13TH MIDDLE EASTERN SIMULATION & MODELLING MULTICONFERENCE

MESM'2012

3RD GAMEON'ARABIA CONFERENCE

EDITED BY

Marwan Al-Akaidi

DECEMBER 10-12, 2012

Muscat, Oman

A Publication of

EUROSIS-ETI

Printed in Ghent, Belgium

13TH MIDDLE EASTERN SIMULATION & MODELLING MULTICONFERENCE

3RD GAMEON-ARABIA CONFERENCE

DECEMBER 10-12, 2012

Organised by

ETI

Sponsored by

EUROSIS

AOU

BITE	IEEE UKRI - SPC	Ghent University
HTI	LMS	University of Skovde

Hosted by

AOU – Arab Open University

Muscat, Oman

EXECUTIVE EDITOR

PHILIPPE GERIL
(BELGIUM)

EDITORS

MESM'2012-GAMEON-ARABIA'2012

General Conference Chair

Prof. Dr Marwan Al-Akaidi
Vice-Rector Academic Affairs
AOU, Kuwait

Local Chair

Dr. Moosa Al-Kindi, AOU, Faculty of Computer Studies, Muscat, Oman

LOCAL COMMITTEE

Dr. Youssef Saad Takroni, AOU, Faculty of Computer Studies, Muscat, Oman
Dr. Ahmed Al Namany, Modern College, Oman
Dr. Saqiq Ali, Sultan Qaboos University, Oman
Mr. Raja Waseem Anwar, AOU, Faculty of Computer Studies, Muscat, Oman
Mr. Sherimon P.C., AOU, Faculty of Computer Studies, Muscat, Oman
Mrs. Farah Abu Shahla, AOU, Faculty of Computer Studies, Muscat, Oman
Mrs. Sumitra Sreenivas, AOU, Faculty of Computer Studies, Muscat, Oman

INTERNATIONAL PROGRAMME COMMITTEE

Modelling Methodology and Modelling Tools
Helge Hagenauer, Universitaet Salzburg, Salzburg, Austria
Brian Hollocks, Bournemouth University, Bournemouth, United Kingdom
Yan Luo, NIST, Gaithersburg, USA
Michel Minoux, Universite P. et M. Curie-CNRS, Paris cedex 05, France

Analysis Methodology
Abdel Aitouche, Hautes Etudes Industrielles, Lille Cedex, France
Bruno Ciciani, Universita di Roma "La Sapienza", Rome, Italy
John Dalton, University of Newcastle, Newcastle-upon-Tyne, United Kingdom
Julie Dugdale, Universite Paul Sabatier, Narbonne, France
Mauro Iacono, SUN - II Universita di Napoli, Aversa (CE) Italy
Jens Kohlmeyer, Universitaet Ulm, Ulm, Germany
Francesco Moscato, Seconda Universita di Napoli, Aversa, Italy
Jairo R. Montoya-Torres, ENS des Mines de Saint-Etienne, Gardanne, France
Nadia Zanzouri, ENIT, Tunis, Tunisia

Simulation of Networks and Communication
Salvatore D'Antonio, Consorzio Interuniversitario Nazionale per l'Informatica, Naples, Italy
Eugen Dedu, LIFC, Montbelliard, France
Carlos Palau, UPV, Valencia, Spain

Complex Systems Simulation
Marta Pla Castells, Universidad de Valencia, Valencia, Spain
Michele Colajanni, Universita di Modena, Modena, Italy
Krzysztof Fujarewicz, Silesian University of Technology, Gliwice, Poland
Tom Dhaene, Ghent University, Ghent, Belgium
Daniel Huber, Heinz Nixdorf Institute, University of Paderborn, Paderborn, Germany
Jan Lemeire, VUB, Brussels, Belgium
Jorge M. Perez, ITP, S.Fernando de Hernares, Spain
Panos Periorellis, University of Newcastle Upon Tyne, Newcastle, United Kingdom
Fatima Rateb, University College London, London, United Kingdom
Marco Remondino, University of Turin, Turin, Italy
Marc Rouff, University of Caen Basse-Normandie, Caen, France
Jan Studzinski, Systems Research Institute of Polish Academy of Science, Warsaw, Poland

Simulation of Signal and Image Processing
Bruno Ciciani, Universita di Roma "La Sapienza", Rome, Italy
Tania Jimenez, LIA, Universite d'Avignon, France
Viorel Nicolau, "Dunarea de Jos" University of Galati, Romania
Bogdan Raducanu, UAB, Cerdanyola, Spain
Renate Sitte, Griffith University, Gold Coast Mail Centre, Australia
Joao Manuel R. S. Tavares, University of Porto, Porto, Portugal
Hanene Trichili, University of Sfax, Sfax, Tunisia
Ahmed M. Zeki, University of Bahrain, Bahrain

Energy Systems Simulation
Musleh B. AL-Otaibi, Alahmadi, Kuwait
Janos Sebestyen Janosy, KFKI, Budapest, Hungary
Patrick Waeger, EMPA, St. Gallen, Switzerland

Simulation in the Chemical and Petrochemical Industry
Ammar Al Khani, Espoo, Finland
Diganta Bhusan Das, Loughborough University, Loughborough, United Kingdom
Eugenio Ferreira, Universidade do Minho, Braga Portugal
Mohammad Al-Nawafleh, Tafila Applied University College, Al-Balqa' Applied University, Tafila, Jordan
Cornel Resteanu, Research Institute of Informatics, Bucharest, Romania
M R Riazi, Kuwait University, Safat, Kuwait

Multimedia and Virtual Reality Systems
Christos Bouras, University of Patras, Rion, Patras, Greece
Zhisheng Huang, Vrije University Amsterdam, The Netherlands
Marco Roccetti, University of Bologna, Bologna, Italy

Decision Processing in Management
Hadhoum Boukachour, Le Havre University, Le Havre, France
Naoufel Cheikhrouhou, Ecole Polytechnique Federale de Lausanne, Lausanne, Switzerland
Joel Colloc, Le Havre University, Le Havre, France
Jeanne Schreurs, Universiteit Hasselt, Diepenbeek, Belgium
Christine Sybord, Lyon 2 Lumiere University, Lyon, France
Behrouz Zarei, Sharif University of Technology, Tehran, Iran

Modelling and Simulation in Biomedical Systems
Manuel Alfonseca, Escuela Politecnica Superior, Universidad Autonoma de Madrid, Madrid, Spain

Modelling and Simulation for Industrial Applications
Luis Felipe Acebes Arconada, Universidad de Valladolid, Valladolid, Spain
Pascal Berruet, Universite Bretagne Sud, Lorient cedex, France
PJ Byrne, DCU Business School, Dublin City University, Dublin, Ireland
A. Chatzinikolaou, Bosch Rexroth SA, Athens, Greece

INTERNATIONAL PROGRAMME COMMITTEE

Olaf Diegel, Massey University, Albany Campus, North Shore Mail Centre, Auckland, New Zealand
Sergej Kalaschnikow, VA TECH ELIN EBG Elektronik GmbH & Co, Vienna, Austria
Markus Koch, C-LAB, Paderborn, Germany
Habtom Mebrahtu, Anglia Polytechnica University, Chelmsford, United Kingdom
Pascal Meyer, Forschungszentrum Karlsruhe, Karlsruhe, Germany
Lars Moench, TU Ilmenau, Ilmenau, Germany
Carlos E. Pereira, Universidad Federal do Rio Grande Do Sul, Porto Alegre, Brazil
Caterina Rizzi, Universita degli Studi di Bergama, Dalmine, Italy
Amir Shirkhodaie, Tennessee State University, Nashville, USA
Hans Veeke, TU Delft, Delft, The Netherlands
Ramiro Velazquez, Universite Paris 6 - CNRS, Fontenay aux Roses, France
Joao de Silva Wellington, UFPE, Recife, Brazil

Simulation in Logistics, Traffic, Transport and Harbour Simulation
Paul Davidsson, Blekinge Institute of Technology, Ronneby, Sweden
Isabel Garcia Gutierrez, Universidad Carlos III de Madrid, Leganes, Madrid, Spain
Thomas Hanne, Fraunhofer-Institut ITWM, Kaiserslautern, Germany
Daniel Krajzewicz, GermanAerospace Centre, Berlin, Germany
Seung Man Lee, NASA Ames Research Center, Moffett Field, USA
Herve Manier, Universite de Technologie de Belfort-Montbeliard (UTBM), Belfort Cedex, France
Vittorio Maniezzo, Universita di Bologna, Bologna, Italy
Rafael J. Martinez, Universitat de Valencia, Valencia, Spain
Philippe Mussi, INRIA Sophia Antipolis, France Frederic Serin, IUT du Havre, Le Havre, France

Web based Simulation
Victor Bassilious, University of Abertay Dundee, United Kingdom
Jorge Marx Gomez, Otto-von-Guericke-Universitaet Magdeburg, Magdeburg, Germany
Wolfgang Kuehn, University of Wuppertal, Wuppertal, Germany
Jose Machado, Universidade do Minho, Braga Codex, Portugal
Mara Nikolaidou, University of Athens, Ilissia, Athens, Greece
Francesco Quaglia, Universita' di Roma "La Sapienza", Roma, Italy
Stefan Sarstedt, University of Applied Sciences Hamburg, Hamburg, Germany
Alfonso Urquia, UNED, Madrid, Spain

Simulation in Information Processing
Julien Bourgeois, LIFC, Montbelliard, France
Christophe Claramunt, Naval Academy Research Institute, Brest Naval, France
Patrick De Causmaecker, KU Leuven, Camous Kortrijk, Kortrijk, Belgium
Khalil Drira, LAAS-CNRS, Toulouse Cedex 04, France
Laurent Guillon, Naval Academy Research Institute, Brest Naval, France
Jean-Claude Hochon, GFI Consulting, Toulouse, France
Peter Lawrence, Deakin University, Bentleigh, Australia
Christophe Roche, Universite de Savoie, Le Bourget du Lac cedex, France
Christian Siemers, TU Clausthal, Clausthal-Zellerfeld, Germany
Jim Torresen, University of Oslo, Oslo, Norway

Simulation in Fuzzy Systems, Neural Networks and Genetic Algorithms
Track Chair: Kifaya Qaddoum, Warwick University, Coventry, United Kingdom
Chrissanthi Angeli, Technological Education Institute of Piraeus, Athens, Greece
Paulo Alexandre Ribeiro Cortez, University of Minho, Guimaraes, Portugal
Viorel Nicolau, University of Galati, Galati, Romania

Simulation in the Built Environment
Bernadette O'Regan, University of Limerick, Limerick, Ireland
Agustin Yague, Universidad Politicnica de Madrid, Madrid, Spain

Simulation in Archaeology
Jean Yves Blaise, UMR CNRS/MCC 694 MAP, Marseille, France
Anders Hast, University of Gavle, Gavle, Sweden
Wolfgang Kreutzer, University of Canterbury, Christchurch, New Zealand
Marcos A Rodrigues, Sheffield Hallam University, Sheffield , United Kingdom

Simulation Projects in the Arab World
Saleh Al-Jufout, Tafila Technical University, Tafila, Jordan
Mohamed-Aly O. Louly, University of Technology, Troyes, France

GAMEON-ARABIA
Marco Block-Berlitz,Dresden University, Dresden, Germany
Pieter Jorissen, Karel de Grote Hogeschool, Hoboken, Belgium
Ian M Marshall, Coventry University, Coventry, United Kingdom
Joseph M. Saur, GTRI, Joint Systems Integration Center, USA

MESM'2012

GAMEON-ARABIA'2012

All articles published in these Proceedings have been peer reviewed

EUROSIS-ETI Publications are ISI-Thomson and IET referenced

A CIP Catalogue record for this book is available from the Royal Library of Belgium under nr.12620

For permission to publish a complete paper write EUROSIS, c/o Philippe Geril, ETI Bvba, Executive Director, Greenbridge Science Park, Ghent University-Ostend Campus, Wetenschapspark 1, Plassendale 1, B-8400 Ostend Belgium.

EUROSIS is a Division of ETI Bvba, The European Technology Institute, Torhoutsesteenweg 162, Box 4, B-8400 Ostend, Belgium

Printed in Belgium by Reproduct NV, Ghent, Belgium
Cover Design by Grafisch Bedrijf Lammaing, Ostend, Belgium

EUROSIS-ETI Publication
ISBN: 978-90-77381-75-5
ISN: 978-90-77381-75-5

Preface

Dear conference delegate,

I have the pleasure to welcome you to the combined 13th Middle Eastern Modeling & Simulation Multiconference (MESM2012) and 3rd annual Pan-Arabic - GAMEON-ARABIA'2012 organized by EUROSIS and hosted and sponsored by the Arab Open University (Oman Branch) in Muscat, Oman. The MESM'2012-GAMEON-ARABIA'2012 is co-sponsored by the IEEE – UKRI SPC, BITE, Ghent University, HTI, LMS and The University of Skovde. This year's event also marks our first visit to Oman with the MESM–GAMEON-ARABIA conferences.

While the MESM'2012 Conference highlights recent and significant advances in many research areas of modeling and simulation related to distributed simulation, electronics simulation, security systems and information processing, GAMEON-ARABIA'2012 looks at the advances in gaming research in the Middle East.

Next to the programme featuring the refereed and selected papers the joint event also features first of all a number of excellent keynotes by Lt.Col Dr. Ahmed Bin Subaih of the Dubai Police General HQ entitled "Gamification in the Dubai Police: Opportunities and Challenges" and by Ken Newman of HTC in Abu Dhabi entitled "Modeling Culture: An overview of attempts to model aspect of human culture and history", and secondly three turorials. Two of these tutorials are presented by Ass. Prof. Uvais Qidwai of Qatar University in Doha, Qatar entitled: "Parametric Modeling of Ultrasonic Signals in Nondestructive Testing Applications in Oil & Gas Industry" and "Modeling ECG and EEG Signals for Possible Classification Applications". The third tutorial is given by Ass. Prof. Muqeem Khan, Northwestern University in Qatar, Doha, Qatar entitled "Autodesk Maya Software and the Workflow in 3D Content Design".

As General Conference Chair of both events, I would like to express my thanks to the Arab Open University for giving me the time to chair this conference and thanks also to the committee members for reviewing the papers and to our local chair Dr. Moosa Al-Kindi, Faculty of Computer Studies, Arab Open University University, Muscat, Oman in organizing this event at our Oman Branch.

Thanks to my colleague Philippe Geril, executive director of EUROSIS office for supporting the event and for his time. Last but not least thanks to all authors without whom the conference would not be a successful conference.

Professor Dr Marwan Al-Akaidi
General Conference Chair
EUROSIS – M. East Chair
Vice-Rector Academic Affairs
Arab Open University
Kuwait

CONTENTS

CONTENTS

INFORMATION PROCESSING

GAMING

LATE PAPER

SCIENTIFIC PROGRAMME

DISTRIBUTED SIMULATION

LOAD BALANCING AND RESOURCE ALLOCATION IN DISTRIBUTED SIMULATION SYSTEMS

Giuseppe Iazeolla and Alessandra Pieroni
Enterprise Engineering Department
University of Rome TorVergata, Rome, Italy
E-mail: {giuseppe.iazeolla,alessandra.pieroni}@uniroma2.it

KEYWORDS
Distributed and parallel simulation methodology, distributed simulation modelling, simulation performance evaluation.

ABSTRACT

Assume a local simulator (LS) of a given system Σ is available and that we wish to turn it into a distributed simulator (DS). In the DS case, the LS is partitioned into segments called federates, each federate being run by a separate host. Before implementing the DS (i.e., at design-time) we ask: will the DS execution time be shorter than LS one? In some cases the DS may run slower than the equivalent LS. To answer this question we are to consider that the execution time of a distributed simulation system depends on 3 interacting factors: 1) the speedup "S" (or run-time gain) resulting from the partitioning of the local simulator into federates. 2) The synchronization-message communication overhead. 3) The data-message communication overhead. The speedup depends on the load balancing between the parallel-operating hosts. The message delays depend on the network resource allocation. The combination of such factors makes very hard predicting the benefits of the LS to DS transformation. In this paper a DS load balancing and resource allocation decision procedure is proposed to evaluate the benefit of the LS to DS transformation. The procedure is guided by a performance model of the DS. The High Level Architecture (HLA) distributed simulation standard is assumed to be used.

1. INTRODUCTION

A simulation model can be seen as consisting of a set of sub-models. In local simulation (LS), a single model exists that simulates the entire system and is run by a single host. In distributed simulation (DS), various sub-models (called federates) simulate distinct parts of the system and are run by separated hosts connected via a local, metropolitan or wide-area network (LAN, MAN or WAN) or a composition thereof.
Predicting at design-time the convenience of implementing the DS version of the LS can be of interest. Indeed, the development of a DS system is a complex and expensive task, including the cost of achieving the necessary know-how of the distributed simulation standard, the cost of the extra-lines of code to develop for each federate [2, 4], the cost of the hosts and the computer networks, the number of design alternatives to face (in terms of simulator partitioning, host capabilities, network potentialities and so on).
This paper introduces a procedure to support the evaluation of the DS convenience before implementation. The method investigates the effects of two elements:
1) The load balancing among the various hosts that operate in parallel; 2) The allocation of network resources for the exchange of synchronization messages and data messages among federates.
The management of the load balancing among the parallel hosts affects the achievable speedup. On the other hand, the management of network resource allocation affects the synch-communication overhead and the data-communication overheads that both work against the speedup, by lowering down the run-time gain obtained with the speedup.
A DS load balancing and resource allocation procedure is proposed to choose (at design-time) whether to remain on the LS version of the simulator or carry out the implementation of its DS version. The procedure is guided by a performance model (PM) of the DS. The model can be used to perform what-if analysis and sensitivity analysis to observe how managing the tradeoff between the parallel load balancing and the network resource allocation may affect the DS execution time. The PM assumes the DS is based on the HLA protocol standard and middleware [8].
The paper is organized as follows: Sect.II presents the problem statement. Sect.III illustrates the PM and its implementation in the OMNet++ simulation language and its use in the DS load balancing and resource allocation procedure. Sect.IV presents the paper contribution with respect to existing literature and finally Sect.V gives concluding remarks.

2. PROBLEM STATEMENT

Assume a local simulator (LS) of a given system Σ is available, and that we wish to turn it into a distributed simulator (DS).
In the DS case, the LS is partitioned into sub-models called federates, each federate being run by a separate host. Fig.1 shows the two federate case, with N_S denoting the network for the exchange of synch messages and N_D the one for data messages.
Before implementing the DS (i.e., at design-time) we wonder: will the DS execution time be shorter than LS one? In some cases the DS may run slower than the equivalent LS. To answer this question a Performance Model (PM) of the DS is introduced.

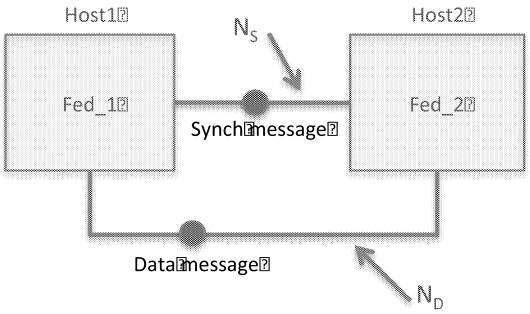


Figure 1. DS system with two federates

For the sake of simplicity a two federate PM is presented. It is understood that as the number of federates grows, the degree of complexity and the amount of communication overhead increases depending on the application and the implementation characteristics. Modeling the k-federate case, however, is out of the scope of this work. In the paper, indeed, we only wish to show that one cannot neglect the trade-off between parallel load-balancing and network resource-allocation even in a simple two federate case.
The following terminology will be used throughout the paper:

Σ = System to be simulated
LS(Σ) = Local Simulator of Σ
T_{LS} = LS execution time
DS(Σ) = Distributed Simulator of Σ
T_{DS} = DS execution time
PM(DS(Σ)) = Performance Model of DS(Σ) to predict the execution time T_{DS}

The question is: when does DS(Σ) run faster than LS(Σ)?

