores, and passing rates were slightly higher for paper-and-pencil test (PPT) than for computer-based test (CBT), although the mode effect was not significant.

Way et al. [19], investigated the comparability of paper and online versions of the Texas statewide tests in mathematics, reading/English language arts, science and social studies at grades 8 and 11. The results of this study showed that the tests were more difficult for the online group than for the paper group.

Keng et al. [9] found that English language arts items that were longer in passage length and math items that required graphing and geometric manipulations or involved scrolling in the online administration tended to favor the paper group.

Over the years, the quality of tests that are done on the computer has changed, also the student experience in using computers. The study of Kingston [10] summarizes the results of eighty-one researches that have been done between 1997 and 2007. All these studies investigated the comparability of classical test and test done on computer. In his study, Kingston applied meta-analysis in order to demonstrate if the grade (elementary, middle or high schools) or subject in which knowledge is checked (English, mathematics, social sciences) have an impact on the comparability of computerized and traditional tests. Research has shown that the grade does not affect the comparability of tests, while in the case of the subject it was shown that the classical tests have a small advantage for math test, while a computerized test of knowledge has an advantage in testing English and social sciences.

The paper of Wang [18] described the research that was done in 2003 in the United States. The subject of study was Stanford Diagnostic Reading Test Fourth Edition (SDRT 4) and the Stanford Diagnostic Mathematics Test Fourth Edition (SDMT 4), each of which has six levels and which are adapted for taking on the computer. The participants were students from U.S. school from second to twelfth grade. In this study, 1863 students have done the test SDRT 4 and 1774 students the test SDMT 4. The results gave solid, unambiguous evidence of reliability and comparability of test results SDRT 4 and SDMT 4 for all grades and levels of the test, regardless of the manner of conducting the test. Differences in the achieved results based on the method of conducting the test do not exceed the expected random errors for most SDRT 4/SDMT 4 subtests.

The project PASS-IT (Project on Assessment in Scotland – using Information Technology) lasted for 27 months (starting from August 2002 until December 2004), its aim was to look into the possibility of formative and summative online knowledge assessment in secondary schools in Scotland [3]. One of the conclusions of the research is that technology must support the educational requirements of specific subjects and levels. For example, in order to reliably and validly determine the success of students in mathematics, the system must provide the possibility

of partial points. Furthermore, for certain subjects such as music, integration of multimedia elements is very important to support the issues in this area.

Today's technology has the ability to do more than just accelerate the process of testing. A growing number of experts involved in education agree that technology can improve teaching and learning. One of the projects that involves new forms of technology in solving problems in real life is the Problem Solving in Technology Rich Environments (TRE project) [17]. The project was started in 2003 in the United States and had a number of participants of 2000 students. TRE tested necessary scientific skills, such as the ability to find information about preset subject, to estimate which information is relevant for experiment, to make the plan and perform an experiment, and to organize and interpret results.

Thus, for example, eighth grade students in the experiment (which was entirely done on computer) had the task of using a balloon charged with helium to solve the problem of the growing complexity. They had to find the relation between power holding balloon at a height, mass and volume. Students were asked to determine the relationship between the mass which is placed in the basket of balloon and height it can reach. To solve this problem, students have gathered the necessary information performing the experiment several times with different masses, and when they had enough data to make conclusions, it was supposed to give the conclusions in the form of answers to multiple responses questions. The TRE project demonstrated several unique capabilities of knowledge assessment provided by technology [17]. First, the technology allows the presentation of much more complex problems to be solved in several steps. Different forms of multimedia, such as an animated helium balloon and an instrument panel that allows setting the parameters of the balloons, can represent the problem much better than if it were only explained in written form or orally.

Another example of technology in setting up and solving problems is the Floaters test which is offered to students in the UK as part of the World Class test [17]. This program allows checking students' knowledge in conditions without paper and pencil. For example, students use interactive simulation to measure the weight of various foods such as carrots, apples and bananas, and their task is to determine whether these pieces of fruit can float on the surface of the water. Students are then asked to set up a hypothesis based on the templates that were found.

Research

The main purpose of this study is based on theoretical research and the use of computer capabilities in the evaluation of knowledge in order to indicate statistically significant possibility of raising the overall level and quality of the teaching process. Some results about using computers for student assessments could be found in Maravic et al. [12] and Maravic et al. [13]. Besides this main purpose, the aim was to detect those characteristics of CBT, which may have a negative effect on students' achievements. The objective was to determine the influence of the way in which computer randomly generates questions (area and weight), i.e. an impact if first the most difficult question appears from a set of selected test questions and inability of browsing back and forth. Also, the intention was to find out if there is influence on students' results if immediately after given the answer the message "answer is correct" or "answer is incorrect" appears.

The main hypothesis of this research is that the results, given by the computer-based tests, are valid and reliable alternative to the classical way of knowledge testing on paper. Therefore, the goal is to find the answer to the question of whether there are differences in the achievements of students which outcome from different modalities of delivery of the test. The following null hypothesis was stated:

"There is no significant difference between the students' score in computer-based test, compared to those obtained with paper-and-pencil test."

In addition to this primary aim, one more objective was formulated: how and whether the way of question presentation (one question at a time on the screen, or more questions and need for scrolling) affects achieved better result. The following auxiliary hypothesis was stated:

"There are differences between the students' score which depend on how the computer-based assessment was built, how the question was presented and which are the answering techniques."

Experimental Results

The experiment was carried out with college students. The objectives were to evaluate students' results and opinion when they take tests on PCs, to see whether there are significant differences in the results obtained with paper-and-pencil tests and with computers, and also to search for differences between diversely assembled computer based assessments. In order to know what the students' opinions are and whether or not they were satisfied, a survey was carried out with specific questions and personal comments.

Participants

The participants were future engineers, i.e. students of computer science (engineering students) at Subotica Tech (Serbia), all about the same age (about 20 years old) and in a similar situation (first year of computer science study). Data was collected in the spring of 2010. The research included 90 students (selected from the Department of Informatics) who took the Object-oriented programming course as a compulsory subject. The students of computer science are predominantly male (which is generally true for Subotica Tech). This is reflected in the gender-percentage: 90% male test subjects and 10% female. The total number of college students at Subotica Tech is 591, of which only 57 are female, or 9.64%, so the sample can be considered representative. The students were divided into two groups, an experimental group with 45 students (computer-based test) and a control group with same number of students (paper-and-pencil test). Students were pre-tested to ensure that the groups are of equal knowledge. All students had previously been given instructions for the examination and related learning material.

Instrument

In order to investigate students' knowledge, a *multiple choice questionnaire* (*MCQ*) with twenty questions was developed. The paper and pencil and the computer based versions of the MCQ test included the same set of twenty questions. The time provided for solving the test was thirty minutes for both groups. Hand scoring was done for the paper-and-pencil version of the test, and automatic scoring by computer for the computer-based test. To make participants familiar with the CBT, they had an opportunity to exercise before the test.

Examination procedure and scoring methodology

All students, participants in the experiment did the same test. The test contained twenty questions. For each question there were several answers offered (usually four) of which only one was correct (i.e. it was an MCQ with one correct answer). Each correct answer carries one point. For incorrect responses there were no negative points given, and questions that remained without answers carried zero points. The negative marking was omitted based on the findings of Bliss [4], namely that negative marking tends to penalize the more able students. The decision to omit questions is influenced by personality characteristics [8]. According to [7], The Royal College of General Practitioners in the UK discontinued negative marking many years ago when they demonstrated that it discriminated female candidates because they tended to be more cautious with regard to guessing.

The maximum number of points that can be obtained on the test was twenty. During the preparation of the paper test, the order of test questions was not associated with their weight (i.e., the questions were not ranged from easier towards the more difficult, or vice versa), but they were randomly selected from a set of questions and compiled to make up the questionnaire. The order of questions for the CBT was left up to the computer to randomly arrange them. Before any of the students used the tests on the computer, all college computers underwent technical checks, to ensure that they had the correct software installed and to check that their display configurations were acceptable. Immediately prior to the test administration, students were asked to access a practice test and practice the question answer submission process.

Analysis

The results that the students achieved on the tests were subjected to statistical analysis. ANOVA (Analysis Of Variance between groups) analysis was applied to test the hypotheses. All statistical analyses reported in this research were conducted with a significant level of .01.

Results

The American Psychological Association (APA), in a document entitled Guidelines for Computer-Based Tests and Interpretations [2], gives specific recommendations for computerized test administrations and score interpretations. The guidelines state the "computerized administration normally should provide test takers with at least the same degree of feedback and editorial control regarding their responses that they would experience in traditional testing formats" [2]. This means that test participants should be able to review their responses to previous items as well as skip ahead to future items, and make any changes they wish along the way. To check the influence of ways of presenting issues two experiments were planned in this research.

In order to check whether or not the way in which questions are presented on the screen may influence achieved results (only one question per screen, or all questions provided for the test), we have carried out two experiments. In the first experiment, participants of the experimental group could only see one question on the computer monitor, there was no possibility of browsing back and forth if an answer to the question was not given, and immediately after submitting the answer the message "true" or "false" appeared on the screen. These factors were obviously available to students of control group who did the PP test.

The distribution of participants' scores in the PPT and CBT is presented in Table 1 and in Figure 1. The mean score was higher for the paper-and-pencil test (M=9.91, SD=5.22) than for the computer-based testing (M=8.84, SD=4.607) by 1.07 points. The goal of this research was to find out whether there are differences in the achievements of students due to the different modalities of the test delivery. The participants' results were not statistically different in the CBT and in the P&P test (F=1.056, p>0.01), as presented in Table 2, in the case when students could see only one question on the computer screen.