A. The management of the speedup/communication overhead trade-off

As with most parallel computations, to obtain a significant speedup the portion of LS that can be parallelized must be large relative to the portion that is inherently serial. Let us denote by S(N) the maximum speedup that can be achieved using N processors and by Q the fraction of computation that is inherently serial. According to Amdahl's law [3,12] even with an arbitrarily large number of processors ($N\to\infty$), S(N) can be no larger than the inverse of the inherently serial portion Q of LS.
Thus, one requirement for the DS code to achieve significant speedups is that the fraction Q should be small. An appropriate partitioning of LS into a set of federates should then be found at design-time that improves S while maintaining the synch- and data-message communication overheads low. In other words, a partitioning that yields a high computation-to-communication ratio (i.e., a large amount of computation between communications). On this basis, a DS load balancing and resource allocation procedure can be foreseen (Fig.2) to choose whether to remain on the LS version of the simulation system or carry out the implementation of its DS version. In other words, assume a LS(Σ) has been developed and that its T_{LS} is not satisfactory.

A search for an appropriate partitioning of LS(Σ) into federates and for an appropriate choice of the N_S and N_D networks has to be performed by the iterative use of the PM(DS(Σ)), to obtain a distributed simulator with $T_{DS} < T_{LS}$.
At each iteration, if the T_{DS} predicted by the PM is sufficiently lower than T_{LS}, the decision to implement the DS(Σ) can be taken.
Otherwise, one may either try a new tentative partitioning and load balancing (in other words, new portions of the original local simulator are assigned to various federate, and thus to various hosts) or try alternative network N_S and N_D allocation with improved capabilities. In both cases, a new performance model of the DS is run and the DS execution time is predicted without the need of really developing the DS. In case no partitioning nor network improvements can be found, one may decide not to implement the DS(Σ).
An example use of the PM in the DS load balancing and resource allocation procedure is illustrated in Sect.III.A. The PM cannot be evaluated by analytic methods and thus its evaluation is simulation-based. The coding of the PM is done in the OMNet++ simulation language [11] and an example coding is provided in Sect.III.C.

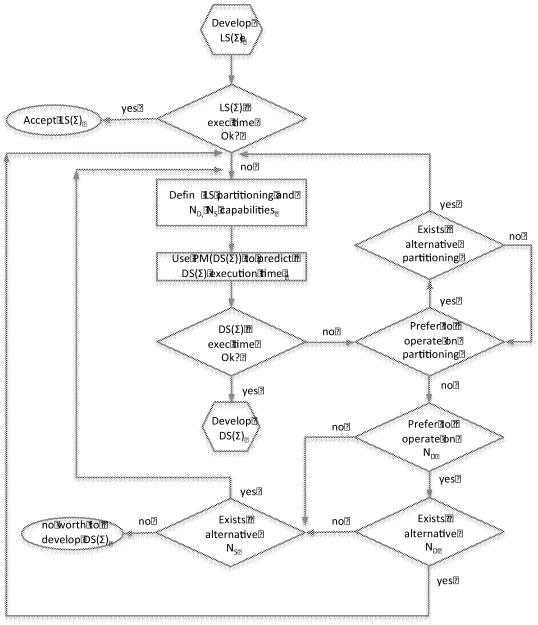


Figure 2. DS load balancing and resource allocation procedure

3. THE PERFORMANCE MODEL OF THE DS(Σ)

It is assumed the reader is familiar with the structure of an HLA federation, based on the so-called Run Time Infrastructure (RTI) [13]. The RTI is the software that allows the federates to execute together. In Fig.3 the interface, between the RTI and the federates, is illustrated [10]. The federates do not talk to each other directly. They are instead connected to the RTI and communicate with each other using services provided by the RTI. The RTI offers to each federate an interface called *RTI Ambassador*.

Each federate on the other hand presents an interface called *Federate Ambassador* to the RTI.

In the following we shall denote by:

- LEX the local execution internal to a federate, in other words, the standard simulation operations such as event processing, event routines, scheduling of local events, etc.

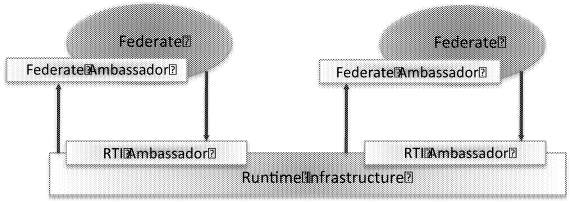


Figure 3. HHLA federation structure

- HLAR the execution internal to a federate of an RTI service, e.g., an invocation of a time advance request.
- HLAF-Ex the execution internal to a federate of a service request coming from the *Federate Ambassador*.

The PM(DS(Σ)) consists of two model: the federation performance model and the network performance model. In this paper we deal, in particular, with the federation performance model (limited to the two federates case, as said above). while giving only a few details of the network performance model.

A. The federation performance model

To answer the Fig.2 question "Exists alternative partitioning?" we shall assume that the LS is partitioned into a 2-federates DS and shall evaluate the PM of such a partitioning.

The PM of the 2-federates DS is shown in Fig.4, where the details of the PM of only one federate (Fed_1) are illustrated. One may partition the LS code into the 2-portions of the DS code in various ways. As we shall better see in Sect.III.C, the effect of the partitioning choice is reflected in the value given to a model *parameter* (parameter p_{SYNC}) at model parameterization time.

The model includes three time-consuming nodes:

- *CPU* that synthetically represents the host that runs the federate.
- N_S representing the synch network, that transports synchronization messages from Fed_1 to Fed_2, and vice versa.
- N_D representing the data network, that transports data messages from Fed_1 to Fed_2, and vice versa.

Besides such time-consuming nodes, a set of non-time consuming nodes can be found, namely:

- *AND* nodes that perform AND-logic operations.
- *OR* nodes that perform OR-logic operations.
- *SPLIT* nodes, that split an incoming job class into two or more outgoing classes.
- *Classifier* nodes that, basing on the class of the input job, forward the job in one or the other direction.
- *Router* nodes that perform probabilistic routing of the incoming jobs.
- *Merge* node that merges two job classes.

The computation performed by the federate is carried out by jobs of various classes that circulate in the PM, namely:

- Class C_1 jobs
- Class C^D_1 jobs
- Class C_{HLA} jobs
- Class C_{RL} jobs
- Class C^D_{RL} jobs
- Class C_{RC} jobs
- Class C^D_{RC} jobs

The only jobs that consume *CPU* time are C_1 and C_{RC}. The relative *parameters*, (e.g. mean CPU time $E(t_{CPU})$, variance, etc.) can be obtained [2] basing on the model of the federate software run by the *CPU* and the *CPU* capacity. The class C_1 job simulates the so-called federate *main thread* [10], performing LEX and HLAR computations. The class C_{RC} job simulates the so-called federate *RTI callback* [10], performing HLAR-Ex computations.

A class C^D_1 job is a job derived from C_1 and holding the data payload to be forwarded to Fed_2 through network N_D, when the RTI-Ack arrives from Fed_2 (see the AND^D_{12} node). A class C_{HLA} is a job derived from C_1 and holding the synch message to be forwarded to Fed_2 through network N_D. A class C_{RL} job represents the so-called federate *request listener thread*, waiting for synch messages from Fed_2 (see the AN^D_{21} node). A class C^D_{RL} job is the federate request listener thread, waiting for data messages from Fed_2 (see the AND^D_{21} node). A class C^D_{RC} job is the federate request callback thread holding the data payload coming from Fed_2 and to be used by the C_1 job class.

It is assumed that the *conservative-time management* HLA option is used, in other words, no federate advances logical time except when it can be guaranteed not to receive any events in its past. If we also assume *zero lookahead* (actually HLA does not accept a null value for lookahead and thus a very small value is given to this parameter) there is guarantee that federates do not receive events in the past and thus that they are fully synchronized. For such choices, the federates will not process events in parallel and parallelism will only be found when federates include intrinsically parallel portions of LS. If this holds, a significant speedup will be obtained when transforming the LS into its DS version.

The computation performed by the federation starts by launching the RTI interface and initializing the HLA components local to each federate. Such initial computations are performed only once and do not substantially affect the federation execution time and thus are omitted from the modeling. They are synthetically represented by the INIT job on the top of Fig.4. The INIT job enters the $Split_0$ node and yields the main thread C_1 for Fed_1 and the main thread C_2 for Fed_2. The main thread C_1 enters the $Split_1$ node and yields three outgoing jobs: C_1 itself again, C_{RL} and C^D_{RL}.

The job of class C_1 enters the *CPU* processing queue from $Split_1$ and circulates in the model (in a way that we shall soon illustrate), so iteratively reentering the CPU processing queue coming from the AND_{MT1} node (in case a synchronous HLA service is invocated (p_{SYNC}), otherwise the C_1 job returns directly to the CPU processing queue without enter AND_{MT1} node, as better explained below). The job of class C_{RL}, instead, enters the OR_{21} node and from here enters the AND_{21} node where it waits for a synch message from Fed_2 to generate a C_{RC} job, which through the $Split_3$ produces both a new C_{RL} job (that waits for future synch messages) and the C_{RC} itself again that enters the CPU processing queue. The same logic applies to the CD_{RL} job coming from $Split_1$, which enters the OR^D_{21} and the AND^D_{21} nodes waiting for a data message from Fed_2. The C^D_{RC} job outgoing $Split_4$ does not enter the CPU processing queue directly but merges itself with the C_1 circulating main thread through the merge node M_1. As said above, the C_1 job entering the CPU performs LEX and HLAR computations, while the C_{RC} job performs HLAR-Ex computations.

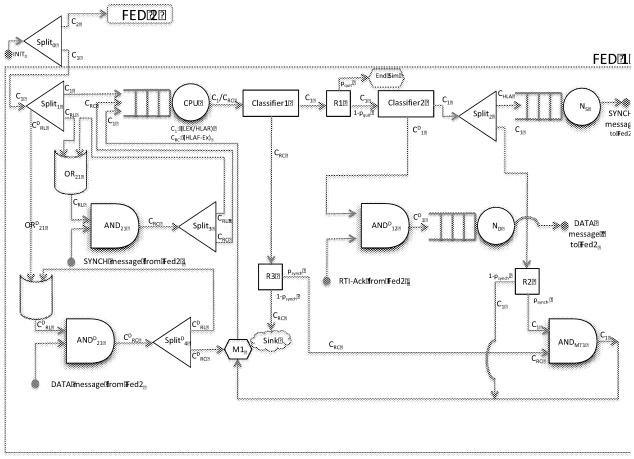


Figure 4. View of the federation PM with details of Fed_1

The job leaving the *CPU* can be a C_{RC} or a C_1 job.
In case the job leaving the *CPU* is a C_{RC} job, the *Classifier1* node forwards it to the router *R3*, which sends the job to the AND_{MT1} node if the HLA synchronous modality [10] is used. Otherwise, the C_{RC} job has no effects and is absorbed by the sink node. If directed to the AND_{MT1} node, the C_{RC} job gives consensus to the circulation of the main thread C_1, which thus re-enters the CPU processing queue.
In case instead, the job leaving the *CPU* is a C_1 job, the *Classifier1* directs it to the *R1* router, which sends the job to *Classifier2* if the simulation is not ended ($1-p_{QUIT}$). Here, if C_1 contains a data-message, a C^D_1 job is produced which enters the AND^D_{12} node, and waits for the RTI-Ack from Fed_2 in order to be forwarded to Fed_2 through network N_D. If instead, the outcome from *Classifier2* is a no-data message C_1, this enters the $Split_2$ node and yields a C_{HLA} job (holding a synch-message to be forwarded to Fed_2 through network N_S) and again a circulating main thread C_1, which (in case a synchronous HLA service is invocated (p_{SYNC})) reaches the aforementioned AND_{MT1} node to iterate the main thread circulation. In case, instead, of no-synchronous HLA service ($1-p_{SYNC}$), the C_1 job does not need the AND_{MT1} consensus to iterate the main thread circulation, and returns directly to the CPU processing queue.
Let us conclude this Section by pointing out that in building the federate model we did not make any mention of the simulated system Σ. This is since the federate model we introduce in this paper is independent from Σ, i.e. it is valid for any Σ. In other words, the paper model can be used for any HLA-based simulation. Only its parameters may depend on Σ, as better seen in Sect.III.C.

B. The network performance model

A further model is necessary to answer the second Fig.2 question of the DS load balancing and resource allocation procedure: "Exists alternative NS (or ND)?". The needed model is the model of the computer network connecting the federation hosts. By use of such a model, the "Exists alternative N_S (or N_D)?" question can be answered by making what-if and sensitivity analysis of the various network components of both N_S and N_D.
Giving details of the network performance model is out of the scope of this paper and the reader can find in [2] the possible approach to such a modeling to identify the N_S and N_D parameters to be used in the PM(DS(Σ)), see Sect.III.C.

C. The OMNet++ version of the PM (DS(Σ)) and model parameterization

To perform an example prediction of the DS(Σ) execution time (T_{DS}) to be used in the Fig.2 DS load balancing and resource allocation procedure, we developed the OMNet++ simulation version (shown in Fig.5) of the model in Fig.4. Only the Fed_1 part and the N_S and N_D nodes are represented in Fig.5.
As said above, the model structure is valid for any system Σ and only its parameters, illustrated in Tab.1, (i.e., the CPU service time, the N_D and N_S service times, the p_{QUIT} and p_{SYNC} routing probabilities) may change with Σ. The derived parameters for a Σ example-case (whose details can be found in [5], but are not essential to be known since the PM structure is independent from Σ) are illustrated in Tab.1. As can be seen from Table1, there exist three types of parameters: the federate Host CPU parameters, the N_S, the N_D parameters and the routing parameters (p_{QUIT} and p_{SYNC}) illustrated in Sect.III.A.

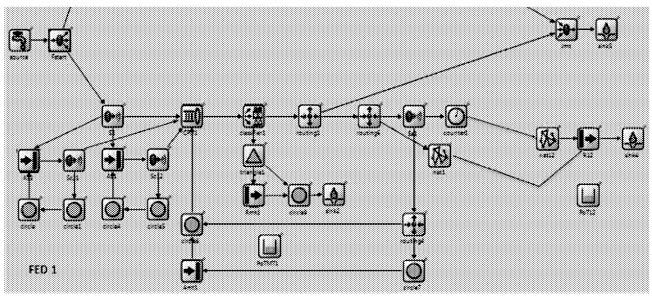


Figure 5. OMNet++ simulation version of the PM(DS(Σ)) Fed_1.

Table 1: Model parameters for a two-Federate DS

	Distribution	Parameters
CPU service time t_{CPU}	positive truncated-Normal	$E(t_{CPU})$ = 10ms (Scen.A) $E(t_{CPU})$ = 500ms (Scen.B) $\sigma^2(t_{CPU}) = 1$
N_S, N_D service time t_S	k-Pareto, k =4	$E(t_S) = 21$ms
Routing parameters	p_{QUIT}	0,001 (Fed_1) 0,001(Fed_2)
	p_{SYNC}	0,82 (Fed_1) 0,74 (Fed_2)

The federate Host CPU service time parameters vary with the job class (C_1 or C_{RC}) and are derived from the CPU capacity and the federate software run by the CPU, as seen in Sect.III.A. For the sake of simplicity, in this example, a common mean $E(t_{CPU})$ of 10ms or 500ms (for Scenarios A and B respectively, see later) is chosen for both classes.
The parameters for the N_D and N_S networks are instead derived from the network models, as briefly seen at the end of Sect.III.B.
The routing parameters p_{QUIT} and p_{SYNC}, finally, can be derived from measurements on LS(Σ), in particular, by counting the number of events $n_{intEvents}$, $n_{disEvents}$, $n_{disToIntEvents}$ which respectively denote the number of local events (internal events), the number of events sent from a potential Fed_1 to a potential Fed_2, and the number of events received from the potential Fed_2. Such counting can be easily performed collecting the number of LS events in a simulation experiment for a given hypothetical LS partitioning into two federates [5].
As said above, the value probability p_{SYNC} takes, reflects the model partitioning. Indeed, it is possible to be convinced [5] that under the *conservative time-management* assumption, its value depends on the number of local events processed within the model partitioning border and on the number of distributed events generated by the model partition.
Basing on the federate Host CPU parameters, the N_S and N_D parameters and the routing parameters, the OMNet++ code simulation model has been run to obtain the T_{DS} predictions shown in Tab.2 [6, 7]. This was carried-out in two scenarios A and B: Scenario A being one in which the fraction Q of inherently serial computation was high and Scenario B in which Q was low. The first column in Tab.2 reports the local simulator execution time T_{LS}. The second column reports the distributed simulator execution time T_{DS} predicted by OMNet++ version of the PM, and the third column the times of the real distributed simulator DS (that was implemented in Java+HLA). Such a column shows how the times predicted by the PM adequately match the real ones. Note that in Scenario B the execution times are in minutes while in Scenario A they are in seconds. This is since Scenario B is built in a way to yield a high computation-to-communication ratio. In other words a large amount of computation between communications.
Tab.2 also shows how in the Scenario B the distributed simulator outperforms the local one.

Table 2: Execution-time results

	T_{LS}	PM results (OMNet++ predictions)	real DS results
A (high Q)	0.7s	T_{DS} = 8.3s	T_{DS} = 8.2s
B (low Q)	33 min	T_{DS} = 12.5 min	T_{DS} = 12.0 min

Indeed, in such a Scenario the DS execution time (T_{DS}) is much lower than the LS time (T_{LS}). Finally by using the expression $S = T_{LS}/T_{DS}$, the results in Tab.2 were used to obtain the speedup results shown in Tab.3:

Table 3: Speedup results

	PM results
A: High Q	S = 0.08
B: low Q	S = 2.64

This table shows that a significant speedup (S = 2.64) is obtained in the B Scenario. In other words, in this case the run-time gain obtained by the parallel execution on two hosts compensates for the data and synch communication overheads.
In the scenario A, instead, the parallelism does not yield a sufficient run-time gain to compensate for the overheads, and the resulting speedup (S = 0.08) is practically irrelevant.
The Tab.2 and 3 results are used by the DS load balancing and resource allocation procedure of Fig.2 to decide at design-time whether to remain on the LS version of the simulator or carry out the implementation of its DS version. In case the T_{DS} execution time is considered not "ok" (see Fig.2), one may either try a new tentative partitioning (that modifies the federation p_{SYNC} parameter, see Fig.4) or try alternative networks N_S and N_D of improved capabilities (that modifies their $E(t_S)$ parameter). In case no partitioning nor network improvements can be found, one may decide not to implement the DS(Σ).

4. RELATED WORK

A number of existing contributions can be found in literature that address the prediction of execution times of simulation systems, see e.g., [1,9,12,14,15,16]. Such contributions deal with parallel simulations, run by a set of processors that are tightly connected by use of multi-processor interconnection network (rather than by a computer network) and coordinated by a dedicated operating system.

Our work does not deal with the parallel case of above, but rather with the distributed case, in other words the case of parallel simulations run by a set of processors connected by a computer network, such as a LAN, a MAN or a WAN, or a composition thereof and coordinated by an appropriate middleware, such as HLA.
Besides being important for the DS load balancing and resource allocation procedure, the knowledge of the effects of the two communication overheads is also of importance to evaluate the *representativeness* of the DS(Σ) at design-time.
Indeed, depending on the nature of system Σ, there are situations in which the data and synch message delays are not critical and thus a communication network of any capability can be used. On other situations, instead, the system Σ can be of such a nature that the synch and data delays become very critical for the representativeness of the system. In other words, the DS(Σ) looses the capability of realistically representing the original Σ in case the N_S and N_D networks are not sufficiently performing.

5. CONCLUSION

The execution time of a distributed simulator (DS) depends on 3 interacting factors: the speedup, the synch-communication overhead, the data-communication overhead. The combination of such 3 factors makes it very hard to predict the advantage of transforming a local version of the simulator (LS) into a distributed version (DS).
A DS load balancing and resource allocation procedure has been proposed to decide at design-time whether to remain on the LS version of the simulator or implement its DS version.
The procedure is guided by a performance model (PM) of the DS. The model can be used to perform what-if analysis and sensitivity analysis to observe how managing the tradeoff between load balancing among the parallel hosts and network resource allocation may affect the DS execution time.
The PM assumes the DS is based on the HLA protocol standard and middleware. The model can be used both to support the DS load balancing and resource allocation procedure and to evaluate the representativeness of the DS(Σ) at design-time.

ACKNOWLEDGEMENT

Work partially supported by funds from the FIRB project on "Software frameworks and technologies for distributed simulation", from the FIRB project on "Performance evaluation of complex systems", from the University of Rome TorVergata research on "Performance modeling of service-oriented architectures" and from the CERTIA Research Center.

REFERENCES

[1] L. Chu-Cheow, L. Yoke-Hean, et al.: "Performance prediction tools for parallel discrete-event simulation", Proceedings of the thirteenth workshop on Parallel and distributed simulation Atlanta, Georgia, United States: IEEE Computer Society, 1999.

[2] A. D'Ambrogio, G. Iazeolla. "Steps towards the Automatic Production of Performance Models of Web-Applications", Computer Networks, n.41, pp 29-39, Elsevier Science, 2003.

[3] R.M. Fujimoto: "Parallel and Distributed Simulation Systems", John Wiley & Sons 1999.

[4] A. D'Ambrogio, D. Gianni, G. Iazeolla: "A Software Architecture to ease the development of Distributed Simulation Systems", Simulation-Transaction of the Society for Modeling and Simulation International, Vol. 87, n.9, pp. 819-836, 2011.

[5] D. Gianni, G. Iazeolla and A. D'Ambrogio: "A methodology to predict the performance of distributed simulation", PADS10 the 24th ACM/IEEE/SCS Workshop on Principles of Advanced and distributed simulation Atlanta May 17-19, 2010.

[6] G. Iazeolla, A. Gentili and F. Ceracchi: "Performance prediction of distributed simulations", Technical Report RI.02.10, Dept. Computer Science, University of Roma TorVergata , 2010.

[7] G. Iazeolla, M. Piccari: "The Speedup in distributed simulation", Technical Report RI.03.10, Software Engineering Lab, Dept. Computer Science, University of Roma TorVergata, 2010.

[8] IEEE Std 1516: "IEEE Standard for Modeling and Simulation (M&S) High Level Architec-ture (HLA) frameworks and rules", 2000.

[9] J. Liu, D. Nicol, et al.: "A Performance prediction of a parallel simulator", Proceedings of the thirteenth workshop on Parallel and distributed simulation Atlanta, Georgia, United States: IEEE Computer Society, 1999.

[10] F. Kuhl, R. Weatherly, J. Dahmann: "Creating Computer Simulation Systems", Prentice-Hall, 1999.

[11] OMNeT++ Discrete event simulation v.4.0. User Manual, http://www.omnetpp.org.

[12] A. Park: "Parallel Discrete Event Simulation", College of Computing. vol. PhD: Georgia Institute of Technology, 2008

[13] Pitch The Certified Runtime Infrastructure for HLA 1516 – User's guide. http://www.pitch.se, 2005.

[14] R. Ewald, D. Chen, et al.: "Performance Analysis of Shared Data Access Algorithms for Distributed Simulation of Multi-Agent Systems", Proceedings of the 20th Workshop on Principles of Advanced and Distributed Simulation: IEEE Computer Society, 2006.

[15] R. Ewald, J. Himmelspach, et al.: "A Simulation Approach to Facilitate Parallel and Distributed Discrete-Event Simulator Development", Proceedings of the 10th IEEE international symposium on Distributed Simulation and Real-Time Applications: IEEE Computer Society, 2006.

[16] K. S. Perumalla, R.M. Fujimoto, et al.: "Performance prediction of large-scale parallel discrete event models of physical systems", Proceedings of the 37th conference on Winter Simulation, Orlando, Florida, Winter Simulation Conference, 2005.

Intelligent Agent Vs. Non Intelligent Agent in Grid Computing Based on Multi-unit Auction

R. Arfa
Department of Mathematics
and Computer Sciences
University of Antwerp
Antwerp, Belgium
email: Arfa.Ramadan@ua.ac.be

P. Hellinckx
Department of Applied Engineering
Karel de Grote University College
Antwerp - Belgium
email: peter.hellinckx@kdg.be

J. Broeckhove
Department of Mathematics
and Computer Sciences
University of Antwerp
Antwerp, Belgium
email: Jan.Broeckhove@ua.ac.be

KEYWORDS
Grid, resource management, grid economies, multi-unit auction, price prediction, neural networks back propagation, fuzzy logic system, agents.

ABSTRACT

Economic principles have been used in the investigation of resource allocation algorithms in computational grids. In this paper we investigate a computational market with a multi-unit auction and agents acting on behalf of resource consumers. Each agent has private resource valuations and budget. We study a model to predict bid prices in the auction, and a model to drive the agent's decisions on accepting or rejecting bids. The aim is to aid resource consumer achieve objectives such as saving budget while meeting a deadline for computational jobs to be finished. The models are evaluated using simulations with a Java-based grid economics simulator. A number of scenarios are checked to see whether the model's intelligence does in effect help in achieving the cost/deadline objective.

INTRODUCTION

A key component of any large-scale distributed system, such as grids or peer-to-peer systems, is a resource management that allows the resources to be allocated and utilized efficiently in a coordinated fashion (Krauter et al. 2002). Traditionally, resource management approaches have favored metrics, such as throughput, that are system-centric to gauge their effectiveness. However, large-scale infrastructures have an diverse user base and thus should use allocation algorithms that take into account the valuations of individual user's of the resources to be used. A user-centric approach improves the utility of the infrastructure to all users, even with diverse requirements.

One such user-centric approach is based on the use of economic principles in resource management and it has been considered by a number of authors (Krauter et al. 2002). An economics based approach involves the definition of a computational market where resources are traded. These resources are often CPU's, but can also be bandwidth, storage space, etc. Prices signal the abundance of the resources provided by suppliers with respect to the demand formulated by the consumers. Willingness to meet a certain price to acquire a resource for completing a task reflects the value attached to the task. The market organization can be of different types such as posted price, auctions, or commodity markets (Buyya et al. 2002).

In this work we focus on a computational market, i.e. the resources are CPU's. All resources are considered equivalent, i.e., prices are not tied to specific machines. Evidently, such an approach is limited to compute-intensive tasks without considerations of data locality for input and output. Market participants communicate their valuation of resource as a willingness to supply (for providers) or acquire (for agents) a certain number of resources.

We consider a multi-unit auction market that uses dutch auction roles for choosing the winners (Leyton-Brown et al. 2000), (Ausubel 2006). The agent bids for resources and the providers offers these resources as a package of multiple speeds. The owner of the highest bid is a winner of each auction.

We introduce agents acting in the market on behalf of resource consumers. Each agent has their private resource valuation and disposes of their own budget. Intelligence is embedded into the agent via a model to predict bid prices in the auction and a model to drive the decisions on accepting or rejecting bids. The former is based on neural network and the latter is based on fuzzy techniques. The price model helps the agents to set a reference bid price to avoid overspending (Vanmechelen et al. 2006). Its basic idea is that the intelligent agents typically have some pre-knowledge about the history of the winning bids. Such pre-knowledge can used to derive better estimates for the bids for new transactions.

The question that we try to answer in this paper is whether the use of intelligence improves the resource consumer performance in achieving objectives such as saving budget while meeting a deadline for computational jobs to be finished. We do so by having an intelligent agent act in a market populated by non-intelligent

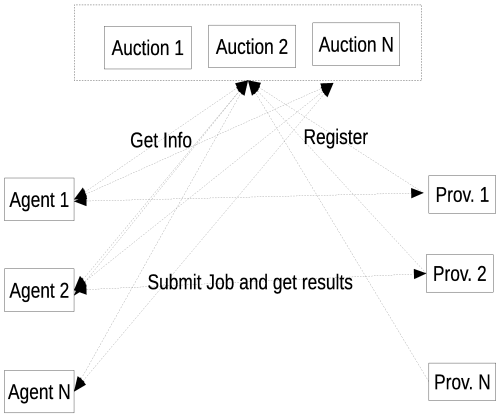


Figure 1: Interaction between agents and providers.

agents and compare their performance in term of budget spent and jobs finished before deadline. A further investigation of what happens when multiple intelligent agents interact in a single computational market is the subject of future work.

SYSTEM MODEL

Our grid environment consists of several participants: resource providers and agents acting for resource consumers. Each agent has one or more independent jobs for execution and is willing to pay. The providers have computational resources (CPUs) and are willing to rent them for profit.
The formal specification of our multi-unit auction includes a number of agents, n, a number of providers p and the mechanism rules. In general our multi-unit auction comprises p separate auctions, one for each provider, that operate over a round of bidding. In each round each agent submits a bid in every auction that chooses to bid in. The bid price for provider i is β_i, the highest bidding price received so far. The β_i is set to zero if the provider i has not received any bid yet.
The winner of every auction is announced at the beginning of next round. If an auction receives multiple admissible bids in a given round the highest bid will win. Figure 1 shows the interaction sequences between the agents and providers. Providers register and get an auction number, and the agent asks for information about each auction and then, if it so chooses, bids in that auction. At the end of each auction, the winner acquires the resource and submits the job for which the CPU resource was obtained for execution. When the job is finished its results are returned to the consumer.

Resource Model

In this study we simulate only one type of resources namely (CPU). Each resource provider willing to participate into the grid can offer multiple CPUs with different speeds that he puts into the resource pool for bidding. The different CPUs with different speeds define categories which we label CPU_i. Resources within each category CPU_i are identical in performance. Performance of the CPU_i is expressed by a performance ratio $PerfRatio_i$ relative to CPU_1 that is considered the reference. We have arbitrarily chosen linear performance ratio's i.e. $PerfRatio_i = i$.

Agents and Jobs

Each agent has a queue of CPU-bound computational jobs that need to be executed. Every job has a nominal running time, i.e. the time it takes to finish on a reference CPU. Jobs are taken to be atomic , in the sense that they are always allocated to a single CPU and are non-pre-emptable. The agents on the system are divided into three groups according to their base allowance in order represent diversity in the user base as far as spending capacity is concerned.

Bidding policies

As indicated before, the intelligent agent operates in an environment of non-intelligent agents. The bidding policy of the former is determined by the prediction model. The intelligent agent uses price returned by the prediction model for next auction to determine its bid. The non-intelligent agents use either of the following policies:

- Uniform policy: in this policy the budget of non-intelligent agent are distributed uniformly over time slots in the time period up to the deadline and the agents use the budget available at any particular time to bid.

- Random policy: in this policy, the non-intelligent agent uses a random amount for bidding in each time step. The bids are created randomly between the low and high number of the membership function of our fuzzy logic function.

Price prediction

The price prediction is based on the use of a neural network trained with historical price information. The historical data are the prices that result from each auction round in the past. The data are divided into two groups of data sets at a ratio of 2:1. The first part is used learning purpose while the last part are reserved for testing. The historical data are updated each time with a recent results from last steps. We have used a fairly common architecture for a small neural network. It is a feed-forward network with input layer, a hidden layer and an output layer, with full connection topology between the layers. For more details on the setup and its use in the context of grid economics, we refer

number of iterations	500
number of agents	150
number of providers	75
resources per provider	min 5, max 20
number of resources type	3
CPU_1 performance ratio	1.0
CPU_2 performance ratio	2.0
CPU_3 performance ratio	3.0
base allowance per group	$1.0M, 2.5M, 5.0M$
budget ratio per group	$\{2.0, 1.5, 1.0\}$
initial number of jobs	$\{1000, 5000\}$
job running time length	$\{2, 3, \cdots, 10\}$

Table 1: GES simulation parameters.

base allowance	$4.0M$
initial number of jobs	1500
job running time length	5
deadLine step time	150

Table 2: Parameters for intelligent and non-intelligent agent.

to (Arfa and Broeckhove 2008b),(Arfa and Broeckhove 2008a) and (Arfa and Broeckhove 2009).
Notice that in our approach, the intelligent agent uses price prediction to determine the bid price, but not to identify specific auctions and time step in which to participate in.

FUZZY MODEL

The fuzzy logic model is defined by a set of logical rules and a set of membership functions. The membership functions describes the range of values that each object can have at any time. The logical rules uses fuzzy variables to make the predictions. The fuzzy variables can have different grades of membership in the interval [0,1] defined by the membership function. Let $\mu_y(x)$ the membership function expressing the degree of membership of object x with respect to to fuzzy set y. The closer the value to 1, the higher the membership of x for set y. Our fuzzy system is constructed using the logical expressions IF-THEN to show the relationship of fuzzy set members (Lin et al. 2006),(He et al. 2004),(Yao 1998).
In our application, the membership values are chosen according to the average of the low and the high bids in the historical data. The prediction model that we build uses the history data to predict the next bid price P_i at each auction round of the market. The returned bid price is then subjected to the fuzzy logic rules to decide the final bid price for the intelligent agent. The constructed rules for our fuzzy logic are:

- **IF** P_i is **low** and (current step **less than** τ) **THEN** bid=bid. Here τ denoted a critical time threshold defined as an 20% of remaining time from the deadline of the job.
- **IF** P_i is **low** and (current step **greater or equal** τ) **THEN** bid=bid+(bid*α). Again τ is the critical threshold defined before, and $\alpha = .1$ increases the current bid level by 10%.
- **IF** P_i is **medium** or is **high** and (deadline-current step **greater than** τ) **THEN** bid=0. Essentially this rule says: under these conditions, abstain from bidding.
- **IF** P_i is **medium** or is **high** and (deadline-current step **less or equal** τ) **THEN** bid=bid+(bid*α). Again τ is the critical threshold defined before, and $\alpha = .1$ increases the current bid level by 10%

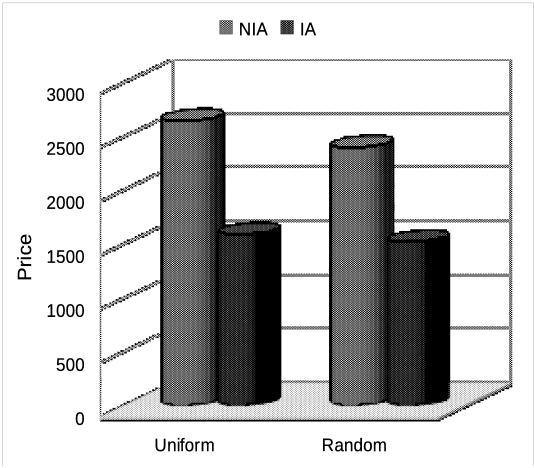


Figure 2: Average over all jobs of budget spent per job for intelligent agent (IA) and non-intelligent agents (NIA) for uniform and random policies.

SIMULATION

The models are evaluated using simulations with a Java-based grid economics simulator (GES). A number of scenarios are checked to see whether embedding the model's intelligence in an agent does in effect help in achieving the cost/deadline objective. The Grid Economics Simulator is a Java-based discrete event simulator designed for the evaluation of economics based resource management algorithms. For details we refer to (Stuer et al. 2007), (Vanmechelen et al. 2008).
The parameters for the GES simulator runs are given in Table 1 and for the agents and jobs in Table 2. Values indicated by brackets indicate a random drawing of a value in that range.
The simulation runs for 500 steps and consist of 150 agents and 75 providers that are willing to participate in the auctions. The providers offers a number of resources as one package which at least contains 5 CPUs

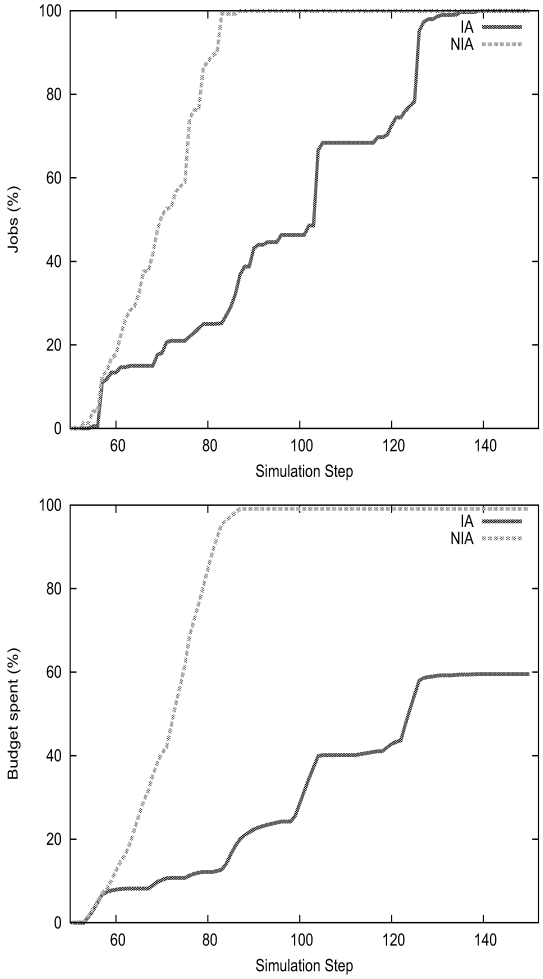


Figure 3: Finished jobs and budget spent per time step for both agent types using uniform policy environment.

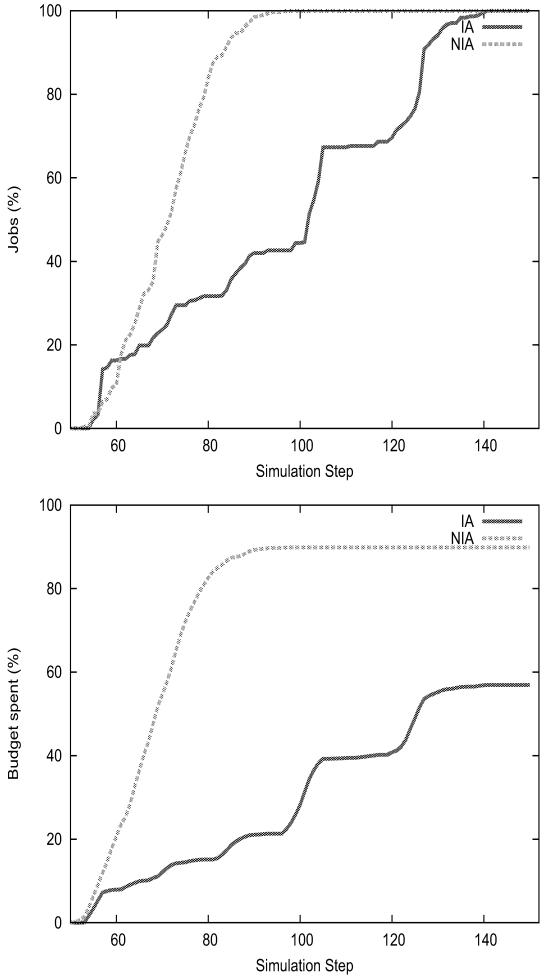


Figure 4: Finished jobs and budget spent per time step for both agent types using random policy.

and maximum 20 CPUs. Each agents starts with a random number of jobs between 1000 and 5000 and with a random running time in job list. The agents are divide into three equal-sized groups that have a small, medium and high budget allowance.
The intelligent agent and non-intelligent agent enter the market at the same time. They are initialized with the same specifications e.g the base allowance budget (4000000), number of jobs (1500) and the deadline time (step 150) see Table 2.

RESULTS

The models described in the previous sections have been executed with variations of the simulation parameters such as number of agents or providers, number of jobs and so on. The results reported here are representative with respect to these variations.
In this section we will look at the effect of the bidding policies on the performance of intelligent and non-intelligent agent with respect to **average budget**, the average over all jobs of budget spent per job for each agent type (see Figure 2) **finished jobs** , the percentage average of finished job per agent type (see Figures 3 and 4), **average spent per time step** , the average budget spent per time step for the execution of jobs (see Figures 3 and 4). The latter figures cover the time span from the start of the market until the final job deadline.

Figure 2 show that the intelligent agent saves more budget than non-intelligent agents with either policy. Figure 3 clearly reflects that the non-intelligent agent has its his jobs finished earlier, way before deadline, and spends more budget while the intelligent agent takes its time to finish jobs and spends considerably less. The figures also show that the non-intelligent agent requires more time to finish its jobs with the random policy.

These results indicate that the intelligence embedded in the consumer agent does indeed assist in achieving the goal of finished jobs within deadline at the lowest cost.