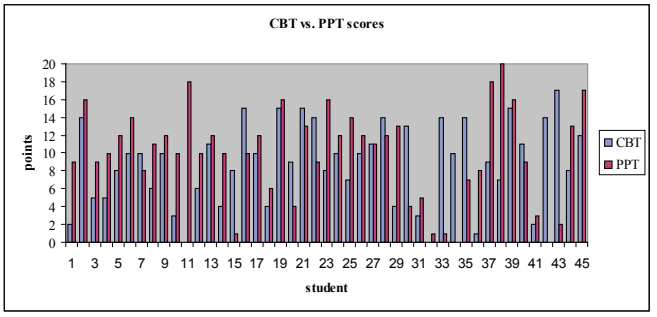


Figure 1. Distributions of students' scores in PPT and CBT, first experiment

Table 1. The distribution of students' scores in the PP test and in the CBT, first experiment

	N	Mean	Standard deviation	Standard error	95%Confidence level		Minimum	Maximum
					Lower bound	Upper bound		
PPT	45	9.91	5.222	0.778	9.13	10.69	0	20
CBT	45	8.84	4.607	0.687	8.15	9.53	0	17
Total	**90**	**9.38**	**4.925**	**0.519**	**8.86**	**9.89**	**0**	**20**

Table 2. One-way ANOVA comparison of scores of participants in the PP and CBT, first experiment

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	25.6	1	25.6	1.056	0.306971	6.932
Within Groups	2133.56	88	24.245			
Total	**2159.16**	**89**				

Table 3. The distribution of students' scores in the paper-and-pencil test and in the computer-based test, second experiment

	N	Mean	Standard deviation	Standard error	95%Confidence level		Minimum	Maximum
					Lower bound	Upper bound		
PPT	45	8.93	5.167	0.77	8.15	9.71	0	19
CBT	45	10	5.117	0.76	9.23	10.77	1	20
Total	**90**	**9.47**	**5.141**	**0.542**	**8.93**	**10.0**	**0**	**20**

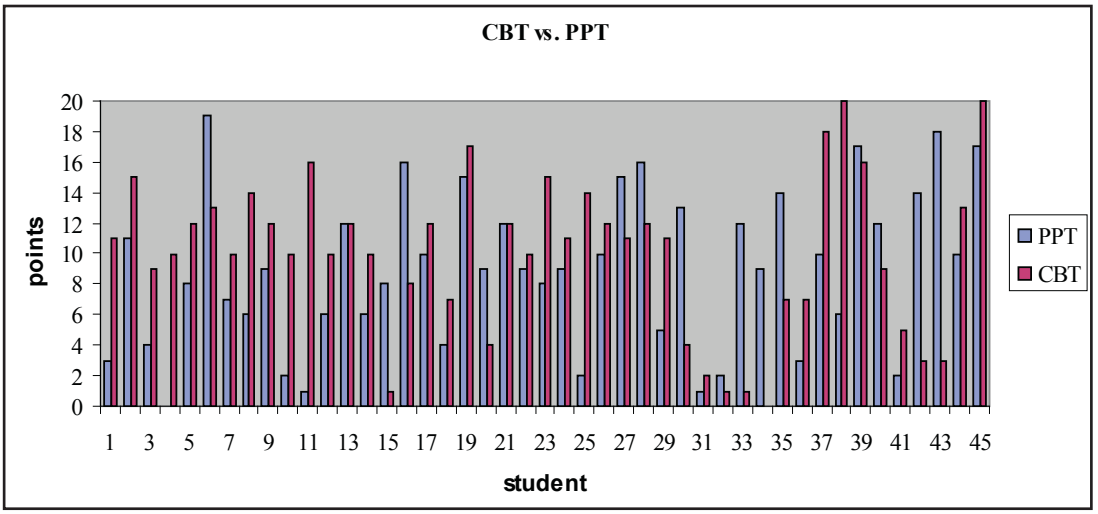


Figure 2. Distributions of students' scores in PPT and CBT, second experiment

To check whether there is a statistically significant difference in the results (which would be the result of the different display modes) the second experiment was conducted two months after the first one. In the second experiment, students who did the test on the computer could see all the questions included in the test at once. After submitting the answer the message "true" or "false" did not appear. The distribution of participants' scores in the PPT and CBT is presented in Table 3 and Figure 2. This time, the mean score and standard deviation for computer-based testing was M=10, SD=5.117, and for the paper-and-pencil test it was M=8.93, SD=5.167. The difference in mean value was the same as in the first experiment, i.e. 1.07 points, but this time students in the experimental group scored better. Data analyses found that there was no statistically significant difference in the results in the CBT and in the PPT (F=0.968, p>0.01), as presented in Table 4.

Based on the results of the first and second experiment we can conclude that there is no statistically significant influence on the students' results due to the way in which questions are presented on the computer screen. The null hypothesis *"There is no significant difference between the students' score in computer-based test, compared to those obtained with paper-and-pencil test"* is confirmed.

After the test, students who did the test on the computer filled out the questionnaire to see what their attitude towards this kind of knowledge testing was, and to find out the answer to the auxiliary hypothesis. The survey was anonymous in order to at-

Table 4. One-way ANOVA comparison of scores of participants in the PPT and CBT, second experiment

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	25.6	1	25.6	0.968197	0.327828	6.931941419
Within Groups	2326.8	88	26.44090909			
Total	**2352.4**	**89**				

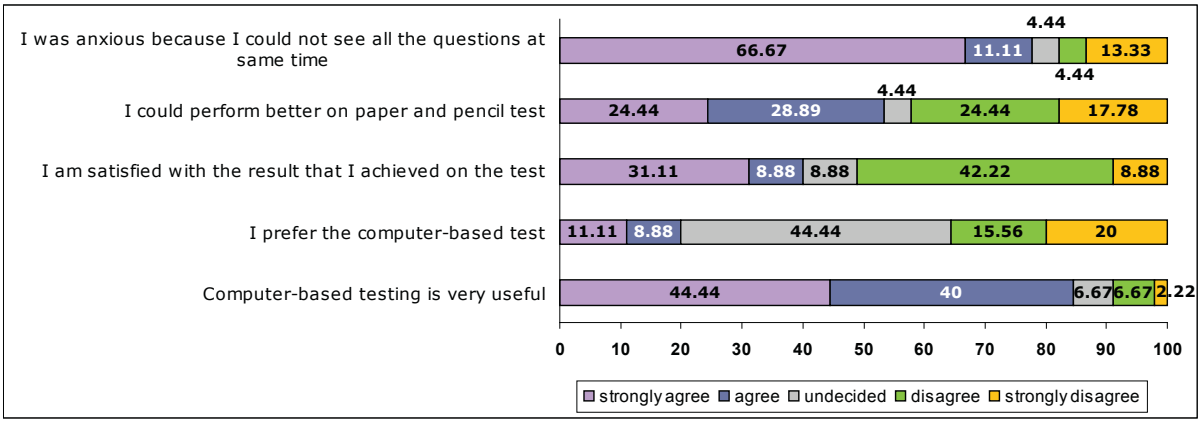


Figure 3. The results of the survey

tain honest answers from the participants. The results of the students' answers are given in Figure 3. The questionnaire was designed to collect information about students' attitudes towards aspects of testing. The survey had five statements: "Computer-based testing is very useful"; "I prefer the computer-based test"; "I am satisfied with the result that I achieved on the test"; "I could perform better on paper-and-pencil test"; "I was anxious because I could not see all the questions at same time". For the evaluation of student responses, the authors used a Likert-type scale with five responses: "strongly agree", "agree", "undecided", "disagree" and "strongly disagree" [5].

Students could also write their personal comments about this kind of testing. The main objection on the part of the students was that they could not see all the questions at once and so could not make a strategy for solving the test. This comment is visible also in their answer to the question "I was anxious because I could not see all the questions at same time", where 66.67% of the students strongly agree and 11.11% agree. Many of the students felt discouraged by the fact that the questions at the beginning of the test seemed too difficult for them, and they would opt for answering them randomly just to get to the next question in line. Later, they had no opportunity to review the test and maybe make an effort to answer the questions that remained unanswered. This attitude may explain the fact that 53.33% of students think that they could perform better on paper-and-pencil test (24.44% strongly agree and 28.89% agree).

As for the results to the statement "I prefer the computer-based test", 11.11% strongly agree, 8.88% agree and even 44.44% were undecided. Students emphasized that they prefer the classical method of solving the test because it gives insight into all the questions for the test. Also, one student "admitted" that he is trying to find a pattern, for example, that the correct answer to every question is under the number 3, and with computer test seeking for patterns was difficult. Despite all the negative comments that were given after the first experiment, students agree that computer-based testing is very useful (44.44% strongly agree and 40% agree). As for the benefits of computer testing, the majority of students pointed out that they liked the fact that after pressing the "submit" button they would find out the result of their achievements. The feedback information after each response about the answer's correctness ("correct" or "incorrect" answer) has a motivational role, but sometimes information that the given answer was the wrong one can negatively affect the further process of solving the test.

After the second experiment, when the students of the experimental group could see all the questions at once, they gave favorable comments. This time the questionnaire had only three statements: "I prefer the computer-based test"; "I am satisfied with the result that I achieved on the test"; "I could perform better on the paper-and-pencil test". The results are given in Figure 4. Students expressed satisfaction because the test now was "a copy of paper test only on the computer". As to the argument of "why" the comments were the following: "I type faster on the keyboard, than I am writing with the pen", "I am more used to use the keyboard than the pencil".

According to the results of the survey, it could be concluded that the following auxiliary hypothesis:

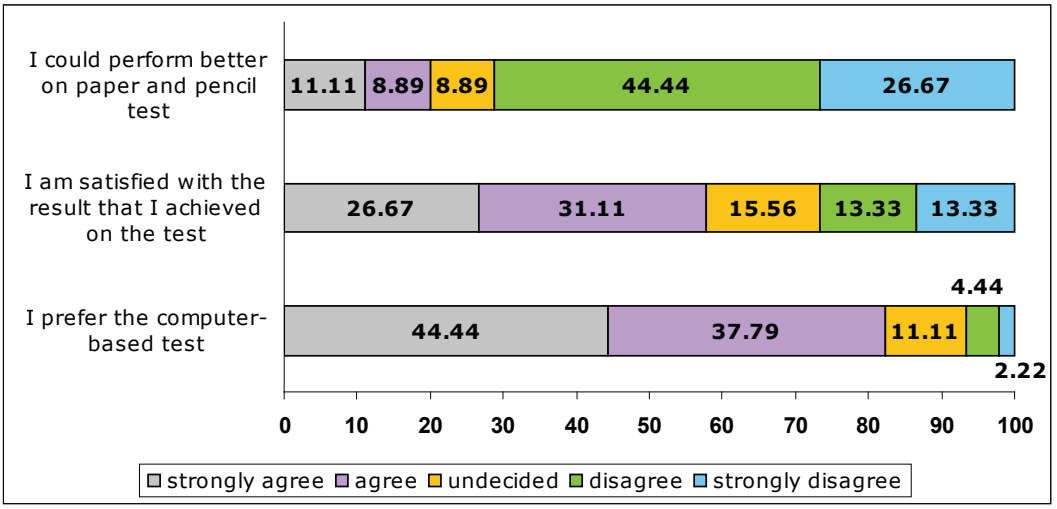


Figure 4. The results of the survey after the second experiment

"There are differences between the students' score which depends on how the computer based assessment was built, how the question was presented and which are the answering techniques." is confirmed.