CONCLUSION

We have investigated a computational market with a multi-unit auction and agents acting on behalf of resource consumers. We have used simulations to investigate whether an agent with intelligence would act in a manner that enables the user to better achieve the goal of finishing all jobs within deadline for lowest cost. The results point to the fact that the price prediction based on neural network together with fuzzy techniques in defining bid levels achieve that objective. The investigations of course need to be extended to include the interaction of multiple intelligent agents.

REFERENCES

Arfa R. and Broeckhove J., 2008a. *Computational Resource Market Price Prediction.* In *Proceedings of 4th International Conference on Information Technology & Multimedia at UNITEN (ICIMU2008).* CoIT, Bangi, Malaysia, 39.

Arfa R. and Broeckhove J., 2008b. *Modeling Prices in Computational Resource Markets.* In *Proceedings of the 9th Middle Eastren Simulation Multiconference (MESM'2008).* eurosis, Amman, Jordan, 69–73.

Arfa R. and Broeckhove J., 2009. *Modeling resource prices in computational commodity markets.* In *Proceedings of the 7th ACM/IEEE International Conference on Computer Systems and Applications (AICCSA09).* IEEE Computer Society Press, Rabat, Morocco, vol. 0, 145–152.

Ausubel L.M., 2006. *An Efficient Dynamic Auction for Heterogeneous Commodities,. American Enonomic Review*, 96, no. 3, 602–629.

Buyya R.; Abramson D.; and Giddy J., 2002. *Economic models for resource management and scheduling in Grid computing. Concurrency and Computation: Practice and Experience*, 14, 1507–1542.

He M.; Jennings N.R.; and Prugel-Bennett A., 2004. *An adaptive bidding agent for multiple English auctions: a neuro-fuzzy approach.* In *IEEE Conference on Fuzzy Systems.* 1519–1524.

Krauter K.; Buyya R.; and Maheswaran M., 2002. *A taxonomy and survey of grid resource management systems for distributed computing. Softw Pract Exper*, 32, 135 – 164.

Leyton-Brown K.; Shoman Y.; and Tennenholtz M., 2000. *An Algorithm for Multi-Unit Combinatorial Auctions.* In *Proceedings of the 17th National Conference on Artificial Intelligence and 12th Conference on Innovative Applications of Artificial Intelligence.* AAAI Press, MIT Press, 56 – 61.

Lim D.; Anthony P.; and Ho C.M., 2008. *Agent for Predicting Online Auction Closing Price in a Simulated Auction Environment.* In *Proceedings of the 10th Pacific Rim International Conference on Artificial Intelligence: Trends in Artificial Intelligence.* Springer-Verlag, Berlin, Heidelberg, PRICAI '08.

Lin C.S.; Chou S.; Chen C.H.; Ho T.R.; and Hsieh Y.C., 2006. *A Final Price Prediction Model for Online English Auctions - A Neuro Fuzzy Approach.* In *JCIS'06.* –1–1.

P. W.M.; Anna O.; K. M.M.J.; and Reeves D., 2008. *Bidding Strategies for Simultaneous Ascending Auctions. The BE Journal of Theoretical Economics*, 8, no. 1, 27.

Stuer G.; Vanmechelen K.; and Broeckhove J., 2007. *A commodity market algorithm for pricing substitutable Grid resources. Future Generation Computer Systems*, 23, no. 5, 688–701.

Vanmechelen K.; Depoorter W.; and Broeckhove J., 2008. *A Simulation Framework for Studying Economic Resource Management in Grids.* In *Proceedings of the International Conference on Computational Science (ICCS 2008).* Springer-Verlag, Berlin Heidelberg, vol. 5101, 226–235.

Vanmechelen K.; Stuer G.; and Broeckhove J., 2006. *Pricing Substitutable Grid Resources using Commodity Market Models.* In H. Lee and S. Miller (Eds.), *Proceedings of the 3rd Int. Workshop on Grid Economics and Business Models (GECON 2006).* World Scientific, Singapore, 103–112.

Yao Y.Y., 1998. *A comparative study of fuzzy sets and rough sets. Information Sciences*, 109, 21–47.

AUTHOR BIOGRAPHIES

Ramadan Arfa is PhD researcher in the Computational Modeling and Programming (CoMP) group of the University of Antwerp, Belgium. He completed his Master degree in 1999 at University of Krakow (AGH), Poland. His research interests include distributed computing, monitoring and scheduling, artificial intelligence, grid computing and grid economics.

Dr. Peter Hellinckx is senior researcher at the Tera Labs of the Karel de Grote University College of Antwerp, Belgium. He obtained his Master in Computer Science and his Ph.D. in Computer Science both from the University of Antwerp in 2002 and 2008 respectively. He then joined the Tera Labs and became responsible for the data communication and distributed computing group.

Dr. Jan Broeckhove is a professor in the Department of Mathematics and Computer Science at the University of Antwerp, Belgium, and head of its Computational Modeling and Programming (CoMP) group. He received his PhD. in Physics in 1982 from the Free University of Brussels, Belgium. His current research interests include distributed computing, in particular grid computing.

ELECTRONICS SIMULATION

PHASE QUANTIZATION EFFECTS ON THE RADIATION OF MEMS BASED RECONFIGURABLE REFLECT ARRAYS

Hassan M.R.Salti
Department of Information Technologies and Computing
Arab Open University
P.O. Box 830, 92400, Al Ardia, Kuwait
E-mail: Hassan.salti@aou.edu.kw

Erwan Fourn and Raphael Gillard
Institute of Electronics and Telecommunications of Rennes IETR
INSA of Rennes
20, Av. des Buttes de Coesmes, CS 14315, 35043 Rennes Cedex, France
E-mail: Erwan.Fourn@insa-rennes.fr

Hervé Legay
Thales Alenia Space
26, Av. Jean François Champollion,
31037 Toulouse Cedex 1, France
Herve.Legay@thalesaleniaspace.com

KEYWORDS
Reflectarray, MEMS, Reconfigurable antenna, satellite communication, quantization.

ABSTRACT

Nowadays, satellite communication is one of the most important technologies. It is used in various applications such as navigation, military applications, scientific missions, etc. One essential device present in every satellite is the antenna. Without it, the satellite is unable to communicate with earth stations and is, hence, useless. Reflectarrays are one of the most promising antennas used in satellite communications. They have several advantages such as the ease of fabrication and the relatively low weight, volume and cost. Recently, many researches have been carried out to add Micro-Electro-Mechanical Systems (MEMS) switches into reflectarrays in order to control their radiation patterns. However, the presence of switches in such antennas means that a digital control is needed. As a consequence, the phase law that controls the radiation pattern becomes quantified. Phase errors are introduced on the radiation aperture which could create severe degradation on the radiation pattern. The main goal of this paper is to estimate the effects of uniform and non-uniform phase quantization errors in MEMS based reconfigurable reflectarrays. For this purpose, several tools have been designed and a large number of simulations are carried out.

INTRODUCTION

Satellite communications are one of the most important technologies of nowadays. Satellites are classified in many different ways: by function (communication, navigation, military, scientific, etc.), type of orbit (synchronous, asynchronous), cost, size, and so forth. However, and whatever its classification is, an artificial satellite is useless unless it can communicate with earth stations on the ground. All satellites are therefore equipped with a radio communication subsystem and accordingly, an essential device that plays the role of interface between the different satellite subsystems, the space channel and earth stations, which is the antenna.

Two wide antenna categories are used in satellite communication: reflector antennas (Balanis 1982) and phased array antenna (Hansen 1998). A reflector antenna consists of a reflecting metallic surface illuminated by a feeding antenna. The radiation pattern of the antenna is based on the geometric properties of the reflecting surface. Technically, the shape of the area to be covered by the satellite on the earth determines the antenna's radiation pattern. Then, geometric optics and physical optics techniques are used to synthesize the geometry of the corresponding reflecting surface (Bergmann et Al. 1988), (De Padova 1978). Nevertheless, due to the nature and the geometry of its reflecting surface, this category of antennas suffers a cumbersome weight and volume, especially when high gain is required. Also, for shaped beams, complex reflector geometries are required which involve a time-consuming and high-cost fabrication process.

The second antenna candidate for satellite communications is phased array antenna. It consists of an array of small printed radiating elements where each element is fed throughout a distinct feeding line (Hansen 1998). Thereby, the phase and the amplitude of the wave radiated by each element can be controlled independently. The phase/amplitude law created on the array surface defines the antenna's radiation pattern. This category has several advantages such as reducing the weight and volume of satellite antennas; however it still suffers from a high complexity and cost of fabrication. The feeding network is also responsible for high losses, especially at high frequencies. This calls for a third category of satellite antennas: the reflectarray.

Reflectarray antenna consists of a feeding antenna illuminating a planar microstrip array (Huang and Encinar 2007). It combines the advantages of both reflector and phased array antennas. Indeed, the presence of a primary antenna as a feeding source cancels the need for a complex feeding network; also, as will be described later, the radiation pattern of such type of antennas is controlled by the phase shift at each element in the aperture (Huang and Encinar 2007), (Pozar et Al. 1997). There is no need for

additional phase shifters behind these elements; each cell creates its own phase shift. A reflectarray is, hence, easier to realize than phased array antennas. Furthermore, the radiation aperture is realized in printed technology which reduces considerably the weight and volume of the antenna (Cadoret et Al. 2005), (Shaker and Cuhaci 1999), (Targonski and Pozar 1994).

Nowadays, reflectarrays face new challenges. In fact, the telecommunication systems are still rapidly evolving in order to meet the greedy needs of the users and their societies who ask for new high quality services, accessible anytime anywhere and that are still in their budgets. For instance, a "living" satellite (offering a given bandwidth, data rate, gain, etc.) that serves an area/country(ies) today and satisfies the needs of its/their society(ies) becomes useless for it/them tomorrow if it does not respond to their needs anymore.

This fact pushes reflectarray designers to create a new generation of intelligent satellite antennas that offer more flexibility to telecommunication systems. They intend to design low cost reconfigurable reflectarray antennas with an adaptive radiation pattern. By this way, for instance, the radiation pattern of the pre-evoked useless satellite antenna could be reconfigured to illuminate another area less demanding than the first greedy one (in terms of services).

Recently, many researches have been carried out to dynamically control the phase shift at each element in the array in order to obtain reconfigurable reflectarrays (Rajagopalan et Al. 2007), (Hu et Al. 2006). MEMS based phase-shifting cells are key elements for such applications (Montoria et Al. 2011), (Carrier-Peruisseau 2008), due to the excellent radio-frequency (RF) properties of MEMS technology such as: low insertion loss, low power consummation, high isolation, etc. MEMS are usually used as switches to control the physical length of resonating elements such as stubs or slots (Legay et Al. 2005), (Papapolymerou et Al. 2003).

MEMS based reconfigurable reflectarrays combine the advantages of both passive reflectarrays and RF-MEMS switches. They are considered as a promising antenna solution for future satellite communication (Vendier et Al. 2003).

However, this type of antennas is still facing some challenges such as the low level of reliability of RF-MEMS switches (Vendier et Al. 2003). MEMS breakdowns can be responsible for severe phase errors on the radiation aperture resulting in large deformations in the radiation patterns. Even so, it has been shown in (Salti et Al. 2010) that increasing the number of MEMS switches in the phase shifting cell creates a given degree of phase redundancy and makes this cell more robust to MEMS failure. This issue has been well discussed in (Salti et Al. 2010) and is hence behind the scope of this paper.

Here, MEMS switches are not supposed to fail, but they still have only two states "ON" or "OFF". The number of MEMS configurations in the phase shifting cell is limited and so is the number of phase shifts. A quantization error that depends on the phase shifts repartition over the range 0°-360° is therefore introduced on the radiation aperture. This error could severely affect the radiation pattern of the reconfigurable reflectarray and should, hence, be assessed.

The main objective of this paper is to study the effects of uniform and non-uniform phase quantization errors in MEMS based reconfigurable reflectarray. It is organized as follows:

At first, MEMS based reconfigurable reflectarrays are introduced and the key element behind phase quantization errors is defined, the "Cell's Characteristic" (CC). To better understand what a cell's characteristic is, a realistic MEMS based phase shifting cell called "Pharmacist Cross", firstly introduced in (Salti et Al. 2010), is recalled. The behavior of this cell, and particularly, the repartition of its phase shifts over the range 0°-360° (i.e. its CC), will serve us as a test-case example for the study presented in the remaining part of this paper.

Secondly, two types of phase quantization are addressed: uniform and non-uniform. The algorithms and tools, used to assess their effects on the radiation pattern of reconfigurable reflectarrays, are presented.

Finally, the results are analyzed, a comparing study is performed and a conclusion is drawn.

MEMS BASED REFLECTARRAYS

Generalities

Figure 1 illustrates how a reflectarray works: the wave, issued from the feeding source antenna to an array element "*i*", arrives with a given phase $\emptyset_i^{inc}$. The array element *i* absorbs the wave, introduces a phase shift $\emptyset_i^{intr}$ and re-radiates the wave with a new phase: $\emptyset_i^{refl} = \emptyset_i^{inc} + \emptyset_i^{intr}$. The group of reflected phases issued from the different array elements is called "Phase law". In fact, it is this phase law that defines the antenna's radiation pattern.

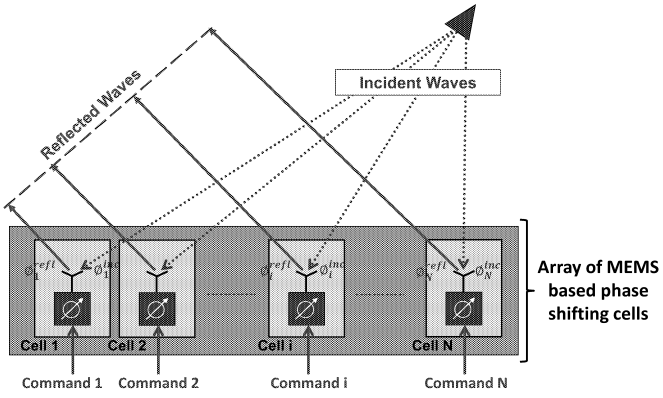


Figure 1: Schematic Representation for MEMS Based Reconfigurable Reflectarrays.

In passive reflectarrays, once the reflectarray has been designed, the phase law and the radiation pattern are unchangeable.

In reconfigurable reflectarrays, the design of the array only fixes the phase of the incident field $\emptyset_i^{inc}$ on any given cell *i* (the wave always travels throughout the same path to reach the cell *i*). So the reflected wave's phase $\emptyset_i^{refl}$ is directly proportional to the phase $\emptyset_i^{intr}$ introduced by the cell *i*. If this cell *i* is capable of introducing any phase shift $\emptyset_i^{intr}$ in the range 0°-360°, at any time and whatever is its position in the array (whatever is the value *i*), any phase law on the

radiation aperture can be created and the reflectarray is considered as reconfigurable.
In MEMS based reconfigurable reflectarrays, the same phase shifting cell is reproduced all over the array to create the radiating aperture (Vendier et Al. 2003). In this cell, RF-MEMS switches are inserted in order to dynamically control the phase shift $\emptyset^{intr}$ (Rajagopalan et Al. 2007).
Since each MEMS switch has only two possible states: ON or OFF, the total number of MEMS configurations is given by 2^M where "M" is the number of MEMS in the phase shifting cell. Each MEMS configuration produces only one phase shift. So the total number of phase shifts that a cell can introduce is also equal to 2^M. The table relating phase shifts to MEMS configurations is called Cell's Characteristic (CC). As we will see later, this CC is the key element behind phase quantization errors.

Typical Example of MEMS Based Phase Shifting Cell

The active cell used here as typical example is called "Pharmacist Cross". It has been firstly introduced in (Salti et Al. 2012). It is obtained by associating several concentric cross loops as in (Chaharmir et Al. 2006). The patch element is implemented on a quartz glass (ε_r=3.78, h=0.5mm) that is separated from ground plane by an air cavity with metallic boundaries (cf. Figure 2.a). Eight MEMS are distributed among the three different slots between the crosses. By changing the states of these MEMS, the electrical lengths of the different slots change and different phase shifts are obtained. The MEMS are capacitive switches having C_{OFF}=20fF (when the switch is up) and C_{ON}=400fF (when the switch is down).
In (Salti et Al. 2010), a didactic method has been applied to optimize the repartition of the available 8 MEMS switches in the cell while maximizing its robustness to MEMS failures at the central frequency f_0=12GHz. This also leads to a linear CC (cf. Figure 2.b). In this paper, this linear CC will serve us as a typical example, when uniform phase quantization errors are addressed.
On the other hand, the authors show in the same paper that the robustness and the CC's linearity are not the only issues to be optimized in MEMS based reflectarray antennas. Bandwidth consideration is also a major issue. As described in (Salti et Al. 2010), some resonant MEMS configurations should be discarded from the original pharmacist cross' linear CC in order to largely increase the reflectarray's bandwidth (the 1dB gain bandwidth passes from 3 to 14%). The resulting CC is non-linear and contains an 80° phase gap around 180° (cf. Figure 2.c). In this paper, this non-linear CC will serve us as a typical example, when non-uniform phase quantization errors are addressed.

PHASE QUANTIZATION ERRORS EFFECTS

Phase Quantization Errors

Suppose that we have an arbitrary phase shifting cell containing M MEMS and producing 2^M different configurations/phase shifts. Let $\emptyset_{th}^{intr}$ be the theoretical phase shift that should be introduced by this cell in order to obtain a theoretical reflected wave with a phase of $\emptyset_{th}^{refl}$. Due to the discrete CC, this phase $\emptyset_{th}^{intr}$ is replaced by a quantified one $\emptyset_{q}^{intr}$ which is the nearest one to $\emptyset_{th}^{intr}$ in the CC.

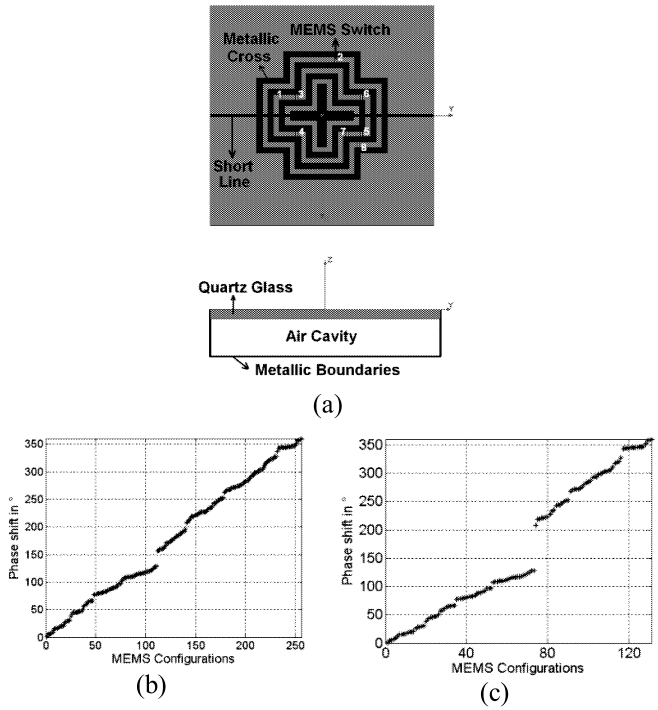


Figure 2: Pharmacist Cross Cell (Salti et Al. 2010): (a) Optimized MEMS Positions, (b) Linear CC, (c) Non-Linear CC.

The cell re-radiates the wave with a phase $\emptyset_{q}^{Refl} = \emptyset^{inc} + \emptyset_{q}^{intr}$. Since the phase of the incident wave on that cell $\emptyset^{inc}$ is constant, the quantization phase error $\Delta\emptyset$ is given by the Equation (1) below.

$$\Delta\emptyset = \emptyset_{th}^{Refl} - \emptyset_{q}^{Refl} = \emptyset_{th}^{intr} - \emptyset_{q}^{intr} \qquad (1)$$

If the CC is linear: for instance, the CC of the optimized pharmacist cross cell (cf. Figure 2.b), the phase quantization is said to be "uniform quantization". In this particular case, the quantization step β is a constant and is directly related to the number of MEMS M by: $\beta=2\pi/2^M$. Figure 3, shows how a given theoretical phase shift $\emptyset_{th}^{intr}$ is quantified. It also shows the corresponding quantization error $\Delta\emptyset$. As we can see in Figure 3.a, if the CC is linear, the phase error $\Delta\emptyset$ uniformly varies between [-β/2 ; β/2] when $\emptyset_{th}^{intr}$ varies between 0° and 360°.
Hence, in reconfigurable reflectarrays, uniform phase quantization creates phase errors on the radiating aperture that are limited between -β/2 and β/2. Reducing β reduces phase errors. This can be achieved by simply adding more MEMS in the phase shifting cell.
On the other hand, if the CC is not linear: for instance, the CC of the optimized pharmacist cross cell but after eliminating dispersive configuration (cf. Figure 2.c); the phase quantization is said to be "non-uniform quantization". As we can see in Figure 3.b, when the quantization is non-uniform, the quantization step β is no more a constant and the phase quantization error $\Delta\emptyset$ varies non-uniformly between $-\beta_{max}/2$ and $\beta_{max}/2$; where β_{max} is the maximum value of β (maximum quantization step). Hence, unlike

uniform quantization, and whatever the number of MEMS in the cell is, large phase errors are possible in a non-uniform quantization. Those are directly related to the larger quantization step; or in other words, directly related to the non-linearity of the CC.

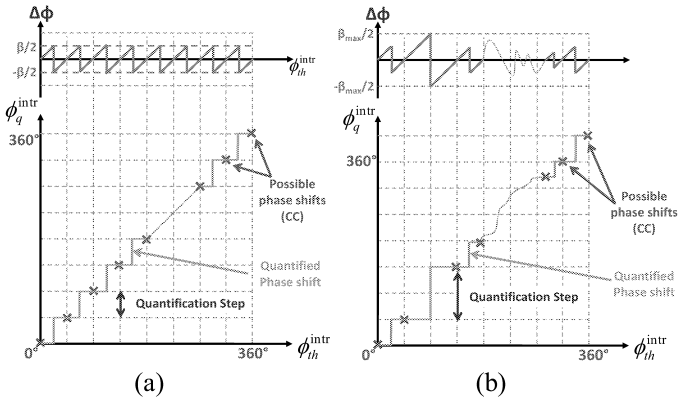


Figure 3: Quantization Error:
(a) Uniform (Linear CC), (b) Non-Uniform (Non-Linear CC)

Test Case Reflectarrays

In order to assess the effects of phase quantization errors, six circular test case reflectarrays are defined (cf. Figure 4). Those differ by their diameter "D" and their inter-element spacing dx = dy = d (the distance between two successive cells). Table 1 summarizes the number of cells in each array. Note that λ_0 is the wave length at the central frequency f_0=12GHz.

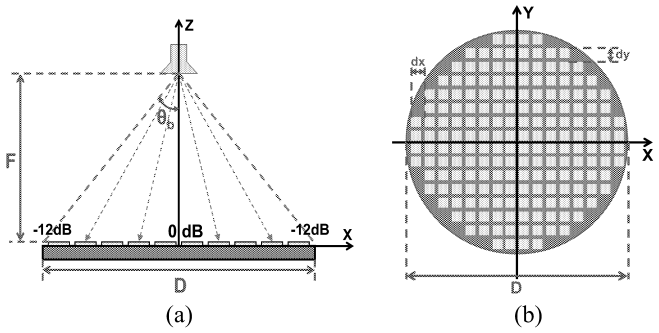


Figure 4: Test Case Reflectarrays: Typical Scheme.

Table 1: Test-Case Reflectarrays: Number of Cells.

D	d	Number of cells on X and Y axis	Total number of cells
$40\lambda_0$	$0.3\lambda_0$	133	13764
	$0.5\lambda_0$	80	4872
	$0.7\lambda_0$	57	2480
$20\lambda_0$	$0.3\lambda_0$	66	3388
	$0.5\lambda_0$	40	1184
	$0.7\lambda_0$	28	608

The feeding source is supposed to be a horn antenna. It is fixed on the central axis of the array, at a distance "F" from this latter and gives -12dB tapering at its outline (outline field's amplitude/central field's amplitude). The ratio F/D is fixed to 0.8 in order to maximize the efficiency of the reflectarray (Pozar et Al. 1997).

For simplicity, the phase law of the reflected waves is calculated at f_0=12GHz for broadside radiation. The reflected phases $\emptyset_{th}^{Refl}$ of all cells are consequently being fixed to a constant.

Theoretical Vs Real Radiation Pattern

In our generalized study, the elementary cell topology is considered as unknown. The radiation pattern at the central frequency f_0 is also approximated to the array's factor $F(\theta,\varphi)$ given by the Equation 2. Where:

- N is the number of cells in the array.
- A_i and A_r are the reflected electrical field's amplitude on, respectively, the current cell *i* and the reference cell *r*. In our study, the reference cell is the nearest one to the center of the array.
- k_0 is the wave number given by $k_0 = 2\pi/\lambda_0$.
- x_i and y_i define the position of the cell *i* 's center.
- θ and φ are the spherical angles, supposing that the axis origin is fixed on the center of the array.

$$F(\theta,\varphi) = \sum_{i=1}^{N} \frac{A_i exp(j\emptyset_i^{Refl})}{A_r exp\,(j\emptyset_r^{Refl})} exp[jk_0(x_i cos\varphi sin\theta + y_i sin\varphi sin\theta)] \quad (2)$$

In addition, $\emptyset_i^{Refl}$ and $\emptyset_r^{Refl}$ are the phases of the reflected waves issued from cells *i* and *r* respectively. As described previously, when calculating the theoretical radiation pattern, those are replaced by the theoretical re-radiated phases $\emptyset_{i\,(th)}^{Refl}$ and $\emptyset_{r\,(th)}^{Refl}$. Thought, in reality, those are replaced with the quantified phases $\emptyset_{i\,(q)}^{Refl}$ and $\emptyset_{r\,(q)}^{Refl}$.

Uniform Phase Quantization Effects

As detailed previously, a uniform quantization is linked to a linear CC. If a cell contains M MEMS, its CC is a uniform repartition of 2^M different phases in the range [0°; 360°]. The quantization step β is, hence, given by: $\beta=360°/2^M$.
This classical case is that of phased array antennas where phase quantization errors result from the use of digital phase shifters (having only 2^{Nb} phases where Nb is the number of control bits in the phase shifter). It has been shown that the uniform phase quantization decreases the theoretical gain G_0 of the antenna. This gain decreases ΔG is directly related to the quantization step β by the Equation 3 (Hansen 1998). The theoretical gain decrease ΔG is drawn in Figure 6 - Theory.

$$\frac{G}{G_0} = \Delta G = \mathrm{sinc}^2(\beta/2) = \mathrm{sinc}^2\left(\frac{\pi}{2^{N_b}}\right) \quad (3)$$

In order to compare with this theoretical curve, the radiation pattern's equation (i.e. Equation 2) and the pre-defined test case reflectarrays have been implemented in Matlab. A quantified phase law is applied on the six test-case reflectarrays instead of the constant theoretical one and the modified radiation pattern $F^q(\theta,\varphi)$ is calculated. This pattern is then compared to the theoretical one $F^{th}(\theta,\varphi)$ (when a theoretical constant phase law is applied). For instance, Figure 5 shows a comparison between the

theoretical and the quantified radiation pattern of one test-case reflectarray (M=3, D=40λ_0 and d=0.5λ_0). This typical example shows us how phase quantization affects the radiation pattern of the antenna. For instance, it decreases the gain of the antenna and increases the level of side lobes.

In our study, side lobes are omitted. At the final stage, only the gain attenuation is calculated from: $\Delta G = \left|F^{th}(0,0) - F^{q}(0,0)\right|$, where $F^{th}(0,0)$ and $F^{q}(0,0)$ are the theoretical and quantified array factors at broadside.

This procedure is repeated for several numbers of MEMS in the cell, and a curve representing the gain attenuation versus the number of MEMS M is plotted. As we can see in Figure 6, the results are quite satisfactory. The simulated gain attenuation is identical to the theoretical one. It is also independent from the inter-element spacing d, especially for large reflectarrays.

Another important issue is that the effects of a uniform phase quantization can be considered as negligible when the number of MEMS in the cell is more than 4 ($\Delta G < 0.014$dB). In other words, though the phase error decreases when adding more MEMS in the phase shifting cell, the quantified gain of the antenna does not significantly improve if more than 4 MEMS are involved.

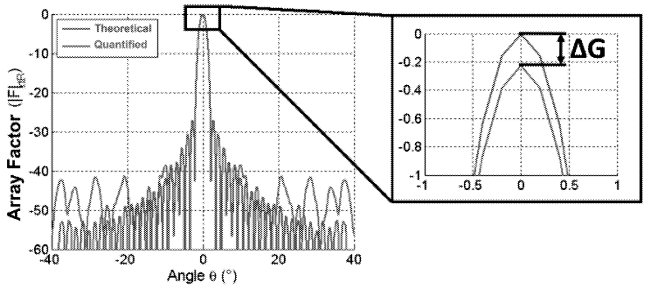


Figure 5: Comparison Between a Theoretical and a Quantified Array Factor
(Test-Case Reflectarray where: M=3, D=40λ_0 and d=0.5λ_0)

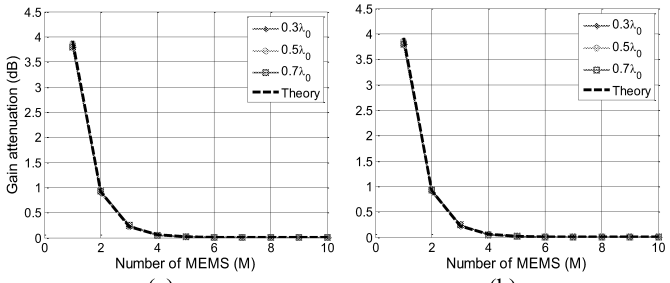


Figure 6: Effects of Uniform Phase Quantization on Different Test-Case Reflectarrays: (a) D=40λ_0; (b) D=20λ_0.

Non-Uniform Phase Quantization Effects

Unlike phased array antennas, the phase quantization is not always uniform in MEMS based reconfigurable reflectarrays. The phase quantization type depends on the linearity of the CC and, consequently, on the cell's topology and the distribution of MEMS in it. For instance, the pharmacist cross cell represents different CCs when the distribution of MEMS changes (Salti et Al. 2010).

Many different non-linear CCs are possible, it is hence impossible to treat them all. Therefore, in this paper, we stick with a realistic case where a non-linear CC results from removing dispersive configurations from an original CC in order to increase the bandwidth of the reflectarray (Salti et Al. 2010). The non-linear CCs studied here are CCs that contain a phase gap. In other words, the cell is unable to introduce an interval of phases and the resulting CC has an abrupt phase discontinuity. For this phase gap, we define two parameters: its length "L_G" and its position in the CC "P_G". Concretely, this means that the cell is unable to introduce phase shifts in the interval [P_G-L_G/2 ; P_G+L_G/2]. All the possible 2^M phases are uniformly distributed outside this phase gap. Hence, the resulting maximum quantization step β_{max} is equal to L_G and the maximum phase error is $\beta_{max}/2 = L_G/2$. A typical example is shown in Figure 7.a where M=5, L_G=90°, P_G=180°.

The algorithm used for uniform quantization is here repeated in order to calculate the gain attenuation due to this non-uniform phase quantization. However, and since the gain attenuation is shown to be independent from the inter-element spacing d, only one test-case reflectarray is considered (D=40λ_0 and d=0.5λ_0). On the other hand, here, three parameters affect the non-linear CC (instead of one):

- The phase gap length L_G: is supposed to vary between 0° and 200°.
- The phase gap position P_G: is supposed to have only four values, 0°, 90°, 180°, 270°.
- The number of MEMS M: varies between 1 and 7.

Figure 7.b shows the gain attenuation curves versus the length of the phase gap L_G for several numbers of MEMS M in the cell. Note that only one position P_G=180° is shown here since the same curve is obtained for all other values. This fact is due to the equi-distribution of phase shifts to be realized on this relatively large test-case reflectarray.

Furthermore, the first value of each curve (for each value M) corresponds to the lowest value that L_G can have. In practice, this value relates to the uniform quantization step β (since a gap in a given CC cannot be lower than $\beta=360°/2^M$). Hence, the gain decrease at that initial point is the one obtained when a uniform quantization is applied. Therefore, linking the first point of the different curves (dark black dashed curve in Figure 7.b) leads us to the gain decrease curve previously drawn in uniform quantization section (cf. Figure 6) but, this time, as a function of the uniform quantization step β instead of the number of MEMS M.

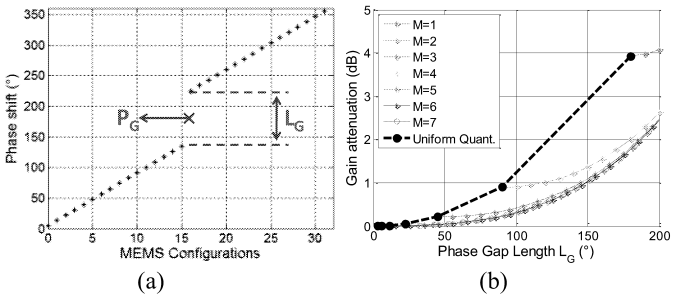


Figure 7: Non-Uniform Quantization: (a) Typical Example of a Non-Linear CC (M=5, L_G=90°, P_G=180°), (b) Gain Attenuation (P_G=180°).

Now, if we want to compare uniform and non-uniform quantization effects, we should fix in mind that the studied cell must contain the same number of MEMS with two different types of CC: one linear and one non-linear. This means that each curve on Figure 7 should be treated and analyzed independently (for a fixed value of M). For instance, if a cell contains M=2 MEMS, the least gain attenuation of 0.9dB is obtained when L_G = 90° which corresponds to a uniform quantization (since $L_G = \beta = 360/2^2 = 90°$); this gain attenuation increases when the non-linearity of the CC becomes more important (i.e. when the phase gap length L_G increases) which corresponds to a non-uniform quantization.
Therefore, analyzing Figure 7 shows that a non-uniform quantization is much more critical than uniform one whatever the number of MEMS in the phase shifting cell is and even for the smallest possible values of the phase gap's length L_G.

Uniform VS Non-Uniform Quantization

For a clearer comparison between uniform and non-uniform quantization, the notion of equivalent MEMS number is here defined. By definition, if a non-uniform quantization introduces a gain attenuation ΔG, the equivalent MEMS number M^{eq} corresponds to the number of MEMS that gives the same gain attenuation ΔG when a uniform quantization is applied. Hence, this ΔG is replaced in Equation 3 and M^{eq} is deduced. However, ΔG is a function of $\text{sinc}^2(\beta/2)$ which is analytically non reversible. Therefore, a graphical method is established.
This method consists in firstly modifying the Equation 3 to a new one (i.e. Equation 4 below).

$$\frac{\beta}{2}\sqrt{\Delta G} = \sin\left(\frac{\beta}{2}\right) \qquad \text{(Eq. 4)}$$

Then this equation is dissociated into two functions f_1 and f_2 given by: $f_1(\beta/2\,;\Delta G) = \frac{\beta}{2}\sqrt{\Delta G}$ and $f_2(\beta/2) = \sin\left(\frac{\beta}{2}\right)$.
For a given value ΔG, the intersection between these two functions bring us back to the uniform quantization step β. Finally, the equivalent MEMS number M^{eq} is deduced from the equation $M^{eq}=\log_2(2\pi/\beta)$.
This method is applied on the predefined non-linear CC with a variable phase gap's length L_G and for different initial MEMS number M. The position of the gap P_G is fixed to 180°.
Figure 8 summarizes the results. We can see clearly the negative effects of a phase gap in the CC. Indeed, for a given number of MEMS M, the equivalent number of MEMS M^{eq} decreases rapidly when L_G increases. For instance, if a cell contains M=6 MEMS and its CC presents a gap of 35°, it will affect the gain of the reflectarray in the same manner as a 5 MEMS' cell that has a linear CC.
Nevertheless, if the number of MEMS in the cell M is more than 4, and the CC presents a gap less than 60°, the equivalent number of MEMS M^{eq} is always more than 4. Going back to the results presented in the previous section (uniform quantification), under these conditions, this means that the gain attenuation can be considered as negligible.
Note finally, that the thick dashed black curve of Figure 8 relates the first initial point of each curve (when L_G=β); it is, hence, simply representing the real number of MEMS in a cell M as a function of the uniform quantization step β.

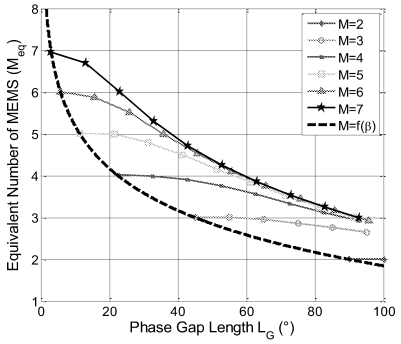


Figure 8: Equivalent Number of MEMS of a Phase Gapped Non-Linear CC.

CONCLUSION

Though MEMS based Reconfigurable reflectarray is considered as a promising antenna solution for future satellite communication, it still confronts several challenges such as the low level of reliability of RF-MEMS switches. One major challenge for reflectarray designers is to create a MEMS based phase shifting cell that is capable of reducing the effects of MEMS failures, increasing the bandwidth of the antenna and reducing its fabrication complexity and cost.
One way to insure robustness to MEMS failures is to increase the number of MEMS switches in the phase shifting cell. However, this increases the chances to have dispersive MEMS configurations that can be responsible for a severe bandwidth limitation and should, hence, be discarded. Discarding configurations means that some phase shifts cannot be introduced by the cell anymore. As a consequence, the repartition of phase shifts over the range 0°-360° presents a phase gap.
Therefore, unlike phased array antennas, the phase quantization is not always uniform in MEMS based reconfigurable reflectarrays. Even with a high number of MEMS in the cell, phase quantization can be responsible of large phase errors distributed over the radiating aperture and, as a consequence, can significantly deteriorate the radiation pattern.
In this paper, several tools have been presented in order to assess the effects of phase quantization on the radiation pattern of a reflectarray.
It has been shown that phase quantization errors depend on the phase-shifting cell and more specifically to the available phase-shifts distribution over the range 0°-360°. If this distribution is linear, the phase quantization is uniform. The effects of such quantization can be easily reduced by increasing the MEMS number M in the phase shifting cell. Nevertheless, the quantified gain of the antenna does not significantly improve if more than 4 MEMS are involved.
On the other hand, if the phase distribution presents a phase gap, the phase quantization is no longer uniform and the pattern deterioration increases considerably with the gap's length. Therefore, it has been shown that, to achieve an acceptable deterioration of the radiation pattern, it is recommended to increase the number of MEMS in the cell while trying to reduce the phase gap to less than 60°.

REFERENCES

Balanis C. A. 1982. "Antenna Theory: Analysis and Design," Wiley, second ed.

Bergmann J., Brown R.C., Clarricoats P.J.B., Zhou H. , 1988. "Synthesis of Shaped-Beam Reflector Antenna Patterns," *IEE Proceedings on Microwaves, Antennas and Propagation*, vol. 35, no. 1, pp. 48-53.

Cadoret D., Laisne A., Gillard R., Legay H. 2005. "A New Reflectarray Cell Using Mmicrostrip Patches Loaded with Slots," *Microwave and Optical Technology Letters*, vol. 44, pp. 270-272.

Chaharmir M.R., et al. 2006. "Broadband Reflectarray Antenna with Double Cross Loops," *Electronic Letters*, vol. 2, pp. 65-66.

De Padova S., Pagana E., Rosenga G. 1978. "Analysis and Synthesis of an Elliptical Paraboloid Offset Reflector Antenna Elliptically Shaped Beam Antenna," *International Symposium on Antennas and Propagation Society*, vol. 16, pp. 381-384.

Hansen R. C. 1998. "Phased Array Antennas," Kay Chang.

Hu W., et al. 2006. "Tunable Liquid Crystal Reflectarray Patch Element," *Electronics Letters*, Vol. 42, Issue 9, pp. 509-511.

Huang J. and Encinar J. A. 2007. "Reflectarray Antennas," Wiley-Interscience series, pp.34-35.