Based on the presented results, it can be concluded that the advances in computer technology and investments in evaluation and testing software, together with the advantages of immediate feedback and automatic grading, make computer-based testing more and more common.

Conclusions

This study makes a contribution to the research on computer-based testing by examining the mode differences between the paper-and-pencil test and computer-based test. The previously conducted researches in this area dealt with the students of primary and secondary schools. In those researches the points of observation were the students' successes in mathematics, English and social sciences; no research was done in the field of programming languages such as C++ with post-secondary students. Also, the majority of studies were conducted with students in highly developed countries like USA and UK. There are only few studies, for example Akdemir and Oguz [1], which were conducted in a developing country such as Serbia.

The main aim of this study was to find out whether there are differences in the achieved results in two ways of testing: computer-based testing and paper-and-pencil test. Also, the intention was to detect those characteristics of CBT, which may have a negative effect on students' achievements. The objective was to determine the influence of the way in which computer random generates questions (area and weight), i.e. an effect if first the most difficult question appears from a set of selected test questions and the inability of browsing back and forth. The intention was also to find out if it will influence the students' results if immediately after giving the answer, the message "your answer is correct" or "your answer is incorrect" appears on the screen. The participants were a representative sample of the population of all engineering students studying computer science at Subotica Tech. The findings of this study led the authors to reach the conclusion that there are no significant differences in scored results for the PPT and CBT. Also, based on the survey results it can be concluded that the way in which questions are presented on the computer screen does have an effect on student satisfaction with CBT.

It is important to mention that the students were more satisfied with the computer-based test when they could see all questions at once (as in the second experiment). In his study Marks [14] observed that algorithms that randomize the order in which the test questions are presented to each candidate automatically control certain computer-based test assessments. If the test was such that in random sequences first the toughest question appeared, it may increase test anxiety for some candidates and influence their scores. Increased anxiety for whatever reason is likely to have a negative effect on that person's performance on the test.

The answer to this problem could be a computer-adaptive test (CAT), as a form of computer-assisted assessment where the level of difficulty of the questions administered to individual test-takers is dynamically tailored to their proficiency levels. Therefore, a logical continuation of this study is to examine the possibilities and advantages that CAT offers.

References:

[1] Akdemir O and Oguz A (2008) Computer based testing: An alternative for the assessment of Turkish undergraduate students. Computers&Education, 51(3), 1198-1204

[2] American Psychological Association Committee on Professional Standards and Committee on Psychological Tests and Assessment (1986) Guidelines for computer-based tests and interpretations. Washington, DC.

[3] Ashton H and Wood C (2006) Use of Online Assessment to Enhance Teaching and Learning: the PASS-IT Project. European Educational Research Journal, 5(2), 122-130

[4] Bliss L (1980) A test of Lord's assumption regarding examinee guessing behavior on multiple-choice tests using elementary school students. Journal of Educational Measurement, 17(2), 147–152

[5] Clason DL and Dormody T J(1994) Analyzing Data Measured by Individual Likert-Type Items. Journal of Agricultural Education, 35(4), 31-35

[6] Drasgow F (2002) The Work Ahead: A sychometric infrastructure for computerized adaptive tests. In C.N. Mills, M.T. Potenza, J.J. Fremer, & W.C. Ward (Eds.), Computer-based testing: Building the foundation for future assessments, Hillsdale, NJ: Lawrence Erlbaum, 67–88

[7] Goldik Z (2008) Abandoning negative marking, European Journal of Anaesthesiology, 25(5), 349-351

[8] Harden RM et al (1976) Multiple choice questions: to guess or not to guess. Medical Education, 10, 27–32. doi:10.1111/j.1365-2923.1976.tb00527.x

[9] Keng L et al. (2008) Item-level comparative analysis of online and paper administrations of the Texas Assessment of Knowledge and Skills, Applied Measurement in Education, 21 (3), 207-26

[10] Kingston N (2009) Comparability of Computer-and Paper-Administered Multiple-Choice Tests for K-12 Populations: A Synthesis, Applied Measurement in Education, 22(1), 22-37

[11] Lee JA and Hopkins L (1985) The effects of training on computerized aptitude test performance and anxiety. Paper presented at the annual meeting of the Eastern Psychological Association. Boston, MA, 1985. Available at http: http://www.eric.ed.gov/PDFS/ED263889.pdf

[12] Maravić Čisar S et al. (2009) True/false Questions Analysis Using Computerized Certainty-based Marking Tests, 7th International Symposium on Intelligent Systems and Informatics SISY 2009, Subotica, Serbia, September 25-26, 2009., Proceedings CD ROM, IEEE Catalog Number: CFP0984C-CDR, ISBN: 978-1-4244-5349-8, Library of Congress: 2009909575

[13] Maravić Čisar S et al. (2010) New Possibilities for Assessment through the Use of Computer Based Testing, 8th International Symposium on Intelligent Systems and Informatics SISY 2010, Subotica, Serbia, September 10-11, 2010., Proceedings CD ROM, IEEE Catalog Number: CFP1084C-CDR, ISBN: 978-1-4244-7395-3, 149-152

[14] Marks AM (2007) Random question sequencing in computer-based testing (CBT) assessments and its effect on individual student performance. Available at http://upetd.up.ac.za/thesis/available/etd-06042008-083644/unrestricted/dissertation.pdf.

[15] Nichols P and Kirkpatrick R (2005) Comparability of the computer administered tests with existing paper-and-pencil tests in reading and mathematics tests. Paper presented at the Annual Meeting of the American Educational Research Association, Montreal, Canada

[16] Shermis MD and Lombard D (1997) *Effects of computer-based test administrations on test anxiety and performance*, Computers in Human Behavior, 14(1), 111-123

[17] Tucker B (2009) Beyond the Bubble: Technology and the Future of Student Assessment, Education Sector Reports. Available at http://www.educationsector.org/usr_doc/Beyond_the_Bubble.pdf

[18] Wang S (2004) Online or Paper: Does Delivery Affect Results? Administration Mode Comparability Study for Stanford Diagnostic Reading and Mathematics Tests, Pearson Education, 2004. Available at http://www.pearsonassessments.com/NR/rdonlyres/D381C2EA-18A6-4B52-A5DC-DD0CEC3B0D40/0/OnlineorPaper.pdf

[19] Way WD et al. (2006) Score comparability of online and paper administrations of the Texas Assessment of Knowledge and Skills. Paper presented at the Annual Meeting of the National Council on Measurement in Education, San Francisco, CA, 2006. Available at http://www.pearsonedmeasurement.com/downloads/ conference/ScoreCompTAKS_cp0601.pdf

Submitted: April 15, 2012
Accepted: May 22, 2012

Relational Model and Missing Information

Siniša Jakovljević
Pension and Disability Insurance Fund, Republic of Srpska
jaksi@teol.net

Contribution to the State of the Art
DOI: **10.7251/JIT1201032J** UDC: **005.5:004.032.6**

Abstract: This paper examines possibilities offered by relational model when using missing information. The overview is conducted and possibilities which occur in practial use were analyzed. The use of predicates in which missing values occur has also been analyzed. Possible effects on system performance have been indicated.

Keywords: relational model, missing information, ***null***, three value logic (3VL), integrity, relational operators.

Introduction

More than 40 years after it was revealed, the relational model still exists in its full capacity on the database scene. Behind the amount of data which is stored for both current needs and long-term storage of important data, there are some systems for Relational Database Management System (RDMBS). Modern RDBMS are based on the implementation of the relational model proposed in the early 70s by E.F. Codd [1]. This document has significantly changed the world of database. Simplicity and understandability of the relational model has enabled it to be generally accepted. The close relationship of the model with the perception of the real world was a key factor of success and user's commitment to it, which by its use, model has achieved. The conception, according to which database users should be freed from the knowledge of the internal data presentation on computers, has opened a wide space for parallel development of technology for internal data storage and technology for access and use of this data. In addition to commercial solutions, many free-of-charge and open-source solutions are present on the market. Even leaders in the commercial segment of the market offer very powerful versions which are fully free-of-charge, while functionality is completely preserved. Differences in relation to the commercial version of the same manufacturer are primarily related to certain limitations in the size of the database, and unavailabilty of additional software tools, which the work with data make more improved and efficient.

Relational Model

The relational data model presents an abstract data theory. It is based on proven mathematical aspects, primarily on the set theory and first-order predicate logic [8]. The original model was consisted of three main components: structure, semantics and manipulation. The structure base is formed by relations. Usually, a graphic relation is presented in the form of a two-dimensial table. Basically, the relation is a mathematical determinant for the table of special type [4].

Relation attributes: Relations are defined by their attributes whose range of values is determined by the domain (type) of those attributes. The domain of a certain attribute represents a set of values which that particular attribute is able to accept. Presentation of the relation in the form of the two-dimensional table represents by itself a simplified approach. The level of relation is determined by the number of at-

tributes that relation contains. Suppose we have relation R which has two attributes, which are presented as two columns in the table. Then, it is a binary relation R. Generally, relation R with an n attribute is viewed as an n-ary, where n ≥ 1. Its presentation consists of the display of the table with n columns. Each presentation has heading and body. The relation heading consists of a set of attributes and their domains, while the relation body consists of a set of tuples which meet the heading structure. Each relation has at least one candidate key which by means of its value uniquely determines one and only one tuple of a given relation.

In the relational model, *Integrity* has a special significance. Integrity is achieved at the following levels: domain, entity and referential. Fulfilling integrity conditions at each of the mentioned levels is presented by the result of a certain logic operation whose result must be evaluated as true. At the domain level, integrity is fulfilled in a way that the value of a certain relation attribute must be within the range which is determined by the type of data for that attribute in particular. For example, if it is stipulated that the relation attribute takes values of a numeric type, then the transaction will be completed and data will be stored into the database if and only if that requirement is fulfilled. Also, if it is stipulated that a certain relation attribute is of an alphanumeric type with the length of n characters, then the transaction of data storage into the database will be successfully completed if and only if the attribute value is presented by alphanumeric characters, and the length is ≤ n.

At the entity level, integrity is fulfilled by defining rules in the form of primary key. Each entity could have more unique sets of attributes which can serve as a candidate key. In accordance with the requirements that are to be addressed, the information system designer can freely choose a set of attributes which will uniquely identify the tuple in that particular entity. Such a set of attributes is called the primary key and refers to the relation, rather than individually to a certain tuple . [7].

Referential integrity is established between the relations. If a certain value appears in one context, then it has to appear in another related context [5]. More formally: if the value of the attribute A of the relation R_1 references the value of the attribute B of the relation R_2, then the value of the attribute B of the relation R_2 must exist.

MISSING INFORMATION

Later, by expanding the relational model [2], the way in which the basic relational model relates to the missing information and how it treats that information is presented. What is missing?

As in the real world, it is possible to expect that at the moment of data collection, the value of a certain data will be unknown. It is very rational to plan the possibility that the database enables later data input, at the moment when it is available. Manipulative possibilities for missing information have predicted a three value logic. Truth is treated in the form of three possible values (3VL - three value logic). Truth values, formulated by Boolean algebra, true or false, are supplemented with a new truth value: unknown. This concept has been criticized immediately. It was suggested that the model solves the problem on the basis of previously accepted practice, old-fashioned, but proven method: instead of the missing value, a constant can be input, whose value according to the needs, is determined by a designer. For numeric data, 0 or 1 can be used taking into account the impact of a constant on arithmetic operations which can be performed, i.e. whether to add or multiply a constant will have an influence on the final result. Bearing in mind these types of cases and possible consequences, the old approach does not leave an impression of a secure solution. Therefore, in [3], the validity of approach by means of *unknown* value is argumented, combining that kind of approach with logic and algebra. The concept which is compatible with logic or algebra is taken over: by solving equations, we do not make use of variable values which at specific stages can be changed in time, but we solve equations by the use of logic.

Two essential questions are pointed out:

- Which type of information is missing?
- What is the reason for information missing?

Regarding the first question, it is possible that one part of a tuple is missing, e.g. date of birth of an employee. Also, it is possible that a complete set of data of an employee is missing. Model must deal with both situations. Regarding the second question, it is possible that at the moment of data collection a certain value is not available, but it will be added to the database immediately after becoming known, without any consequences. However, it is also possible that the data can not be related to the entity structure, meaning that the data in particular is not applicable in a given structure. In this case, we have a situation where a complete tuple is missing. Solution is achieved by introducing markers which will at that place hold an empty space for data storage, which meets the requirements of entity semantics. This marker is called the *null* or *null-value*. If the value is known, it will be stored, if not, a space for the storage will be reserved until the storage is possible. The space for the missing value is left, but the problem of manipulation of the missing value in operations that are performed on the data (arithmetic operations with numeric data, manipulation with character data, etc.) appears. Suppose that in the data about the employee we need to use the data about his/her total employment service. After one year of employment service has passed, total employment service must be increased by 1. This operation is trivial, but its result has a significant impact on the employee. What to do if the initial data about the employment service is missing? Situation arises: *null* + 1 = ? What is the solution to the equation? 1? Maybe! 15? Maybe! But maybe is not the answer to the question. The correct answer would be – total value is unknown. It is known that the result would be of a numeric type, but nothing else is known. The solution will be known once the previous value is known, the missing value. A similar situation occurs when the missing value appears in the alphanumeric values. What is the result in the following situation: *null* + 'BCD'? 'BCD'? Maybe! 'ABCD'? Maybe! But maybe is not the answer to this question either.

The missing values in the context of integrity maintenance are a specific problem. Making reference to an unknown value can not be performed until it becomes known. But then, it is not a missing value anymore. Introducing an unknown value into the primary key is opposite to the relational concept. By establishing the primary key over the relation, a firm rule of integrity is being established, thus, the use of an unknown value in the primary key is excluded. The situation is similar for the referential integrity. Referencing to something that is unknown means that there is a possibility that after determining the values of previously unknown, it can happen that the referenced value does not exist. This conflicts with the integrity rule.

Three Value Logic (3VL)

Three value logic introduces a lot of order and unifies the handling of unknown values.

Comparing scalar values of which at least one is an unknown value as a result gives an unknown value. The unknown value of the logical result represents the third value of the logic truth. In the following tables, the truth values for logic operations of conjuction, disjunction and negation are given.

AND	t	u	f
t	t	u	f
u	u	u	f
f	f	f	f

OR	t	u	f
t	t	t	t
u	t	u	f
f	t	u	f

NOT	
t	f
u	u
f	t

In the tables, usual English abbreviations are used: true, unknown, false.

To illustrate the usage of truth table, let us look at an example:

Let X=1, Y=2, Z is unknown.

X > Y	AND	Y > Z	=> false
X > Y	OR	Y > Z	=> unknown
X < Y	OR	Y > Z	=> true
NOT(X=Z)			=> unknown

Null and scalar operators

Let us consider the following example:

Let the value A be unknown.

A + 1	+ A	=> unknown
A – 1	– A	=> unknown
A * 1	* A	=> unknown
A / 1	/ A	=> unknown
+ A	- A	=> unknown

It can be concluded that if at least one operand of a numeric expression is unknown, the result is an unknown value.

If an unknown value is presented with a null, by analyzing the difference or division by zero, we get intuitively an unexpected result:

null - null	=> unknown
null / 0	=> unknown

In the first case, it is obvious that the two *null* values are not treated as equal ones. They are not values but placeholders for the values which will later appear. Therefore, the result of subtraction is unknown.

In the second case, although we would expect a message to try to divide by zero, regardless of when a new value is available, still, if a divisor is *null*, the result of division by zero is unknown.

Null and relational algebra

Relational algebra as part of the relational model has a very clear development of missing information usage in the relational operations. In order to consider the impact of the missing information on the relational operations, it is necessary to bear in mind the following:

- operands of the relational operations are relations
- the result of the relational operations are relations

Projection by definition eliminates tuples that are duplicates. Hence, if a relation contains more identical tuples, meaning that all correspondent attributes of one tuple have the same values as the attributes of another tuple, the result of the projection will display only one occurence of the tuple, regardeless of their number.

For example, let us imagine the relation R, which is a ternary one, meaning that it has 3 attributes: ID#, NAME, NUMBER, and that we have 4 ternary tuples of this relation in total:

ID#	NAME	NUMBER
4001	PETER	5000
4002	PAUL	4000
4003	MARY	3000
4002	PAUL	4000

Projection π(R) produces new relation which has 3 ternary tuples in total:

ID#	NAME	NUMBER
4001	PETER	5000
4002	PAUL	4000
4003	MARY	3000

Ternary tuple of the relation R whose value is ID=4002, NAME=PAUL, NUMBER=4000, in a resulting relation is summarized in one instance of that ternary tuple.

Let us now imagine that in the same example, for ID=4002 we have missing values:

ID#	NAME	NUMBER
4001	PETER	5000
4002	PAUL	NULL
4003	MARY	3000
4002	PAUL	NULL

Applying logic where *null* is equal to nothing, it follows that NULL ≠ NULL, it would be expected that the projection π(R) in this particular case will produce the resulting relation which will contain 4 ternary tuples:

ID#	NAME	NUMBER
4001	PETER	5000
4002	PAUL	NULL
4003	MARY	3000
4002	PAUL	NULL

However, it is determinated that even those tuples that in correspondent attributes have equal values are considered duplicates or that the pairs are established, to whose values, in database, missing values are added. Also, a certain criticism is expressed in terms of removing duplicated tuples because this expanded definition of duplicates does not meet semantic conditions – *null* ≠ *null* [3].

Missing information has no influence on the relational product. [3] [4]

Restriction operation is subjected to the influence of the missing information. The resulting relation contains only those tuples for which the condition of restriction meets the condition of truth, and tuples whose truth is false or unknown are discarded. T h e relational union operation – *union*, eliminates duplicated tuples in the same manner as explained for the projection.

Expansion of the union into the operation *union all* implies that removing of duplicated tuples is not being performed. Relational difference R_1 MINUS R_2 does not include removing of duplicated tuples. Tuple N_x can appear as a result of relational difference R_1 minus R_2 only if N_x is a duplicate of a certain tuple in relation R_1, and is not a duplicate of any tuple in relation R_2.

Intersect between two relations R_1 and R_2, will be the tuple N_x if and only if N_x is a duplicate of a certain tuple and in relation R_1 and R_2.

Join relations which for the requirement of join include attributes whose value allows the existence of an unknown value, will not execute the join. However, if there is a reason for an outer join of two relations, outer relation will contain missing values for tuples which do not meet the requirement of the join, and a designer will use that circumstance in accordance with the solution of a practical problem.

The core of this practical issue is reflected in the two and three value logic. According to their different nature, it is clear that they produce significantly different results.

Null and keys

Implementations of the relational model in commercial systems for database management solve the use of *null* values in a suitable, we could even say a pleasant manner. At the level of physical and logical database design, a designer is left with a choice whether he/she will at the level of domain integrity allow *null* as a possibility or not. Realization of this rule is simple and achieved by a trivial ban of the tuple which contains *null* value of a concrete attribute.

When it comes to the primary key, it by its nature and purpose should perform a unique identification of one tuple, and by doing so, the possibility of *null* value appearing as a part of the primary key is excluded.

Simply put, strictly defined and binding.

However, it is emphasized that each entity can have alternatives for the primary key, i.e. candidate keys. This suggests that there may be different combinations which can provide unity of a certain tuple. It was pointed out that in reality, sometimes at the moment of data storage, the value of each attribute can not be obtained each time. Therefore, such a combination of attributes excludes the possibility of choice for the primary key. However, once the missing value is obtained, and which will definitely ensure unity, such a formed tuple can have its value and can ensure full unity of the tuple. Still, the primary key must be a restrictive one, previously formed and known, so that the option of subsequently fulfilling the requirements of unity will not be an option. The possibility that a certain unity value is required,

starting from the moment when the missing value is available, seems rational. ANSI SQL-92 standard prescribes that the two values are not distinct if they are *null* values and if they fulfill the requirements of the standard prescribed in the clause 8.2. of the standard ANSI SQL-92. From the moment they cease to be *null* values, there is a real chance that they can become distinct.

Based on the above mentioned, there is still a possibility to technically ensure unity, and a definite checking could be performed when a missing value becomes known. At the moment of value recognition, checking is performed whether it is a distinct one and a unique one in a set of all values of that attribute. If this condition is fulfilled, this particular value can be stored into database. For this checking, integrity checking using a *unique* key is introduced.

Implementation of this part of the ANSI SQL-92 standard varies from manufacturer to manufacturer of Database Management System.. Oracle enables establishment of a *unique* key over a certain attribute in a manner that it allows the storage of *null* value by n times. Microsoft SQL server also enables establishment of a *unique* key, but in a more restrictive manner – in one relation, only one tuple which can in a certain attribute accept only one *null* value is enabled. When that specific tuple receives its missing value of that attribute, only then some other tuple can accept *null* value in that same attribute. This example of different limitation usage points out to a large specificity of the missing values and contradictions with which the world of missing values and the relational model is filled.