Legay H., et al. 2005. "MEMS Controlled Phase-Shift Elements for a Linear Polarized Reflectarray," *28th ESA Antenna Technology Workshop on Space Antenna Systems Technologies*, Netherlands, pp. 443-448.

Montoria S., Chiuppesia E., Farinellia P., Marcacciolia L., Gattia R. V., Sorrentinoa R. 2011. "W-band beam-steerable MEMS-based reflectarray", *International Journal of Microwave and Wireless Technologies*, Volume3, Special Issue05, pp 521-532.

Papapolymerou J., Lange K. L., Goldsmith C.L., Malczewski A., Kleber J. 2003. "Reconfigurable Double-Stub Tuners Using MEMS Switches for Intelligent RF Front-Ends," *IEEE transactions on microwave theory and techniques*, vol. 51, no. 1, pp. 271-278.

Peruisseau-Carrier J. and Skrivervi A. K. 2008. "Monolithic MEMS-Based Reflectarray Cell Digitally Reconfigurable Over a 360° Phase Range," *IEEE Antennas and Wireless Propagation Letters*, Vol. 7, pp. 138-141.

Pozar D., Targonski S., and Syrigos H. 1997. "Design of Millimeter Wave Microstrip Reflectarrays," *IEEE trans. Antennas and Propagation*, vol. 45, no. 2, pp. 287-296.

Rajagopalan H., Rahmat-Samii Y., Imbriale W.A. 2007. "Reconfigurable Patch-Slot Reflectarray Elements Using RF MEMS Switches: A subreflector wavefront controller," *IEEE Antennas and Propagation Society International Symposium*, pp. 5023-5206.

Salti H., Fourn E., Gillard R. 2010. "Minimization of MEMS Breakdowns Effects on the Radiation of a MEMS Based Reconfigurable Reflectarray," *IEEE Transactions on Antennas and Propagation*, Vol. 58, No. 7, pp. 2281-2287.

Salti H., Fourn E., Gillard R., Girard E., Legay H. 2010. "Pharmacist Cross Phase-Shifting Cell Loaded with MEMS Switches for Reconfigurable Reflectarrays," *European Conference on Antennas and Propagation EUCAP*, Barcelona, Spain, pp. 1-4.

Shaker J. and Cuhaci M., 1999. "Planar Reflector for LMCS Applications," *Electronics Letters*, vol. 35, no. 2, pp. 103-104.

Targonski S.D. and Pozar D.M. 1994. "Analysis and Design of a Microstrip Reflectarray using Patches of Variable Size," *IEEE symposium on antenna and propagation digest*, vol.3, pp 1820-1823.

Vendier O., et al. 2003. "Need for Robust RF MEMS in Future Space Applications," *4rth MEMSWAVE*, Toulouse, France.

BIOGRAPHY

HASSAN SALTI was born in Tripoli, Lebanon, in December 1984. He received his PhD degree in Electronics from INSA of Rennes, France, in 2010.

After a postdoctoral research at the Laboratory for Electronics & Information Technology (CEA - LETI), Grenoble, France, he joined the Arab Open University – Kuwait Branch in October 2011. His research activities are mainly focused on the design of reconfigurable antennas for short, high range and satellite applications. He is also interested in Visual Light Communication, his new field of interest in his actual research group at the AOU.

Trust Based Reliable and Secure Routing in Peer-to-Peer Overlay Networks

Raja Waseem Anwar
Arab Open University
Muscat

Farah Abu Shahla
Arab Open University
Muscat

Saqib Ali
Sultan Qaboos University
Muscat

Sumitra Sirinvas
Arab Open University
Muscat

KEYWORDS
Peer-to-Peer, Trust, Security, Overlay Networks

ABSTRACT

Increasingly on the Internet, applications are supported by sets of loosely connected machines operating without any form of central coordination and more complex than classical distributed system. Peer-to-peer networking, a disruptive technology for large scale distributed applications, has gained widespread attention due to the successes of peer-to-peer (P2P) content sharing, media streaming, and telephony applications. Peer-to-peer systems are implemented using an abstract overlay network (An overlay network is a computer network which is built on the top of another network) which makes the P2P system independent from the physical network topology. But current overlays network are not secure; even a small segment of malicious nodes can avoid accurate message delivery throughout the overlay network. This problem is particularly serious in open peer-to-peer systems, where many diverse, sovereign parties without previous trust relationships wish to pool their resources.

This paper studies types, categories, characteristics and network applications of peer-to-peer (P2P) overlay networks. Also will review and analyze the type of peer-to-peer (P2P) current security attacks and then will present our Hybrid secure routing mechanism in peer- to- peer overlay networks to overcome the security challenges.

1 INTRODUCTION

Peer-to-peer (P2P) overlay networks are distributed systems in nature, without any hierarchical organization or centralized control. Peers form self-organizing overlay networks that are overplayed on the Internet Protocol (IP) networks, offering a mix of various features such as robust wide-area routing architecture, efficient search of data items, selection of nearby peers, redundant storage, permanence, hierarchical naming, trust and authentication, anonymity, massive scalability, and fault tolerance [**1**].

1.1 Categories of Peer-to-Peer overlay Networks

P2P systems can be classified into two types such as, **[2].**

- **Unstructured P2P**

In unstructured P2P system, files can be stored in any peer, that is, the file storage has no certain structure. Unstructured P2P system includes Centralized P2P, pure P2P and Hybrid P2P.

- **Structured P2P**

Structured P2P system maintains a link between file contents and IP addresses of the peers using Distributed Hash Table (DHT). Therefore, the whole P2P system holds a certain structure. Normally, structured P2P means DHT-base P2P.

1.2 Characteristics of Peer- to- Peer overlay networks

Peer- to- Peer overlay networks are dynamic and distributed systems. It has its own characteristics, which are described below,

- Peer-to-peer overlay systems go beyond services offered by client-server systems by having symmetry in roles where a client may also be a server **[3].**
- By having symmetry in roles where a client can also act as a server, they do not have a centralized control or fully responsible storage so that the system can have more robustness for any server node failure and for any congestion or bottleneck problem due to a server **[4].**
- They can also have a scalable storage by utilizing storages of each peer as a part of the whole p2p system's storage **[4].**
- The peer-to-peer networking mechanism enables all the nodes in the system can look up and consumes the shared content or files stored in other peers **[4].**
- It allows access to its resources by other systems and supports resource sharing, which requires fault-tolerance, self-organization and massive scalability properties **[3].**

1.3 Peer- to- Peer overlay networks applications

Peer-to-Peer (P2P) overlays have used popularly with the advent of file-sharing applications such as Napster, KaZaa and Bittorrent. In addition to file-sharing and content distribution, P2P networks are also being used for applications such as Voice over IP (VoIP) and television **[5].**

2 Background and Literature Review

Table 1, below presents the literature review and analysis on the literature carried out on Peer-to-Peer overlay networks with their areas of focus.

Article	Area of Focus	Discussion
[11] Xu Xiang et al.	Proposed a secure protocol for peer-to-peer networks, which consists of two phases. One phase is used for detecting malicious nodes and another is for bypassing them. To achieve this they have presented an efficient routing strategy called tracer routing to control routing progress, which is accompanied by a node-ID based signature scheme. Combining that two phases, the source of each query can verify each step and malicious nodes can be identified. They have also presented a scheme to create a secure path to bypass malicious nodes. By using their scheme, each query can successfully reach the destination even if there are a large number of malicious nodes in the system.	In this work authors describe how an efficient routing strategy in peer-to-peer network. Security issues are up to certain level, but nothing related to secure and reliable routing in peer-to-peer networks.
[12] Aina Ravoaja et al.	Presented a scalable protocol for efficient and secure information gathering in a system which high churn and collusive peers. Their proposed system is called as STORM. According to their model, efficiency comes from two ingredients: The first one is self-organization. By making peers to be adjacent to other peers sharing the same interest, they have increased the likelihood of finding relevant information within a small number of numbers of steps. The second one is Distributed Hash Tables (DHT)-based overlay.	In this work they proposed self-organization of peers to adjacent peers and use of Distributed Hash Tables but again no discussion related to trust and security.
[13] Prashant Dewan et al.	Have investigated the reputation systems for P2P networks. In addition, they have presented self-certification, an identity management mechanism, reputation model, and a cryptographic protocol, which facilitates generation of global reputation data in a P2P network, in order to expedite detection of rogues. They have used the identity mechanism based on the ranks of the peers are more relevant than the absolute value of their reputation. Their proposed Self-certification-based identity generation mechanism reduces the threat of liar farms by binding the network. m.	In this work, authors have discussed issues related to self-certification and identity management mechanism using reputation model. Moreover issues raised will work up to certain levels, lack of major security issues and no discussion related to Trust in peer-to-peer overlay network.
[14] *Joao Leitao et al.*	Proposed a multiplicative model, which is proposed to measure reliability of a P2P overlay routing substrate. Also, they have proposed an algorithm, which is based on multiplicative model that is used to send reliable route event notifications to client. Their algorithm, which leverages the Pastry routing algorithm, is used to determine a reliable path, greater than or equal to the threshold, between a pair of overlay broker nodes. They have employed a technique called 'pruning' by which they have restricted flooding the entire overlay routing substrate, when finding a reliable path.	Proposed an approach to measure reliability in peer-to-peer networks along with routing algorithm by employing "pruning" technique, but nothing is suggested or discussed related to Trust.
[16] *Wojciech Galuba et al.*	Proposed a Forward Feedback Protocol (FFP), which routes only a single copy of the message and detects the message loss and excessive delays while routing. According to their model, failures are signaled along the routing paths. Based on these binary signals, each overlay node locally and independently learns to route to avoid failures. Despite the independence of its	In this research the authors proposed a Forward Feedback Protocol which routes only a single copy of the message.

	nodes, the overlay as a whole converges on reliable loop-free routes. These local node interactions lead to the emergence of fast reliable overlay routes. This is a continuous process; the system constantly self-organizes in response to changing delay and loss conditions.	No such issues related to Trust have been discussed or highlighted.
[17] *Ruichuan Chen et al.*	Developed a novel poisoning resistant security framework for P2P content sharing service. Their framework is based on the fundamental fact that the content providers are the only sources to accurately distinguish poisoned contents and verify the integrity of the requested contents. Also, their work is built upon a distributed hash table (DHT) based overlay to provide the mechanisms of availability and scalability. According to their work, a content provider disseminates his shared contents' information to a group of content maintainers self-organized in the DHT-based overlay. This allows a content requestor to perform integrity verification on the requested content by first looking up in the DHT-based overlay for the associated content maintainers who have maintained the corresponding content information.	In this research authors have developed Security Framework for P2P content sharing service which is based on the fact that content providers are the only source, again lack of Trust based security and routing.

Based on literature review and analysis, this research supports the initial findings that most of the studies are based on target specific domains in peer-to-peer overlay networks, while lacking the trust based reliable and secure routing in peer-to-peer overlay networks is the key issue and challenge.

2.1 Security and Failure issues in Peer-to-Peer overlay networks:

The security issues are a major concern in a peer-to-peer environment because partners do not trust each other and the system is vulnerable to third party attacks. Possible attacks in P2P network are described below, **[2].**

- **Denial-of-service (DoS)**

In DoS attack, attackers utilize reasonable service requests to exhaust the resources of a target host. Therefore, the victim host cannot provide any service to other legal intended users.

- **TCP Syn Flooding Attack**

It may happen in the stage when the attacker is downloading files from the victim host. Here, the attacker uses a forged IP address to send a SYN request to the target host. It send continuous SYN request so that the victim host exhausts the resources and cannot perform service to the legal peers.

- **Query Flooding Attack**

The query flooding attack can also happen in application layer in P2P systems, such as Gnudella. A malicious node will constantly generate as much queries as possible to flood the network.

- **Poisoning Attacks**

Poisoning attacks can happen in the P2P networks. Attackers use false information, for example, false file indexes, false IP addresses, or false routing tables, to break the integrity of P2P systems.

- **Sybil Attack**

In P2P network, each entity has a unique identity. It can form a one-one mapping pair to actual entry. Most P2P networks use the virtual addressing scheme based on the logical identifiers to manage and organize the network. However, if the relation of one-one mapping of entity to identity is destroyed by malicious peer, in other word, a malicious entity acts as a number of multiple identities. The entity can control a significant part of networks. Such attack is defined as Sybil attack.

- **Routing attack**

Attacks upon the routing structure of a P2P network come in many forms such as:

a. Malicious agents might silently drop messages that they should forward.
b. Malicious agents might misroute messages to delay or prevent delivery **[6].**

- **Byzantine failures**

A router with a Byzantine failure may exhibit arbitrary behavior. For example, it may calculate incorrect routes, fail to forward packets at random, or inject arbitrary packets into the network. A router with a Byzantine failure is capable of disrupting most or the entire network layer routing protocol components and causing the network not to deliver packets with a high probability of success **[7].**

3 Requirements for routing in Peer-to-Peer overlay networks

An efficient and secure routing mechanism for peer- to-peer overlay networks should satisfy the following metrics

- **Availability**

In peer-to-peer network, availability denotes existence of multiple guarantees for location information to avoid a single point of failure and the availability of multiple paths to destination **[8].**

- **Integrity**

It denotes trustworthiness and consistent of data that are transmitted through the p2p networks. To enhance the integrity of routing in peer-to-peer networks, it is important to reduce the number of queries forwarded to malicious nodes **[5].**

- **Reliable**

A reliable system is in p2p networks should be capable of recovering when a failure occurs. The factors, which should be taken into account for reliability, are data replication, node failure detection and recovery **[8].**

- **Confidentiality**

Data Confidentiality in p2p networks means whether the information in the network is protected against unintended or unauthorized access. Data Confidentiality is a measure of the ability of the system to protect its data. Accordingly, this is an integral component of security **[9].**

- **Fault-tolerance**

Fault-tolerance support makes P2P systems capable of detecting certain categories of failures (e.g. hardware failure, network connectivity) and recover from these without, or with minimal, disruption of services **[10].**

4 Proposed Solution based on the problem identified

For peer- to- peer overlay networks, some routing mechanisms have already been proposed in the literature. The existing routing mechanism in peer- to- peer network can be categorized into two types as recursive routing and iterative routing. Recursive routing is efficient but not secure, on the other hand, iterative routing is not efficient but it provides security [11]. All the routing mechanism in the area of peer- to- peer overlay networks considers either recursive or iterative method and not both.

In paper [11] Xu Xiang et al. have proposed secure routing protocol for peer- to- peer network. In that, each intermediate node sends feedback message to the source node. Since it is based on iterative routing, it is not efficient and increases the overhead.

In paper [16], Wojciech Galuba et al. have introduced fault tolerant routing based on recursive routing in peer- to- peer network. They did not consider any security scheme to provide secured routing.

To avoid the above-mentioned problems of both types of routing, a reliable and secured routing protocol is to be designed. In this paper /proposal, we propose to employ hybrid secure routing mechanism in peer- to- peer overlay networks. Our mechanism is hybrid for it considers both recursive and iterative routing strategies.

Our proposed mechanism consists of three phases:

- Key initialization,
- Routing.
- Trust estimation

Key Initialization

During deployment of nodes in the network, each node sends JOIN-REQ to the DHT (Data Hash Table). JOIN-REQ is a request that contains signature generated by corresponding node. By authorizing the signature, the Private Key Generator (PKG) generates the private key (PK-) and issues to each appropriate nodes. Now, each node in the network has a pair of keys namely private and public key (PK- , PK+). In our mechanism, node ID is used as Public key (PK+).

Routing

When a node (source) needs service from another node (destination), it constructs service request in the format of < SER_REQ, PKS-, H (SER_REQ)>, where PKS- is the private key of the source node, H (SER_REQ) represents hash value of service request. Finally the entire message is signed by the source node. After constructing the SER_REQ, the source node looks at trust counter and selects the neighbor node that has higher trust value and then forwards to that node.

When an intermediate node receives the message, it verifies the signature by using PK+S (public key of the source node). If the signature is authentic, then it looks its trust counter and chooses the node, which has higher trust value. Finally, the service request reaches the destination node. The destination node verifies the signature and checks the hash value. Upon successful verification, the destination node accesses the service request.

Upon successful transmission, the destination node sends response message to the source node. By receiving the response message from the destination node, the source node sends positive feedback message to all intermediate nodes. If the source does not receive response message even after TRES, it sends negative feedback message to all intermediate nodes and retransmit the SER_REQ in another reliable path.

Trust Estimation

According to our mechanism, each node has trust counter, which has the trust value of their neighboring nodes. Initially trust counter value is set to one. And, the counter value is updated while receiving feedback message from the source (sender) node. If the feedback message is positive then the counter value is incremented by one. Otherwise, the counter value is decremented by one.

CONCLUSION & FUTURE WORK

This exploratory study provided a broad overview of previous research and the literature related to Peer-to-Peer overlay networks with their areas of focus. Analysis shows that keep the requirement metrics: Availability, Integrity, Reliable, Confidentiality and Fault-tolerance in mind, our proposed hybrid secure routing mechanism in peer- to- peer overlay networks can provide secure and reliable routing. Our proposed routing mechanism consists of three phases: Key initialization, Routing and Trust estimation.

As part of future work, we will implement the proposed work in Network simulator (NS-2) and prove that it attains good throughput with reduced overhead and delay in presence of attacks and failures properties can be achieved with very low overhead in synthetic network topologies. Thus, our proposed routing mechanism provides secured and reliable routing in peer- to- peer networks.

REFERENCES

1. Eng Keong Lua, Jon Crowcroft, Marcelo Pias, Ravi Sharma and Steven Lim, "A Survey and Comparison of Peer-To-Peer Overlay Network Schemes", IEEE Communications Surveys and Tutorials, 2005.

2. Lin Wang, "Attacks Against Peer-to-peer Networks and Countermeasures", TKK Seminar on Network Security, PP: 51-61, 2006.

3. Eng Keong Lua, Jon Crowcroft, Marcelo Pias, Ravi Sharma and Steven Lim, "A Survey and Comparison of Peer-to-Peer Overlay Network Schemes", IEEE Communications Surveys and Tutorial, 2004.

4. Hyojin Park, Jinhong Yang, Juyoung Park, Shin Gak Kang and Jun Kyun Choi, "A Survey on Peer-to-Peer Overlay Network Schemes", IEEE 10th International Conference on Advanced Communication Technology, (ICACT), 2008.

5. Dhruv Chopra, Henning Schulzrinne, Enrico Marocco, and Emil Ivov, "Peer-to-Peer Overlays for Real-Time Communication: Security Issues and Solutions", IEEE Communications Surveys & Tutorials, 2009.

6. Kevin W. Hamlen and Bhavani Thuraisingham, **"**Secure Peer-to-peer Networks for Trusted Collaboration", IEEE International Conference on Collaborative Computing: Networking, Applications and Worksharing, CollaborateCom , 2007.

7. Matthew Kerner, "Reliable Routing Protocols in Networks with Byzantine Failures", 2007.
8. B. Pourebrahimi K, Bertels S. and Vassiliadis, "A Survey of Peer-to-Peer Networks", ACM Computing Surveys, 2004.

9. http://hitachi-id.com/concepts/confidentiality.html

10. Rolando Martins, Priya Narasimhany, Lu´ıs Lopesz and Fernando Silva, **"**Lightweight Fault-Tolerance for Peer-to-Peer Middleware", 29th IEEE Symposium on Reliable Distributed Systems, 2010.

11. Xu Xiang and Jianguo Chen," A Secure Routing Protocol for Peer-to-Peer Network", Journal of Convergence Information Technology, 2010.

12. Aina Ravoaja and Emmanuelle Anceaume, "STORM: A Secure Overlay for P2P Reputation Management", IEEE First International Conference on Self-Adaptive and Self-Organizing Systems, (SASO '07), 2007.

13. Prashant Dewan and Partha Dasgupta, "P2P Reputation Management Using Distributed Identities and Decentralized Recommendation Chains", IEEE Transactions on Knowledge and Data Engineering, 2010.

14. Jo˜ao Leit˜ao, Liliana Rosa and Lu´ıs Rodrigues, "Large-Scale Peer-to-Peer Autonomic Monitoring", IEEE GLOBECOM Workshops, 2008.

15. Shruti P. Mahambre and Umesh Bellur, **"**Reliable Routing of Event Notifications over P2P Overlay Routing Substrate in Event Based Middleware", IEEE International Parallel and Distributed Processing Symposium, (IPDPS), 2007.

16. Wojciech Galuba, Karl Aberer, Zoran Despotovic and Wolfgang Kellerer, **"**Self-organized fault-tolerant routing in peer-to-peer overlays", Third IEEE International Conference on Self-Adaptive and Self-Organizing Systems, (SASO '09), 2009.

17. Ruichuan Chen, Eng Keong Lua, Jon Crowcroft ,Wenjia Guo, Liyong Tang and Zhong Chen, "Securing Peer-to-Peer Content Sharing Service from Poisoning Attacks", IEEE Eighth International Conference on Peer-to-Peer Computing (P2P'08), 2008.

BIOGRAPHIES

Raja Waseem Anwar faculty of Computer Studies at Arab Open University (AOU), Muscat, Sultanate of Oman. He holds the Master's Degree in Computer Science from Karachi - Pakistan. His research interests lies in the area of Information Security and Computer Networks. He has presented research papers in several National and International Conferences. He has publications in Intrenational journals.

Farah Abu Shahla faculty of Computer Studies at Arab Open University (AOU), Muscat, Sultanate of Oman. She holds a Master's Degree in Computer Science from the American University of Beirut since 2009 and a Bachelor Degree in Computer Science from the Notre Dame University since 2006. Her research interests are mobile and wireless networks

Saqib Ali Ph.D., is an Asisstant Professor at Sultan Qaboos University (SQU), Muscat, Sultanate of Oman. He holds a Ph.D. degree in Computer Science and masters degree in Information Systems from LaTrobe University, Australia. Dr. Ali is the authored and co-authored of several refereed journal and proceedings articles in various areas of Business Process re-engineering, e-commerce, mobile computing and Information Systems Security disciplines.

Sumitra Srinivas Kotipalli faculty of Computer Studies at Arab Open University (AOU), Muscat, Sultanate of Oman. She holds the Master's Degree in Computer Science from SPMV University, Tirupati, India. She worked with Indian space Research Organisation, and worked with Expert system development with GUI for PSLV Tracking System. Her research interest is Artificial Intelligence in mobile computing.

Characterizing the Scheduling Algorithms in IEEE 802.16

Farah Abu Shahla
Arab Open University
Muscat

Raja Waseem Anwar
Arab Open University
Muscat

Mathieu Moussa
Lebanon

KEYWORDS
WiMax, Scheduling, QoS, RR, Uplink scheduling, downlink bandwidth, DFPQ, WFQ.

ABSTRACT
The IEEE 802.16 standard (WiMax) defined for Broadband Wireless Access technologies has proven to be a cost effective wireless alternative to cabled networks especially for large range coverage in developing countries. This standard is widely accepted because of its more sophisticated support for QoS at the MAC layer. Although this standard defines the signaling mechanisms between the various parties in the network, it does not define any algorithm for packet scheduling. Such algorithms are left open for researchers and manufacturers to come up with. In this paper, we start by surveying and characterizing the different scheduling algorithms proposed in the literature while focusing on their limitations and advantages. After that and based on the pros and cons of existing methods, we will present the guidelines of work that suggest steps to follow when devising a scheduling algorithm for WiMax.

INTRODUCTION

The WiMax [12,13] technology plays a vital role in fixed broadband wireless metropolitan area networks (MAN). It has proven to be a cost effective wireless alternative to cabled access networks such as fiber optic links, and digital subscriber line (DSL). The main advantages of the WiMAX technology is its ability of covering larger ranges and its more sophisticated support for Quality-of-Service (QoS) at the level of the Medium Access Control (MAC) layer [2].

The basic architecture of the 802.16 standard consists of two types of fixed stations: a *Base Station* (BS), and a *Subscriber Station* (SS). The BS is a central equipment set providing connectivity, management and control of several SSs situated at varying distances while an SS can represent a building equipped with conventional wireless or wired LAN.

The 802.16 defines two basic operation modes for sharing the wireless medium: a mandatory point-to-multipoint or *PMP mode*, and an optional *mesh mode*. In the PMP mode, a controlling BS connects several SSs to various public networks including the internet. An SS cannot send to or receive from another SS unless it passes through the BS first. In the mesh mode, the SSs can directly communicate without the need for passing through the BS. In this paper, we will be focusing on the PMP mode alone since it is anticipated that the ISPs will prefer the use of the PMP mode to connect the customer to the internet.

The communication path between the BS and an SS is bi-directional: uplink (from SS to BS) and downlink (from BS to SS). The MAC protocol is connection-oriented. Each packet has to be associated with a connection at the MAC level which provides a reliable way for bandwidth request, QoS negotiation and other data transfer related actions. An SS registers to the BS first: a process in which the SS negotiates the initial QoS requirements.

The requested bandwidth requirements may be granted by the BS on two different bases: either per connection-based (GPC) or per subscriber station based (GPSS). After that process, an SS can start sending and/or receiving data. The QoS requirements can be changed dynamically either when an SS is polled by the BS or by exchanging *Dynamic Service Change* (DSC) messages [8].

Although the IEEE 802.16 defines the signaling mechanism for information exchange between BS and SS (connection set-up, BW-request, and MAP messages), it does not define a scheduling algorithm neither in the downlink direction nor in the uplink direction. In addition, the admission control in the BS is left undefined. All in all, the IEEE 802.16 specification does not define a scheduling algorithm at the BS that allocates resources (including time slots) for each SS. The BS should implement an efficient scheduling algorithm in order to achieve better utilization of the resources [3]. Because of the increasing demand on traffic that requires a lot of bandwidth (VoIP, MPEG, VoD…); the allocation of the slots should be based on both QoS requirements and bandwidth request sizes as not to hinder those services that require real-time access and at the same time not to starve the least priority services.

To improve on this issue, IEEE 802.16 groups applications according to their QoS requirements, their packet arrival pattern, and their means to request bandwidth for the BS. The resulting groups are: *Unsolicited Grant Services* or (UGS), *Real-time Polling Service* or (rtPS), *Extended rtPS* (ertPS), *Non real-time Polling Service* (nrtPS), *Best Effort* (BE).

IEEE 802.16 defines the UGS downlink scheduling (from BS to SS) as well as the UGS uplink scheduling and leaves the definitions of the uplink scheduling for the rest of the services (rtPS, ertPS, nrtPS, BE) undefined. The implementation of such algorithms is left open for different researches and manufacturers to come up with.

In the section II, we will present different scheduling algorithms that are found in the literature. These scheduling approaches are categorized into three different groups and discussed theoretically and through examples. In section III, we evaluate these techniques based on their advantages and disadvantages while in section IV, we propose a model to follow when designing an efficient yet simple scheduling algorithm for the WiMax IEEE 802.16. Finally, section V summarizes and concludes.

CATEGORIES OF SCHEDULING ALGORITHMS

This section characterizes existing scheduling algorithms into three major categories: Simple Techniques based on adopted techniques (Round Robin, Weighted Round Robin…), Layered Techniques which use a hierarchy of different simple techniques, and Sophisticated Non-Layered Techniques that define new formulations based on the QoS parameters (minimum bandwidth required …) of the various service flows.

Simple Techniques

When thinking about developing scheduling algorithms for an emerging technology such as that introduced in WiMAX, it is obvious that researches and developers would try to exploit the scheduling techniques already existing in other technologies such as wired networks.

In many network protocols, such algorithms prove to be very useful, unfortunately in WiMax, nobody uses these algorithms as is, because when used in their exact form they cannot guarantee the complex QoS properties that this standard requires.

The following are some of the simple algorithms that although are not used in there simple form, appear in more complex architectures:

Round Robin (RR) scheduling is one of the simplest algorithms for scheduling packets in a computer network. It assigns time slots (also called quantum) to each packet in equal portions and in order. Round Robin is very simple to implement and has several advantages. However sometimes, in more advanced networks, the simple round-robin scheduling is not sufficient where other parameters like priority should be taken into account. Variants of RR exist such as DRR (Deficit Round Robin) and DWRR (Deficit Weighted Round Robin).

The Weighted Fair Queueing (WFQ) is a generalization of the original Fair Queueing algorithm. It allows several data flows to fairly share the link capacity. The advantage over conventional First In First Out (FIFO) queuing is that an ill-behaved flow (consisting of large data packets or bursts of many packets) will only punish itself and not other flows. WF^2Q (Worst Case Fair Weighted Fair Queueing) was introduced to keep the delay bound.

The above techniques can be further classified under channel unaware schedulers [1]. They make no use of channel state conditions and other information such as power level, channel error and loss rates, in making the scheduling decisions. They basically assume an error free channel.

However since in reality, channel conditions cannot be guaranteed to be error-free, channel aware schedulers come to take these into consideration [1].

Although some would use these algorithms in their simplest form, others would prefer to merge, combine, or modify these algorithms to obtain more sophisticated scheduling algorithms that can guarantee QoS in IEEE 802.16.

Layered Techniques

Such methods use different types of simple techniques, some of which are mentioned above. The aim is to create a hierarchy of schedulers that can guarantee service and fairness for the 5 different QoS classes. Basically, the intuition is to couple the different service classes with the simple scheduling technique that fits it best. The co-existence of these simple techniques with or without the help of other techniques imposes priority among the services as a whole. These Hierarchal techniques are the most commonly used since they are the result of changes done to already popular and well defined algorithms.

Several of such approaches can be found in the literature [10, 14, 15, 18]. The most important of these approaches are to be explained in the remaining of this part.

The architecture proposed for uplink scheduling in [9] uses a hierarchy of schedulers. Changes are done to the Uplink Packet Scheduling (UPS) module in the BS as well as adding a traffic policing module at the SS. Three modules are added to the BS's UPS. The Information Module is responsible for retrieving the queue size of each connection from the BW-Request messages according to the corresponding service classes. It determines the arrival time and deadline for each rtPS packet, and will then update the Scheduling Database Module with the most recent information about all rtPS, nrtPS, and BE connections. The Scheduling Database Module will, in turn, store all the information concerning each connection, and for rtPS service flow it also stores the deadlines of each corresponding connection. Finally, the Service Assignment Module will allocate bandwidth by first passing on all service classes in strict priority (from highest to lowest: UGS, rtPS, nrtPS, and BE), and then, for each service class independently, will employ the following disciplines: fixed bandwidth allocation for UGS, earliest deadline first (EDF) for rtPS, weighted fair queuing (WFQ) for nrtPS, and equal bandwidth allocation for BE.

Although the algorithm in [9] improves on simple techniques by arranging them into a hierarchy, it still overlooks some issues. First, it does not consider downlink scheduling and supposes that it is rather simple since in the downlink subframe the only sender is the BS. Although the BS is the only sender, an algorithm for the downlink subframe is still important since it defines the amount as well as the type of data to pass to multiple SSs (or a single SS) and when. All these information will be generated in the form of a DL-MAP that will be passed during the beginning of a frame to specify the duration of the downlink subframe.

The authors of [15] also present a double layer service flow scheduling architecture. Their proposed algorithm resembles in many ways the algorithm just explained [9].

In the architecture proposed in [15], hereby referred to as DFPQ, the authors define an admission control policy to control the number of service flows, and propose a bandwidth allocation mechanism to provide QoS for different types of applications.

DFPQ, unlike in [9], defines not only uplink scheduling but also downlink scheduling at the BS which provides a more efficient utilization of the bandwidth. In the DFPQ algorithm, the QoS requirements are granted by the BS on a GPC basis.

At the highest layer, and instead of using strict priority among classes, the authors use a *Deficit Fair Priority Queuing* (DFPQ), thus in other words introduces the simple *Deficit Fair Queuing* (DFQ) technique. At the second level, they propose simple scheduling techniques for different classes: EDF for rtPS, WFQ for nrtPS and RR for BE.

The *Preemptive Deficit Fair Priority Queuing* (PDFPQ) in [10], the authors propose a variation of the above algorithm, and uses an idea similar to [16].

The PDFPQ architecture maintains the same scheduling queues at the lowest level (layer 2). However, at the first layer, they add the notion of preemption to the DFPQ architecture. The aim is to give the rtPS service flow packets higher chances to meet their deadlines while decreasing their delay to guarantee the QoS requirements.

All in all, the PDFPQ algorithm enhances the DFPQ technique by improving the performance of the rtPS service flow class.

In [11], an Enhanced Packet Scheduling Algorithm for IEEE 802.16 is defined based on Bennett_Zang model [6]. J. Bennett and H. Zhang had already proposed a good H_WF2Q+ scheduling framework that distributes weighted bandwidth to various sets of flows based on some criteria. The algorithm of [11] adjusts the model of Bennett_Zang in order to serve multimedia traffic thus providing QoS support for most of the real time applications as defined in the WiMax standard. In [11], traffic (represented by a connection) is grouped into two major categories: *best effort traffic* and *QoS traffic*. They then classify *QoS traffic* as being either *hard-QoS* or *soft-QoS* traffic. Soft-QoS traffics are traffics whose minimum bandwidth requirement is less than its maximum bandwidth requirement. The rtPS and nrtPS traffic are considered as Soft-QoS traffic. Unlike Soft-QoS, Hard-QoS traffics are those traffics whose minimum and maximum bandwidth requirements are equal. UGS is considered as Hard-QoS because its bit rate is constant.

The EDF-BWA Scheduling Algorithm [14], hereby referred to as EBSA, matches the CBR-like and VBR-like traffic. It is used to guarantee the UGS, rtPS and nrtPS types of services with their specific requirements. The SS QoS policy module is used at the first layer to analyze the QoS requirements. It does so to make sure that these requirements are not violating the QoS contract including the maximum bandwidth requirement. The services are then classified in different queues using the Classifiers module. EBSA will provide UGS services periodic grants allocated using the EDF scheduling scheme.

Many researchers have tried to find newer independent ideas that do not rely on any of the above mentioned techniques, but instead use the exisiting QoS parameters of WiMax to define original, more sophisticated and non-traditional techniques.

Sophisticated Non-layered Techniques

These are algorithms that involve non-traditional methods of scheduling. They do not depend upon a hierarchy of other simple scheduling algorithms and yet they cannot be considered as simple techniques.

A scheduling solution for the IEEE 802.16 BS to ensure the QoS requirements of SSs in the uplink and downlink directions has been presented in [4]. Slots are allocated based on QoS requirements, bandwidth request sizes and WiMAX network parameters. The algorithm is conceptually similar to WRR (Weighted Round Robin). The BS scheduler holds the burden of allocating slots for each individual connection in a way to allow polling (by the SSs) as well. The algorithm works as follows: the BS scheduler serves the connections starting from the highest priority class. For each connection, the number of slots is calculated based on the class it belongs to e.g. UGS, rtPS, ertPS … Initially the authors of [4] came up with a formula to calculate the number of slots Ni that fit in each frame.

The scheduling algorithm is simple and fast since it is based on the round-robin scheduling. Furthermore it takes into consideration the minimum/maximum bandwidth requirements, class type, slot size, the bandwidth request size, frames-per-second and modulation.

Algorithms on how to allocate free slots and order the slots to decrease the jitter have also been discussed in [4]. All in all, the paper presents an efficient solution for the BS to be able to allocate slots based on the QoS requirements, bandwidth requests and network parameters.

Another approach such as that proposed in [7], presents an uplink packet scheduling with call admission control (CAC) mechanism that is token bucket based. In order to characterize traffic flows they suggest a mathematical model. In the proposed model, bandwidth needed by real-time flows can be properly reserved while promising their delay requirements. In the uplink packet scheduling algorithm, they adopt Earliest Deadline First (EDF) mechanism in its simplest form.

OVERALL CRITIQUE

Although the simple techniques that we presented in section I part A serve as good stand-alone techniques when used in scheduling various flows with minor differences in characteristics, they prove inefficient when used in WiMax. This is due to the fact that various classes in WiMax differ on a large scale when it comes to QoS parameters [5]. The priority imposed among these classes does not allow for a single simple scheduling algorithm to be used alone since some classes (UGS, rtPS, ertPS) have more stringent delays than others. All in all, the simple techniques for scheduling in WiMax show to be inconvenient when QoS guarantees are needed.

The layered architecture uses a hierarchy of layers each one focusing on a certain goal. The advantage introduced by combining different levels of simple techniques comes in terms of guaranteeing QoS for various classes. The multiple layers can impose priority among the service classes in WiMax. It also prevents starvation of low priority classes (nrtPS and BE) by introducing fairness.

The problem with such layered techniques is the fact that they introduce implementation complexity. The modifications and layering done to simpler techniques can yield such complexity that can, in turn, introduce some unnecessary overhead during implementation and execution. Although these techniques have many benefits it's always necessary to ensure that the introduced complexity does not affect the validity and efficiency of scheduling decisions and subsequently jeopardize the QoS requirements. The DFPQ and its Preemptive extension proposed in [15] and [10] respectively use a two-layer structure which is based on the GPC mode where the bulk of the scheduling work is done at the BS. It is a challenging task to use a hierarchy of schedulers because the per-connection QoS requirements should be translated into the scheduler configuration at each level. Also, it is not enough to calculate the scheduler configuration only once when an SS joins or leaves the network. As SSs send data, their request sizes change all the time. As a result, the scheduler at the BS should reassign slots very frequently.

A more sophisticated method that does not rely on layering could be a better solution since there is not much time to do the scheduling decision. The advantage of such methods is decreasing the overhead imposed by having layers of scheduling decisions to be performed before actually assigning a slot to the frame. The authors of [4] actually provide a simple to understand, multiple algorithms based architecture that guarantees QoS while exploiting the network resources. This gives us the intuition that the simpler the algorithm is, the less scheduling decision time it takes and thus the more efficient it is.

The problem with this algorithm is that new arrivals with more stringent delays might have to wait at the SS while they should be running instantaneously instead of the other connections. This is due to the fact that the SS does not make any decision itself (no scheduler at the SS).

The token bucket algorithm proposed in [7] is original and new of its kind. As far as our knowledge, no such similar work has been applied in this literature. This algorithm sets forth a kind of traffic conditioning especially on the rtPS class. Although the paper in hand defines mathematical formulas to calculate various parameters concerned with the flow, it directs its attention mostly to the rtPS class storing its related information in a database. The fact that they turned their attention to one class would lead to neglecting the priorities of the other classes and jeopardizing the overall fairness.

Each of the surveyed approaches in the previous sections consists of interesting ideas and techniques that exhibit minor pitfalls in various points of concern. Having a global view of all such methods can help researchers formulate a better model that can provide sufficient help when devising a good and efficient scheduling algorithm.

The proposed guidelines in the following section exploits both the advantages and disadvantages of all the mentioned approaches.

PROPOSED GUIDELINES FOR DESIGNING SCHEDULING ALGORITHMS IN IEEE 802.16

In this section we develop main guidelines to follow in order to develop a new, efficient and well organized scheduling architecture in IEEE 802.16. Our strategy makes use of different categories mentioned throughout the paper. We will start by investigating the differences between GPC and GPSS. Based on our preferences of which mode to use, we will define accurate models for the BS and SS schedulers respectively.

In the definition of the IEEE 802.16 standard, they state that the SS sends the request for bandwidth according to its pending connections, however the BS allocates the total bandwidth required per SS while trying to ensure its minimum required backlog. In other term the WiMax standard proposes the usage of GPSS.

The GPSS grants bandwidth to the SS instead of to each connection, the SS scheduler will then re-distribute bandwidth among its connections, thus off-loading the base station's work. This will contribute to a low overhead at the BS side however it requires a more intelligent subscriber station. The GPSS mode is suitable for many connections per terminal, and it allows more sophisticated reaction to QoS needs.

Some researches tried to use a radical approach which instead of allocating the backlog for the SS as a whole, the bandwidth is allocated per connection using the CID (Connection ID) thus omitting the need for a sophisticated scheduler at the SS. This approach, known as the GPC mode, leaves a higher burden on the BS scheduler to allocate slots for each of the connections, implying a higher overhead at the BS and a simpler scheduler at the SS. The GPC mode is mostly suitable for few users per subscriber station.