Null and performance

Satisfaction of a final user of a certain applicative solution that relies on database is reflected in the speed of response at which the user, from the database gets an answer to his/her query. Modern optimizers, implemented in RDBMS to select a set of results from all data, use techniques which rely on very complex mathematical models.

In [6], it is demonstrated that the presence of *null* values does not affect the number of distinct values which are generated in the form of system and object statistics. *Null*s are ignored while creating statistics. However, if the percentage of *null*s' participation is significant, a danger may appear in the form that the optimizer incorrectly interprets data, which would result in incorrect instructions for query execution which arise as a result of optimizer's action. Execution plan of the SQL query, which is based on incorrect optimizer's assumptions, inevitably leads to the performance degradation and a longer period of query execution. In relatively static data, initial optimizer's assumptions will give solid and stable results, however, unless missing values are regularly updated, the appearance of real data instead of *null*s will affect optimizer's work.

The issue of comparing *null* with an existing value is already mentioned. The only checking related to the *null* and which can be described as a trivial is the checking whether a certain attribute has *null* instead of value. Everything except that kind of checking is related to the potentially serious problems. The manner in which the *null* in the predicates of the SQL statement should be treated is an obligation which has to be considered by a designer of the applicative solution. Practical question would be: should *null* be treated as to fulfill the condition of truth or not? The answer depends on business logic on which a designer must create solution. If it is necessary to answer the question how many employees have an income between the values of I_1 and I_2, then a serious confusion may be caused by the result which will not consider the presence of the *null*s in the income data. Does the *null* meet the condition of comparison in this particular example is a question that should be answered by business logic, used by a designer. In the SQL query itself, *null* is processed using functions which treat *null*s in a sense that if the value is unknown, some previously adopted constant is applied. Possible solution is that even *null* does not meet the requirement of comparison. In the first as well as in the second case, it is necessary to further process missing value. Further danger presents the possibility that the total result of comparison ends up being unknown. It seems practical that the *null* should be mapped in some appropriate value. This additional process takes up part of the time which is needed for the result to be produced, but still presents serious

danger which threatens from multiple angles.

As said earlier, system statistics ignore tuples which contain *null*. If the attributes with the *null*s appear in index which is B*tree organized, this implies that

table1.number_of_rows > index1(table1).number_of_rows

which can, again, produce wrong conclusions obtained by the optimizer.

Conclusion

The importance of the relational model lies in the fact that the real world is very practically mapped into a technical form. Even in life, we often have situations where something is missing, out of objective and subjective reasons. Databases are necessities of reality, and therefore, the world of missing values is a completely natural phenomenon. The manner in which the relational model has supported this part of reality is very practical and detailed. It represents a powerful framework in which a designer must adjust a way in which he/she will manage something that is at a given moment unknown. The influence of the missing can be small, insignificant, but at the same time very big. The need exists, solutions also. Maybe missing information will be banned and maybe not. The world of missing information is condemned to a lot of answers in the form of – may be.

Reference

[1] Codd, E. (1970). A Relational Model of Data for Large Shared Data Banks. *Communications of the ACM , 13* (6), 377-387.

[2] Codd, E. (1979). Extending the Database Relational Model to Capture More Meaning. *ACM Transactions on Database Systems , 4* (4).

[3] Codd, E. F. (1990). *The Relational Model for DataBase Management Version 2.* Reading, Massachusetts: Addison-Wesley.

[4] Date, C. (2004). *An Introduction to Database Systems* (8th ed.). Boston: Pearson / Addison Wesley.

[5] Garcia-Molina, H., Ullman, J. D., & Widom, J. (2009). *Database Systems - The complete book* (2nd ed.). New Jersey: Pearson Prentice Hall.

[6] Lewis, J. (2006). *"Cost-Based Oracle Fundamentals".* New York: Apress.

[7] Silberschatz, A., Korth, H. F., & Sudarshan, S. (2011). *Database System Concepts, 6th ed.* New York: McGraw-Hill.

[8] Wolfram Research, Inc. (1999-2012). Retrieved from Wolfram MathWorld: http://mathworld.wolfram.com/First-Order-Logic.html

Submitted: May 08, 2012
Accepted: May 30, 2012

History-Enriched Digital Objects as a Factor of Improvement of Adaptive Educational Web Site Navigation

Željko Eremić
Technical College of Applied Sciences, Zrenjanin, Serbia
zeljko.eremic@gmail.com

Case study
DOI: **10.7251/JIT1201039E** UDC: **004.032.6**

Abstract: Modern educational websites offer a wealth of information and content intended for both students and teachers. Such facilities often are not grouped in a single location. While students are in need of fast and efficient access to certain content, teachers are in need for an insight into the learning process of students. By using capabilities of Ajax, it is possible to implement a system for mutual support, where teachers and students who have knowledge of the desired resource would share it with students who are in need of such information in real time. History-enriched digital objects can be used to store information about knowledge sharing. In combination with the records of user's behavior from the log files, this shared knowledge can make a significant contribution to the successful design and navigation of adaptive web sites. Adaptive web sites can change their content and presentation based on the previously recorded user's behavior.

Key words: History-enriched digital objects, Web design, Internet technology, Data mining.

Introduction

The educational process is largely based on the opportunities offered by information technology. Teaching content and information about lectures are given on the web sites of educational institutions, and their amount increases daily. The increase in the volume of content is beginning to arise as a problem when there is a need of separation of the useful content.

"One of the tendencies in education is the continually growing amount of learning content which must be acquired by the student. Almost every generation's curricula are extended by a certain amount of new, updated or revised material. With this infestation of learning content, another issue arises, namely that the time which is intended for learning this amount of content is growing ever shorter for each subsequent generation."[8]

„Systems such as educational experiences change in terms of modernization and globalization. Students are required to improve the style of self - teaching and their skills."[1]

Educational web sites are faced with frequent changes, which have to be met. In this sense there is a clear similarity with the situation in the business. "Contemporary business millieu is faced up with frequent changes. The market has its own challenges, and companies are making efforts to meet them in the best possible way."[9]

History-enriched digital objects can be used to keep the history of user's interactions with Web site. Such objects could be achieved by using Ajax technology. In combination with the log files, they could significantly contribute to the improvement of adaptive navigation of web sites. Students would be able to rely on knowledge of their colleagues and teachers in the process of finding useful content in real time, while once recorded knowledge could be used in future processes of student's learning.

Adaptive Websites and Log Files

Adaptive web sites are kinds of sites which form a model of user behavior based on the previously recorded user behaviors, and then based on these models they change their structure and presentation. Some of the first papers dealing with the theme of adaptive web sites are [7] and [6].This idea is being further developed in many other papers.

Log files are being used as a source of information. These files contain information about requests of the website user in a format that is standardized and suitable for computer processing. These files are automatically created and filled by the web server. Log files are discussed in [5].

History-Enriched Digital Objects

The main disadvantage of log file information is that relatively small amount of information is recorded about the user requests. There is justification for the usage of history-enriched digital objects in situations where information about user behaviors that are not recorded in the log files have to be collected.

"The notion of history-enriched digital objects is similar to physical wear. Usage leaves wear. Physical wear is an emergent property and though it generally remains unremarked upon until it causes a problem, it is also tattooed directly on the worn object, appearing exactly where it can make an informative difference."[4]

In this paper, the usage of such objects is intended for keeping the record of knowledge sharing among users. When a user gives a suggestion to another user, details of knowledge sharing are recorded in a database and used for improvement of the adaptive website navigation in future.

Ajax

Ajax combines several technologies that work together (XHTML, CSS, DOM, XML, XSLT, XML, HttpRequest). Ajax engine is located between the user and the server and it is designed for rendering pages for users and communicating with the server on behalf of users.

"The Ajax engine allows the user's interaction with the application to happen asynchronously - independent of communication with the server. So the user never stares at a blank browser window and an hourglass icon, waiting around for the server to do something."[3]. Comparison of the traditional model of web applications and Ajax model is given in Figure 1.

Ajax represents the basis for interpersonal support systems which are referred in this paper. The activities of this system are recorded and used for improvement of adaptive website navigation in combination with data from log files. Ajax allows periodic refresh of the section with the proposed shortcuts, while the rest remains seemingly unchanged, and as such it is suitable for study by the user (student or teacher).

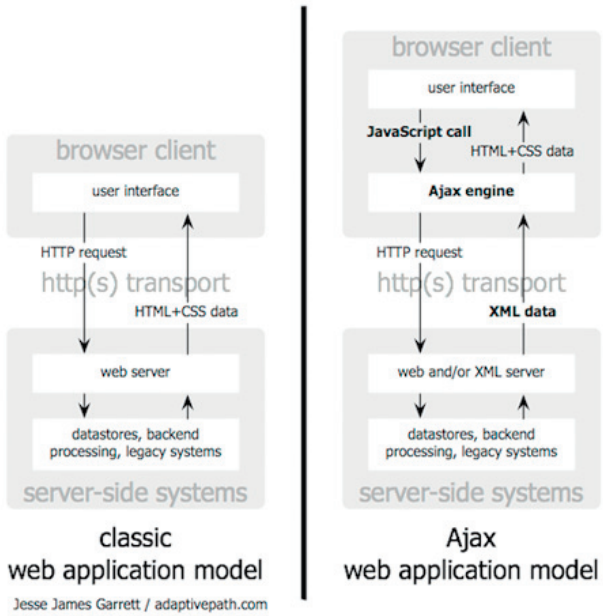


Figure 1: "The Traditional Model For Web Applications (Left) Compared To The Ajax Model (Right)." [3]

Interpersonal Support

In the article [2], possibilities of interpersonal support in real time are discussed. Block with suggested links is located on the web site pages and it has been implemented by using Ajax. There is a button which enables users to request assistance from other users regarding the selection of some of the proposed

shortcuts. The procedure is described in the example illustrated in Figures 2, 3 and 4. User Tom, clicking on the button, requests help in selecting some of the proposed connections, and user Tim gives him a suggestion about connection that leads to the document with the important content..

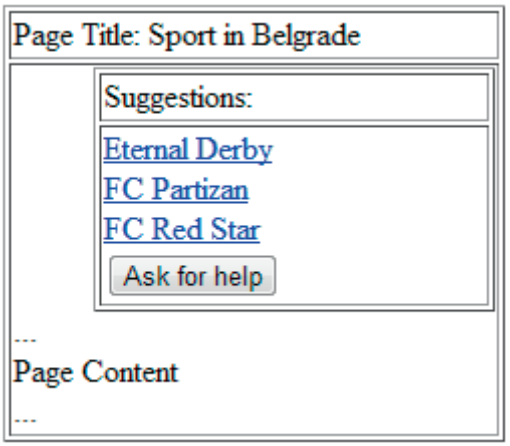


Figure 2: Page Entitled "Sport In Belgrade", Which Was Visited by the User Called "Tom" Contains Three Links to Recommended Web Pages

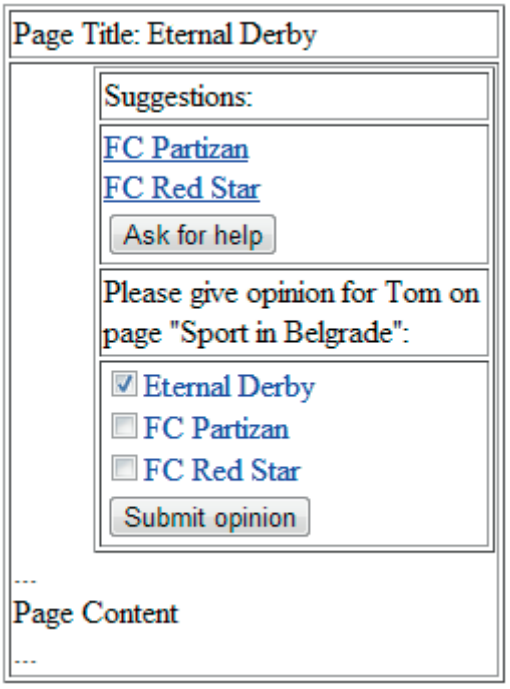
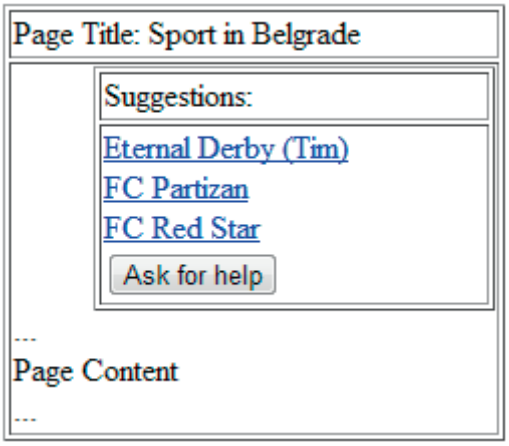


Figure 3: Page Entitled "Eternal Derby", Which Was Visited By "Tim" Contain Two Recommended Links to Web Pages, But it Also Contains the Request For the Suggestion to "Tom"

Figure 4: Page Entitled "Sport In Belgrade", Which Was



Visited by the User "Tom", Contains Three Recommended Links to Web Pages, With the Stated Recommendation Made by the Provider of Assistance Identified as "Tim"

All user requests for suggestions and responses to those requests can be recorded in the database. This data can be archived in a certain period of time and then used to improve adaptive web sites. It would be very interesting to compare the observed relationships between documents by using the model of previous user behaviors (the source would be log files) versus relationships between documents identified by using data from the system proposed in this paper.

History-Enriched Digital Objects and Improvement of Navigation

Interpersonal support, which is illustrated in the previous section, allows users to share knowledge almost instantaneously. Depending on the speed of response and refresh rate of Ajax section, it is possible to get answers within the time limit which is measured in seconds. This knowledge brings immediate benefits to the applicant.

Another possibility that arises is the archiving of these interactions and the subsequent use of data mining techniques in order to obtain useful information which could be used to improve the navigation of the (adaptive) web site. Conceptual model is presented in Figure 5 and physical model designed for the archiving of interactions which are the result of interpersonal cooperation activities, is presented in Figure 6.

The table "User" is intended to keep a record of all known users of the system identified by their

user names and IP addresses in case when the user is logged onto the system. If the IP address is the same (access from the same computer, for example) it is possible to distinguish the users according to their user names. Table "Session" keeps a record of any session where a request for assistance or respond to a request for assistance came from. Users can be unambiguously determined on the basis of the session, since each session has exactly one user. During the session, user can send a request for help while being on a document (web page), which is recorded in the table "Request". A document from where the request originates, a session from which the request is sent and time when the request is sent to the table "Request" are being recorded for each request. For each request there could be one, none or more responses from one or more user sessions (each session has its own user). The response to the request is identified by id values of the request which has been responded to, id document which is indicated by the response and the id of the session where the response came from. Also the time, when each of the responses is received, is being recorded in the table "Answer". Id value, the path and the name if there is any, are determined for each document (web page). These values are being recorded in the table "Document".

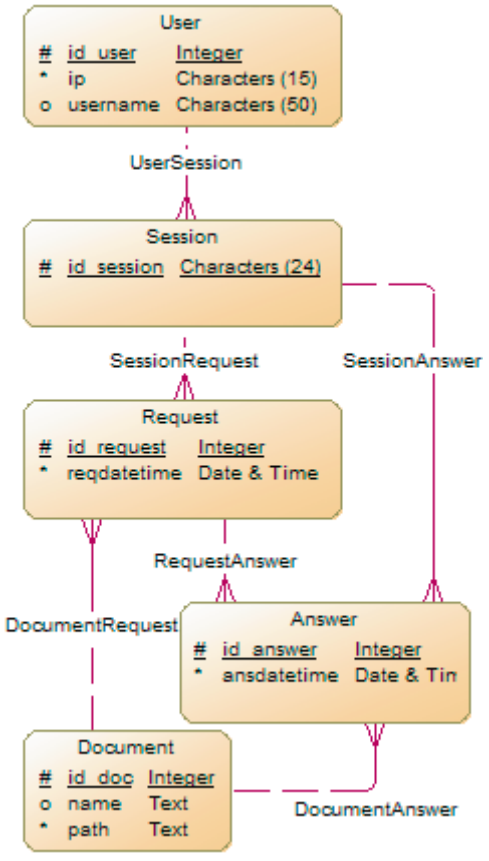


Figure 5: Conceptual Data Model

These tables can form a small database that could be placed on a web server. This database could record any interactions between users in a certain period of time. Data that would be kept in such a database might not be extensive as that from the log files, but would more directly reflect the needs of users.

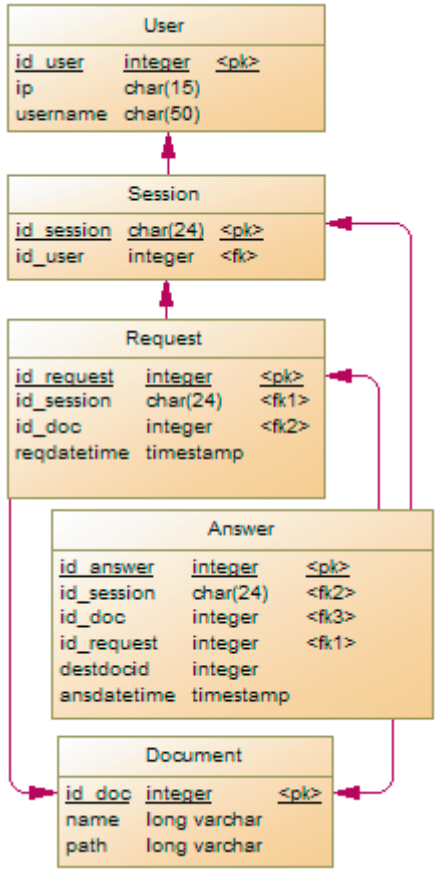


Figure 6: Physical Data Model

Such recorded data is ready for later application of Data mining techniques with the aim of extracting the useful knowledge. One of the applications is the determining of promising shortcuts based not only on the previously recorded user behaviors in the log files but also on interpersonal support. In the example shown in [2], the suggestion which user Tim gave could be used in the future if some user had the same request from the same page.

Another application could be found in the validation of existing system for suggesting shortcuts. Specifically in Figure 2, Tom gets a list of shortcuts offered by the system. This list of shortcuts is obtained from the previous models of user behaviors. Such a list could be, for example, received from large amounts of data recorded in log files. Although the data from the log files is very extensive and easily generated, it is often limited to a small number of attributes. On the other hand, the usage of History-enriched digital objects often requires specialized solutions such as solutions offered in this article. Such solutions require some effort but can offer additional benefits. In the example in [2], the shortcuts can be ranked on the basis of the frequency of the registered user paths. On the other hand, when using History-enriched digital objects it is possible to take into account the suggestions of users recorded in the system proposed in this paper.

Conclusion

The approach proposed in this paper should provide the necessary prerequisites for the archiving of user knowledge, which are related to the access to desired resources. Interpersonal support that has been proposed in [2] allows users to help each other when choosing a link to the proposed documents. In this way, users who seek help can get it very quickly, while still not having decided about where they could continue their session. Archiving of these suggestions can help future users. Small information system that is proposed in this paper allows such archiving and it is directly related to the solutions given in the article [2]. Data, which is expected as a result, is ready for applying data mining techniques in order to improve the navigation structure of adaptive web sites.

References

[1] Brtka, E., Brtka, V., and Radosav, D., (2012). The Clustering Data Mining Module as a Part of the E-Learning System. METALURGIA INTERNATIONAL. No. 17, Vol.4: 220-223.

[2] Eremić, Ž., and Radosav, D., (2012). Collaborative user support as a contribution to navigation improvement of adaptive websites for distance learning. International Conference on Information Technology and Development of Education – ITRO 2012, Zrenjanin.

[3] Garrett, J. J., (2005). Ajax: A new approach to web applications. Available At: http://www.adaptivepath.com/ideas/ajax-newapproach-web-applications.

[4] Hill, W., and Hollan, J., (1993). History-Enriched Digital Objects. Proceedings of Computers, Freedom: and Privacy (CFP'93).

[5] Pamnani, R., and Chawan, P., (2010). Web Usage Mining: A Research Area In Web Mining. International Conference on Recent Trends in Computer Engineering ISCET 2010, Punjab, India.

[6] Perkowitz, M., and Etzioni, O., (1997). Adaptive Sites: Automatically Learning from User Access Patterns. In Proc. 6th Int. World Wide Web Conf, Santa Clara, California.

[7] Perkowitz, M., and Etzioni, O., (1997). Adaptive Web Sites: an AI Challenge. Fifteenth International Joint Conference on Artificial Intelligence, Nagoya, Japan:16-21.

[8] Pinter, R., Radosav, D., and Maravic Cisar, S., (2012). Analyzing the impact of using interactive animations in teaching. International Journal of Computers, Communications & Control, 7 , 1: 157-172.

[9] Radosav, D., and Eremić, Ž., (2010). Relevant factors for promotion on-line shopping in Serbia analysis. TECHNICS TECHNOLOGIES EDUCATION MANAGEMENT - TTEM, No. 2, Vol. 5: 367-377.

[10] Željko Eremić (1981) is graduate engineer in IT area. He graduated at Technical Faculty "Mihajlo Pupin" in Zrenjanin in 2010 and got his title of master of science in Informatics. He is currently teaching at Technical College of Applied Sciences in Zrenjanin. His interests include developing software, e-business and e-marketing.

Submitted: April 05, 2012
Accepted: May 21, 2012

E-Commerce in DinaCard System

Dalibor Vučić[1], Saša Salapura[2]

[1]Bank of Moscow Belgrade, vucicd@msn.com

[2]Hypo-Alpe-Adria Bank Banja Luka, sasa.salapura@gmail.com

Case study

DOI: **10.7251/JIT1201044V** UDC: **004.738.5:65.012.2**

Abstract: *This paper presents the status of e-commerce in Serbia with the focus on the domestic DinaCard system, its architecture and participants in the system. We reported results on Internet transaction in DinaCard system in 2009, 2010 and 2011. We found that the number of all participants, including banks with the license for acquiring, banks with the license for issuing and Internet merchants was extremely low (up to 5) and showed no significant positive trend. As a consequence, the number of transactions with the DinaCard cards was also unacceptably low. Based on these results, we concluded that the DinaCard system for Internet transactions have a great potential, but all the participants have to make an effort to significantly increase the use of the domestic card in e-commerce.*

Key words: *e-commerce, DinaCard, Internet payments, e-business*

Introduction

In recent years of constant and rapid changes, especially in the field of Internet technology, companies need to incorporate e-business in their business, to keep or even improve their position on the market. In order to achieve this aim, companies should quickly adapt to a changing environment [6].

Although e-business has markedly been involved in many parts of business in developed countries, situation in Serbia is different. Electronic business has recently started to develop in our country and no literature data on the status of e-commerce on Serbian market has been published so far. In order to promote and support the use of e-business, the National Bank of Serbia made a great effort to introduce Internet transaction through the domestic DinaCard payment system in our country. Thus the aim of this paper was to systematize information from relevant literature on the e-business and payment models in electronic commerce, and to present the situation in Serbia related to e-commerce using the DinaCard payment system.

Literature Review

The literature which describes and defines the concepts of electronic business and electronic commerce, their similarities and differences is extensive, but on the other hand, the data related to this issue in Serbia are very sparse.

The meaning and limitations of both e-commerce and e-business are still a matter of the debate among consultants and academics worldwide. Laudon and Traver [8] defined e-commerce as the use of the Internet and Web for business (business transactions), or digitally enabled commercial transactions among organizations and/or individuals. The most frequent forms of e-commerce are: 1. transactional forms, which provide online sales, 2. service-oriented forms which encourage shopping and make closer relations between buyers and sellers, 3. brand building which promote and develop brands and 4. portals which provide various types of information [3].

Electronic business is a transformation of a business based on a connection of the company, customers and partners in the form of association process

(integration), process of cooperation (collaboration), and a global network connection (aggregation) using the Internet as a medium [16]. There have been many polemics in the literature on the difference between e-commerce and e-business, but the following three states are mostly cited: Electronic commerce overlaps with the electronic business in some levels [1], electronic commerce is equivalent with electronic business in scope [15] and electronic commerce is a subset of electronic business [4].

E-business and e-commerce are completely dependent on technology, from the hardware to the application layer. There are 4 layers in the infrastructure needed for e-business and e-commerce: network technology substrate (telecommunication networks and protocols), transmission (transportation services and representation standards), middleware (the connection between the transmission and application layer, including services related to security and authentication) and application (client application) [8].

The appearance of electronic commerce has also brought new financial needs, which often could not be met by the traditional payment system. For example, in electronic commerce it was necessary to find a solution for payment peer-to-peer, and for so-called micro-payments, which could not have been paid by existing credit card and classical payment systems. Based on these needs, the following payment systems in e-commerce have been extracted: Internet transactions with credit card, digital wallet, online stored value, digital accumulating balance, digital checks and contactless payment systems [5].

System of Internet transactions with credit cards is one of the dominant forms of Internet payments. The process of Internet transaction with credit card is almost identical to process of classic transactions with credit card directly used at merchant's place. The main differences between Internet transactions and standard transactions with credit card are physical absence of the both cardholder and the card (CNP - cardholder not present), as well as the absence of a signature, i.e. physical authentication of the client.

Privacy and security are very important parts for the clients, and it has been proven empirically that privacy (protection of personal data) and security (protection of users from fraud and financial loss) have strong impact on trust of the online financial services [9].

DinaCard System

To support the development of card business in Serbia, the National Bank of Serbia implemented a national payment card project in 2003. The new domestic card has been called DinaCard. The National Bank of Serbia provided the conditions for full implementation (technical, regulatory, organizational) of the national system for payment cards, through the newly established organizational unit (National Center for Payment Card, NCPC). All the terms and conditions for involving banks and other participants in DinaCard system have been precisely defined by the release of the DinaCard Operating rules document and DinaCard Technical documentation. In this way, banks in Serbia were able to introduce payment card business significantly faster, simpler and cheaper, and the other participants could have been involved in this system easily. Thereby card business in Serbia made a great progress and end users got used to payment cards and their use through a network of ATM and POS terminals. NCPC has constantly been working to introduce new services in DinaCard system, according to the trends of international card systems, and users' needs. Basic Card Services as payments and cash withdrawals were extended with additional services, including m-commerce and e-commerce services by introducing the Internet payments in DinaCard system.

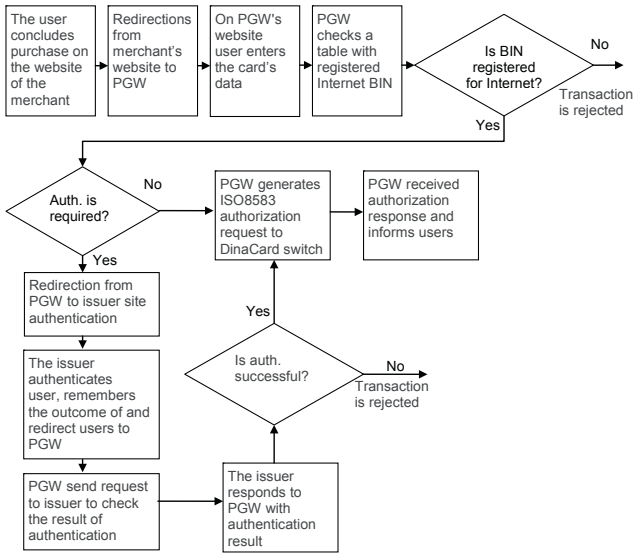
The DinaCard system for Internet payments was developed based on existing solutions in other international card systems, but taking into account specific needs and requirements of the local market. One of the specificity of the implemented system was a bigger involvement of the Payment gateway provider (PGW) in the process of internet transactions, unlike the other card systems, in which the biggest role had merchants and card issuers [7].

Participants in DinaCard Internet payment system, in addition to bank card issuers which should

enable payment via the Internet with their cards, are the bank acquirers for Internet transactions, Internet merchants, the bank issuer's processors, the bank acquirer's processors, PGW, the NCPC's central switch and customers or users of the services [10].

The payment flow on the Internet in DinaCard system is shown in Figure 1. As it can be seen on Figure 1, PGW actively participates in conversion of started transaction, but also in user authentication, which is usually done by the Internet merchants. The aim of this bigger involvement of PGW is to minimize the participation of Internet merchants in the technical part and to facilitate their inclusion in the Internet payment system.

Figure 1 Payment flow for DinaCard Internet transactions



Methodology

This study was designed as an interpretive case study. The main reason for this design was the lack of the relevant literature for Serbia. Our guide for investigation was specific research questions during the interviewing of the participants in the DinaCard Internet payment system, document analysis, and web site analysis for collecting data. We collected the data on the status of Internet payments in DinaCard system for 2009-2011, and compared the data during this period.

Results

The number of banks which have been licensed for Internet issuing and Internet acquiring, the number of the Internet merchants and the number of PGW and processors in the DinaCard Internet payment system are presented in Table 1. As it can be seen from Table 1, the number of all these participants in the system is very small and doesn't show any significant positive trend.

Table 2 presents the data on the number of successful and rejected Internet transactions in DinaCard system during the period from 2009 to the end of 2011. The number of both successful and rejected transactions was very low and in case of successful transactions even decreasing during this period.

Table 2 Number of Internet transaction in DinaCard system

Year	No. of successful Internet transactions	No. of rejected Internet transactions
2009.	79	59
2010.	50	109
2011.	12	110

In order to compare these results with the Internet transactions in Serbia made with other card's brands and with the total number of the DinaCard transactions in the same period, the data related to these types of transactions are shown in Table 3.

Table 1 The number of participants involved in the DinaCard Internet payment system

Year	No. of the Banks with licence for Internet acquiring	No. of the Banks with licence for Internet issuing	No. of Internet merchants	No. of licensed PGW	No. of processors
2009	1	3	3	1	1
2010	2	3	5	1	1
2011	2	4	5	1	1

Table 3 Number of Internet transactions with international brand cards and total number of DinaCard transactions

Year	No. of successful VISA and MasterCard Internet transactions	Total No. transactions which are made with DinaCard cards
2009.	72,467	28.7 millions
2010.	71,184	29.1 millions
2011.	104,183	31.8 millions

Discusion

This study is the first study on the e-commerce in Serbia and region. Although we showed the data on Internet transactions using Visa and Master cards in our country, we were mostly focused on our domestic DinaCard Internet payment system in the period from 2009 to 2011.

Based on the data of the Statistical Office of the Republic of Serbia on the use of Information and communication technologies (ICT) in the Republic of Serbia, the number of households who had Internet access increased from 36.7% in 2009 [12], and 39.0% in 2010 [13] to 41.2% in 2011 [14]. The same source reported Frequency of use e-commerce services by individuals and showed that the number of people who used e-commerce increased from 13% in 2009, to 18% in 2011 [12-14]. However, more than 80% of people have never used services for e-commerce.

The results of our study showed that the number of all participants in Internet payment system of DinaCard is extremely low. Moreover, these numbers showed no significant positive trend. As a consequence, number of Internet transaction in the DinaCard system was also unacceptably low. At the same time, total number of transactions in the DinaCard system, as well as the number of Internet transactions using other card brands, was far much higher. These results indicate that cardholders were familiar with the use of cards in traditional way, but also for Internet transactions, although they use DinaCard cards for Internet payments very rarely.

One of the possible reasons for this situation in Serbia is an insufficient number of the banks issuers which need to allow their DinaCard cards to be used for Internet transactions in DinaCard system. At the end of 2009, 26 out of 34 banks have been licensed for issuing any type of DinaCard cards, and at the end of 2011 there were 27 of the 33 banks in Serbia with this license [11]. However, only 1-2 banks had the license for Internet acquiring and 3-4 had the license for Internet issuing in this period. Additionally, the number of Internet merchant activated in the system was similar, up to 5. All these participants are important for the process of Internet transactions. The banks with acquiring license for Internet transactions are necessary because Internet merchants must be legally connected to the system through the banks with that type of license. Considering that only 1 or 2 banks could have provided the DinaCards for Internet payments, it is clear why the number of cards, transactions and Internet merchant is so small.

Taking into account that our region passed transition in the last two decades, it would be interesting to compare these results with the neighbouring countries. However, according to our knowledge, no studies related to e-commerce have been conducted so far in the region.

The importance of this research is in closer understanding of the need to transform business processes in line with new market requests and economic circumstances, including electronic business, electronic commerce and Internet payment systems. Since customers are the core of bank's existence [2], banks have to respond to their needs.

Conclusions

We concluded that the potential of the DinaCard system for Internet transactions is huge but untapped. The market is still developing, and the use for the DinaCards in e-commerce is negligible when compared with classical transactions with the same

card, as well as with the use of other brands cards (Visa, MasterCard) for Internet payments. In order to significantly increase the use of DinaCards for this purpose, all participants involved in this process must actively contribute by promoting the system.

References

[1] Afuah, A. and Tucci, C. (2003). Internet Business Models and Strategies. Second Edition, McGraw-Hill, New York.

[2] Bradley, L. and Stewart, K. (2003). The diffusion of online banking. *Journal of Marketing Management*, 19(10), 1087–1109.

[3] Chaffey, D. (2006). E-Business and E-Commerce Management. Third Edition, Pearson Education Limited, London.

[4] Kalakota, R. and Robinson, M. (2003). E-Business 2.0: Roadmap for Success. 2nd edition. Addison-Wesley Professional, Boston, MA.

[5] Kao, D. (2003). A strategy-based model for e-commerce planning. *Industrial Management & Data Systems*, 103(4), 238-252.

[6] Kotler, P. and Keller, K.L. (2006) Marketing menadžment. 12th edition, Data Status, Beograd. (*in serbian*)

[7] Lagator A. et al. (2007). Plaćanje na Internetu u DinaCard sistemu, VII međunarodna konferencija E-trgovina, Palic, 18-20. april (*in serbian*)

[8] Laudon K. C. and Traver C.G. (2007). E-Commerce, Business, Technology, Society, Fifth Edition, Pearson Prentice Hall, New Jersey.

[9] Montoya-Weiss, M. et al. (2003). Determinants of Online Channel Use and Overall Satisfaction with a Relational, Multi-channel Service Provider. *Journal of the Academy of Marketing Science,* 31(4), 448-58.

[10] Narodna banka Srbije, NCPK. (2009). Operativna pravila DinaCard sistema OP-V05. Beograd (*in serbian*)

[11] Narodna banka Srbije NCPK. (2010). www.dinacard.nbs.rs (*in serbian*)

[12] Republički zavod za statistiku Republike Srbije. (2009). Upotreba informaciono-komunikacionih tehnologija u Republici Srbiji, 2009. http://webrzs.stat.gov.rs. (*in serbian*)

[13] Republički zavod za statistiku Republike Srbije. (2010). Upotreba informaciono-komunikacionih tehnologija u Republici Srbiji, 2010. http://webrzs.stat.gov.rs. (*in serbian*)

[14] Republički zavod za statistiku Republike Srbije. (2011). Upotreba informaciono-komunikacionih tehnologija u Republici Srbiji, 2011. http://webrzs.stat.gov.rs. (*in serbian*)

[15] urban, T. (2000). Electronic Commerce – A managerial Perspective. Pearson Prentice Hall, New Jersey. (*in serbian*)

[16] Vukmirović, D. (2003). Internet Marketing. Fakultet Organizacionih Nauka, Beograd (*in serbian*)

Submitted: May 20, 2012
Accepted: June 06, 2012

INSTRUCTIONS FOR AUTHORS

The *Journal of Information Technology and Application (JITA)* publishes quality, original papers that contribute to the methodology of IT research as well as good examples of practical applications.

Authors are advised that adherence to the Instructions to Authors will help speed up the refereeing and production stages for most papers.

- Language and presentation
- Length of submissions
- Submission
- Contact details/biographies
- Title of the paper
- Abstract and keywords
- Figures and tables
- Sections
- Footnotes
- Special characters
- Spelling
- References
- Proofs
- PDF offprint
- Copyright and permissions
- Final material
- Correspondence
- Publication ethics

LANGUAGE AND PRESENTATION

Manuscripts should be written in English. All authors should obtain assistance in the editing of their papers for correct spelling and use of English grammar. Manuscripts should have double spacing, with ample margins and pages should be numbered consecutively. The Editors reserve the right to make changes that may clarify or condense papers where this is considered desirable.

LENGTH OF SUBMISSIONS

Papers should not normally exceed 12 Journal pages (about 8000 words). However, in certain circumstances (e.g., review papers) longer papers will be published.

SUBMISSION

Manuscripts must be submitted through the JITA online submission system.

Please read the instructions carefully before submitting your manuscript and ensure the main article files do not contain any author identifiable information.

Although PDF is acceptable for initial submission original source (i.e. MS Word) files will be required for typesetting etc.

CONTACT DETAILS/BIOGRAPHIES

A separate file containing the names and addresses of the authors, and the name and full contact details (full postal address, telephone, fax and e-mail) of the author to whom correspondence is to be directed should be uploaded at the time of submission (you should select Contact details/Biographies as the file type). This file is not shown to reviewers. This file should also contain short biographies for each author (50 words maximum each) which will appear at the end of their paper.

The authors' names and addresses must not appear in the body of the manuscript, to preserve anonymity. Manuscripts containing author details of any kind will be returned for correction.

TITLE OF THE PAPER

The title of the paper should not be longer than 16 words.

ABSTRACT AND KEYWORDS

The first page of the manuscript should contain a summary of not more than 200 words. This should be self-contained and understandable by the general reader outside the context of the full paper. You should also add 3 to 6 keywords.

Figures and tables

Figures which contain only textual rather than diagrammatic information should be designated Tables. Figures and tables should be numbered consecutively as they appear in the text. All figures and tables should have a caption.

Sections

Sections and subsections should be clearly differentiated but should not be numbered.

Footnotes

Papers must be written without the use of footnotes.

Special characters

Mathematical expressions and Greek or other symbols should be written clearly with ample spacing. Any unusual characters should be indicated on a separate sheet.

Spelling

Spelling must be consistent with the Concise Oxford Dictionary.

References

References in the text are indicated by the number in square brackets. If a referenced paper has three or more authors the reference should always appear as the first author followed by et al. References are listed alphabetically. All document types, both printed and electronic, are in the same list. References to the same author are listed chronologically, with the oldest on top. Journal titles should not be abbreviated.

Journal

Avramović ZŽ (1995) Method for evaluating the strength of retarding steps on a marshalling yard hump. *European Journal of Operational Research*, 85(1), 504–514.

Book

Walsham G (1993) Interpreting Information Systems in Organizations. Wiley, Chichester.

Contributed volume

Huberman AM and Miles MB (1994) Data Management and analysis methods. In Handbook of Qualitative Research (Denzin NK and Linkoln YS, Eds), pp 428-444, Sage, Thousand Oaks, California.

Conference Paper

Huberman AM and Miles MB (1994) Data Management and analysis methods. In Handbook of Qualitative Research (Denzin NK and Linkoln YS, Eds), pp 428-444, Sage, Thousand Oaks, California.

Unpublished reports/theses

Nandhakumar JJ (1993) The practice of executive information systems development: and in-depth case study. PhD Thesis, Department of Engineering, University of Cambridge.

Proofs

Proofs of papers will be sent to authors for checking. Alterations to diagrams should be avoided where possible. It will not be possible to accept major textual changes at this stage. Proofs must be returned to the publishers within 48 hours of receipt by fax, first-class post, airmail or courier. Failure to return the proof will result in the paper being delayed.

PDF offprint

Corresponding authors will receive a PDF of their article. This PDF offprint is provided for personal use. It is the responsibility of the corresponding author to pass the PDF offprint onto co-authors (if relevant) and ensure that they are aware of the conditions pertaining to its use.

The PDF must not be placed on a publicly-available website for general viewing, or otherwise distributed without seeking our permission, as this

would contravene our copyright policy and potentially damage the journal's circulation. Please visit **http://www.apeiron-journals.com/JITA/authors/rights_and_permissions.html** to see our latest copyright policy.

Copyright and permissions

The copyright of all material published in the Journal is held by Paneuropean University APEIRON. The author must complete and return the copyright form enclosed with the proofs.

Authors may submit papers which have been published elsewhere in a foreign language, provided permission has been obtained from the original publisher before submission.

Authors wishing to use material previously published in JITA should consult the publisher.

Final material

All final material must be submitted electronically in its original application format (MS Word is preferred). The file must correspond exactly to the final version of the manuscript.

Correspondence

Business correspondence and enquiries relating to advertising, subscriptions, back numbers or reprints should be addressed to the relevant person at:

Paneuropean University APEIRON
Journal JITA
Pere Krece 13, P.O.Box 51
78102 Banja Luka
Bosnia and Hercegovina / RS

Publication ethics

We take an active interest in issues and developments relating to publication ethics, such as plagiarism, falsification of data, fabrication of results and other areas of ethical misconduct. Please note that submitted manuscripts may be subject to checks using the corresponding service, in order to detect instances of overlapping and similar text.

JITA

PUBLISHER
Paneuropean University APEIRON,
College of Information Technology
Banja Luka, Republic of Srpska, B&H
www.apeiron-uni.eu
Darko Uremović, Person Responsible for the Publisher
Aleksandra Vidović, Editor of University Publications

ISSN 2232-9625