After investigation, we conclude that the GPSS mode is better than the GPC mode for the following reasons. First, the bandwidth granted to the SS will remove a high burden off the BS. Since the BS is already responsible for synchronizing the communication between all the SSs in a network, it is a good idea to remove the load of assigning slots per connection. Second, offering the chance for SS to redistribute the total received bandwidth will give it the ability to decide upon which connections are more important than others especially if the connection is a newly arriving one with more stringent delays and it has not requested bandwidth yet.

Since we decided to go with the GPSS mode, the rest of this section will be focusing on how a good BS and SS schedulers should look like in such mode.

The Base Station should contain the following functionalities and mechanisms. First and foremost it should remain as simple as possible so that it doesn't get affected when scaling the network (i.e. adding more SSs). For this idea to hold feasible, the BS should not use a hierarchy of schedulers because of the decision overhead that such architecture imposes. The main aim of this coordinating

station is to calculate the minimum required bandwidth of each SS in its network. The best way to execute this function is by using mathematical formulations such as those used in [4, 20] due to the fact that they are fast and yield the most approximate results. Another worry for the BS is to minimize the distance between slots that are allocated for the same SS, and this is to decrease possible jitter and delay factors that might hinder pending real time applications. Any remaining slots that are not allocated in a sub-frame, should be redistributed fairly amongst stations with bottom-priority classes as not to cause their starvation.

When the SS receives the total bandwidth that is allocated for it in the upcoming frame, it is supposed to redistribute this backlog equally and fairly amongst the pending connections. To do this the SS scheduler should re-examine the service class queues to check for any newly-arrived unaccounted for connections. If such connections exist, and exhibit high priority and less tolerance for delay, the SS should re-evaluate their needs and assign part of the granted bandwidth to these connections. The other less demanding connections will be delayed to another upcoming bandwidth request. The idea here is to maintain the QoS service parameters as effective as possible. The SS can use some kind of the simple techniques (such as EDF) to evaluate which of the pending connections is more critical. As long as the SS does not use a merge of many simple techniques, it will remain efficient and responsive to real time updates.

In the following we present the summarized list of the mentioned guidelines for modeling scheduler at the BS and SS in IEEE 802.16 functioning under the GPSS mode.

Guidelines for Modeling Schedulers at BS and SS

Based on our choice of the GPSS mode the guidelines for the schedulers come as follow.

1. ***BS Scheduler:***
   a. *Do not include a hierarchy of schedulers due to the overhead it causes.*
   b. *Calculate the minimum bandwidth required for each SS using certain accurate mathematical formulations.*
   c. *Minimize the distance between slots that are allocated for the same SS to decrease possible jitter and delay.*
   d. *Redistribute the remaining bandwidth fairly amongst stations with bottom-priority classes as not to cause their starvation.*
2. ***SS Scheduler***
   a. *Use some kind of the simple techniques (such as EDF) to evaluate which of the pending connections is more critical as long as the SS does not use a merge of many simple techniques.*
   b. *Redistribute the granted bandwidth equally and fairly amongst the pending connections by re-evaluating their needs and assign part of the granted bandwidth to newly- arrived unaccounted for connections if they exist.*

Our proposed framework is flexible thus allowing the developer of a scheduling algorithm in IEEE 802.16 to apply minor changes as he sees necessary during implementation or analysis of the specific network topology in hand.

CONCLUSION

In this paper, we characterize the various types of scheduling algorithms in IEEE 802.16 according to their functionalities and design. The resulting characterization yields three groups of scheduling algorithms: Simple and less used techniques, Layered more popular mechanisms, and newly designed Sophisticated Non-layered techniques. We substantiate our classification by presenting the most popular approaches in the literature and explaining their main intuition. We continue by stating the advantages and disadvantages of the mentioned techniques and categories, as well as comparing and criticizing them.

Finally we make use of the presented classification and present interesting guidelines to follow while developing a scheduling architecture for IEEE 802.16. Our guidelines originate from a GPSS mode and continue by defining the respective BS and SS scheduler design strategy.

REFERENCES

[1] S. Chakchai, R. Jain, A. Tamimi, Scheduling in IEEE 802.16e mobile WiMAX networks: key issues and a survey, IEEE Journal on Selected Areas in Communications, vol.27, no.2, pp.156-171, February 2009.

[2] S. ZhenTao, G. Abdullah, S. XiuYing, L. Ning, Improving QoS Of WiMAX By On_Demand Bandwidth Allocation Based On PMP Mode, Journal Of Computers, October 2011.

[3] P. Chandur, R.M. Karthic, K. M. Sivalingam, Performance Evaluation of Scheduling Algorithms for Mobile WiMax Networks, Eight IEEE Workshop on Pervasive Wireless Networking 2012, Lugano, March 2012.

[4] A. Sayenko, O. Alanen, J. Karhula, T. Hamalainen, Ensuring the QoS Requirements in 802.16 Scheduling, ACM 2006

[5] N. Abu Ali, P. Dhrona, H. Hassanein, A Performance Study of Uplink Scheduling Algorithms in Point-to-Multipoint WiMax Networks, Computer Communications, 2009.

[6] J. Bennett, H. Zhang, Hierarchical Packet Fair Queueing Algorithms, ACM SIGCOMM 1996.

[7] T. Tsai, C. Jiang, C. Wang, CAC and Packet Scheduling Using Token Bucket for IEEE 802.16 Networks, Journal of Communication, Vol 1, No 2, May 2006.

[8] H. Safa, M. Watfa, Book Chapter: Quality of Service in Wireless Local and Metropolitan Area Networks

[9] K. wongthavarawat, A. Ganz, IEEE 802.16 Based Last Mile Broadband wireless Military Networks with Quality of Service Support, In Proc of. IEEE Milcom 2003.

[10] H. Safa, H. Artail, M. Karam, R. Soudah, S. Khayat, New Scheduling Architecture for IEEE 802.16 Wireless Metropolitan Area Network, IEEE 2007

[11] Y. Shang, S. Cheng , An Enhanced Packet Scheduling Algorithm for QoS Support in IEEE 802.16 wireless Network, Lecture Note in Computer Science, Volume 3619/2005.

[12] IEEE Std 802.16-2004, IEEE Standard for Local and Metropolitan Area Networks Part 16: Air Interface for Fixed Broadband Wireless Access Systems, 2004.

[13] IEEE Std 802.16e-2005, IEEE Standard for Local and Metropolitan Area Networks Part 16: Air Interface for Fixed Broadband Wireless Access Systems, Amendment 2: Physical and Medium Access Control Layers for Combined Fixed and Mobile Operation in Licensed Bands, 2006.

[14] V. Rangel, J. Ortiz, J. Gomez, Performance Analysis of QoS Scheduling in Broadband IEEE 802.16 Based Networks.

[15] J. Chen, W. Jiao, H. Wang, A Service Flow Management Strategy for IEEE 802.16 Broadband Wireless Access Systems

in TDD Mode, Proc. IEEE International Conf. on Communications, (ICC 2005), May 2005, pp. 3422 – 3426.
[16] M. Hawa, D. W. Petr, Quality of Service Scheduling in Cable and Broadband Wireless Access Systems, Proc. of the 10th IEEE Int'l Workshop on Quality of Service 2002, May 2002, pp. 247 – 255.

BIOGRAPHIES

Farah Abu Shahla faculty of Computer Studies at Arab Open University (AOU), Muscat, Sultanate of Oman. She holds the Master's Degree in Computer Science from Lebenese American University - Beiruit. Her research interests lies in the area of WiMax and Computer Networks. She has presented research papers in several National and International Conferences. She has publications in Intrenational journals.

Raja Waseem Anwar is a faculty of Computer Studies at Arab Open University (AOU), Muscat, Sultanate of Oman. He holds the Master's Degree in Computer Science from Karachi University. His research interests lies in the area of Information Security and Computer Networks. He has presented research papers in several National and International Conferences. He has publications in Intrenational journals.

Mathieu Moussa received a M.Sc. in Computer Science from the American University of Beirut in 2009. His research interests include WiMax and Wireless Sensor Networks. Mathieu is currently a Software Developer with a focus on web technologies.

SECURITY SYSTEMS

SIMULATION OF LIBRARY SECURITY SYSTEM USING IRIS SCANNING

R. Bremananth and Ahmad Sharieh
Information Systems and Technology Department
Sur University College
P.O. Box 440 P.C 411,
Sur, Oman.
E-mail: *bremresearch@gmail.com / bremananth@suc.edu.om, **ahmadsharieh@suc.edu.om / sharieh@ju.edu.jo

KEYWORDS
Authentication, Biometric, Iris scanning, Library security, Security Access.

ABSTRACT

Authentication is an essential task in the library management system while transaction is carried out by different customers either in on-line or off-line. This paper mainly focuses on how to provide the authentication for library accessing users when each circulation has taken place. Furthermore, suggested algorithm will easily be involved in the traditional library systems. In this work, we have proposed a frame work for library security using iris scan and implemented authentication algorithms using Java. Experiments are performed based on 1250 diverse subjects in different angles of variations in the authentication process. The results motivates that the proposed methodology will deploy in the present process of library management system in order to improve the secure transaction.

INTRODUCTION

Library is a common resource sharing center for entire community in the world. It could safely maintain by the administrator and librarian. However providing authentication to the authorized person is a crucial process in order to minimize the loss of materials. Usually, in a library, materials include books, periodicals, newspapers, journals, lecturer notes, prints, documents, CDs, e-books, videotapes, DVDs, cassettes, and others. All these materials are required a secure access by the user with different age levels and authorization level. A stock verification was reported that a small library in some countries lost 139 books per month when the average attendance in the library is about 300 to 400 users daily (Mark 2004). Even though, certain library has multiple of security towers with paper sensors or RFID (Radio Frequency Identification) readers, it is quite obvious that some users can pull the wool over somebody's eyes. The security issues in most of the libraries still exist. These security problems motivated us to incorporate a new methodology of iris scanning in the library entry and exit systems.

Biometric identification can be utilized in a wide range of applications such as passport administration, security entry, and access to privileged information, computer login or any other bank transaction. Biometrics is branch of computer vision that is user-friendly, intelligent and reliable for a person recognition system. However, one of the most hazardous security coercion is masquerade in that some person claims somebody's materials. Due to lack of security, unauthorized persons may get right to use confidential data or important accessories can be stolen. Hence, there is an essential requirement for person characteristic based authentication attributable to the fact that it can endow with the utmost fortification against masquerade. Among other biometrics, iris is a unique pattern in nature. In fact, no two irises in the world are identical. Even a person's left and right eyes have completely diverse patterns. For this reason, as compare with other biometrics such as ears, fingerprint, and face, iris can aptly be used as a kind of living password throughout the life time of human with one-time enrolment.

A sample iris image is shown in Figure 1, that acquired in diverse angle of variation. The biological features of the iris have a complex pattern which contains many distinguishing features such as arching ligaments, crypts, radial furrows, pigment frill, pupillary area, ciliary area, rings, corona, freckles and zigzag collarette. These spatial patterns in the iris are unique in nature (Bremananth and Chitra 2006). The key issue of biometric pattern recognition problem is the relation between bury and non-bury variability. That is, pattern categories can be efficiently differentiated only, if the variability between features of a given cluster is less than the variability between other clusters. In the iris recognition process, variability of iris non-bury cluster features are less than the bury cluster variability (Zhang 2009).

The iris recognition is used in this paper, initially acquires eye images by testing the changes in diameter of pupil boundary from the same distance of capturing. Subsequently, iris segmentation, normalization, and enrichment phases are performed to make the iris patterns suitable for feature extraction process. Finally, the iris discriminator design phase classifies the irises. The core process of iris scanning is then incorporated with library security system in order to authenticate the library users based on the diverse authorization of level of users.

The main contributions to this paper are: Iris recognition with certain distance and a frame work for library security system based on Iris pattern. The remainder of the paper is organized as follows. Section II describes the proposed method of library security system using iris recognition system. Implementation and simulation results are illustrated in Section III. Section IV consists of concluding remarks on our paper.

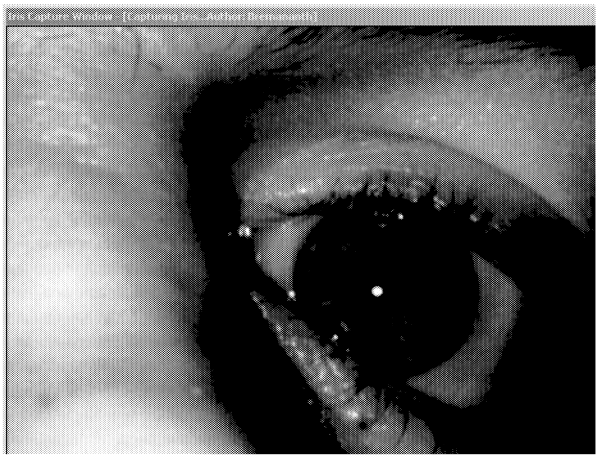


Figure 1: A sample Iris captured at a distance with diverse angle.

LIBRARY SECURITY USING IRIS SCAN

Library entry and exit users are monitored by iris cameras. The user has to show his or her left or right iris to access the library gate. Once authentication is done by the iris scan system, door of the single user has been opened and user is allowed inside the library. On exit, user iris is scanned along with the opted library materials. They are scanned with respect to its access code. If the opted library materials are authorized to be taken outside by the scanned iris, then the system gives process the opted material otherwise rejects the current process and waiting for the next user request. A frame work of the proposed system is shown in Figure 2. It consists of two main processes for on exit user from the library. The first process is the scanning of library access codes. The second process is the scanning of iris of opted user. If the user is authenticated, then opted material is also eligible to access by the user. Then chosen material will be delivered otherwise the current process is terminated and process goes to next user request.

Authentication and Process of iris Recognition

The authentication has been performed with two main operations such as:

• To store iris features, this is called as person's iris characters (PIC). The proposed system has PIC of 96 features.

• The second main operation is real-time matching of iris features.

For these two main processes, the system requires various steps of following operations in order to extract the features and iris matching.

An important and complex step of iris recognition system is image acquisition in the library environment. It is difficult to acquire clear images using the standard camera with ordinary lighting. In this module, a biometric camera is used to capture the user's eye images. It acquires images by passing NIR (near infrared) waves. The acquisition distance is normally between 19 and 36 inches and the average capturing time is 1•5 seconds (Kumar 2008) (Tsai and J. Luo 2011) (Zhao 2010).

The proposed system here can work both in outdoor and indoor environments without any hot spot of lighting intensities. However, unlike face, palm and fingerprints acquisition, irises are internal organs which are present inside the closed area of the eyelids; for this reason users must provide full cooperation for acquiring their eye images.

Iris extraction from eye image

Iris extraction is the process to remove unnecessary data such as pupil, eyelashes and other portions from the eye images. They are not required for the Iris feature code generation and iris identification or recognition process. The first step to extract the iris is inner localization. In this process, we have utilized eight-way symmetry circular method for filtering the inner periphery of the pupil (Bremananth and Chitra 2006, 2008) (Daugman 2004). Once inner periphery is localized then the extracted image is passed on to outer periphery localization module. This module has four-way symmetry circular method that localizes the outer periphery of the iris (Bremananth and Chitra 2006, 2008) (Daugman 2004).

The portion of eyelid or eyelashes is removed from the eye images since they are not a part of iris. They may tend to lead misclassification in the iris recognition process. For this reason, eyelid/eyelashes removal is also an important process in the iris authentication. These partitions are removed by looking the contrasted low-level pixel operation because the iris portion and the eyelashes/eyelids overlap each other and it is very hard to eliminate these properly.

After localization, iris image is now ready for feature extraction. The feature code of iris is extracted based on applying Gabor filter banks (Bremananth and Chitra 2006, 2008, 2011). Each bank of filter has 24 kernels, and then iris pattern is subdivided into 4 consecutive signals. In this system, 96 features are extracted from a person in order to accomplish the identification process.

The following java source code is developed for creating the Gaussian envelope in order to wrap the Gabor filter on the gray pixel surface of the iris patterns.

```
public void createGaussianFreqDirKernel(float sigma)
{
int n = (int) Math.ceil(4.0f*sigma);
int size = 2*n+1;
```

```
float[] data1 = new float[size*size*24];
double s = 2.0*sigma*sigma;
float norm = 0.0f;
int k = 0,f;
for (f =2;f<65;f*=2) {
for (int t = 0;t<136;t+=45){
for (int y = -n; y <= n; ++y){
for (int x = -n; x <= n; ++x, ++k)
```

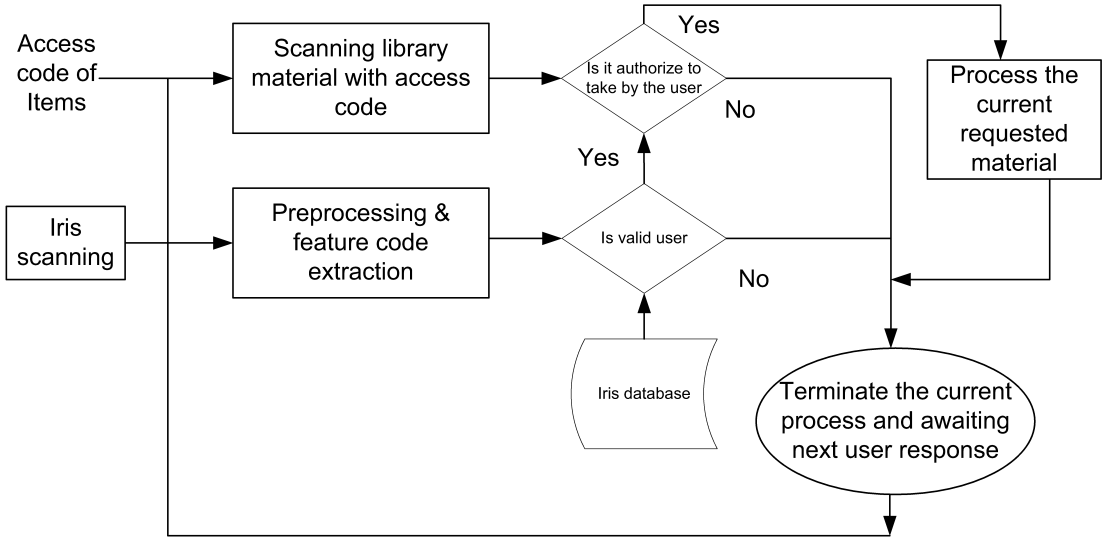


Figure 2: A Frame work for Library security using Iris scanning.

```
{
double r = Math.sqrt(x*x + y*y);
data1[k]=(float)(Math.exp(-(r*r)/s)*(
Math.cos(2*3.14*f*((x*Math.cos(t))+(y*Math.sin(t))))));
norm += data1[k];
}}}}
for (k = 0; k < size*size*24; ++k)
{
data1[k] /= norm;
}
k=0;
for (int i=0;i<24;i++ )
{
for (int j=0;j<size*size;j++)
{
gaussianKerData[i][j]=data1[k];
k++;
}}}
```

The aforesaid source code Gaussian envelope has a set of frequency changes ranging from 2 to 64 and incremented by power of 2. The rotation angle theta ranges from 0 degree to 135 degree by incrementing 45 degree each. Once the kernels have been formed then extracted iris patterns are operated by the Gabor filters in order to fetch the feature codes.

IRIS ADMINISTRATION

Registration of new user: After user personal information is entered and then biometric camera is enabled to acquire eye images of users. Initially, both left and right eye images of a user are captured, preprocessed and their iris code features are extracted. The feature set is stored in the database for the verification or recognition. Iris Recognition is called as one-to-many testing of irises in the database. User can opt to give either left or right eye to recognize the system as a registered user. Once authentication is provide by checking the appropriate iris feature code, then user is allowed to access the chosen library materials.

In the iris verification, user has to enter his/her user code or number in the system and then biometric camera enable to capture either left or right eye in order to do the matching process. This process is called as one-to-one matching.

Iris code modification: Registered user iris may reenroll after five to six years period of time however based on the literatures and state-of-art techniques. It is an optional one and iris pattern is stable throughout his/her life time.

Iris code deletion: Registered users can remove from the database once their record is no longer required by the administration.

Iris code and Privileged users: Based on the iris code certain materials in the library are restricted to use by diverse classes of users.

Data flow of Enrolment process

During the enrolment process, users are required to remove their eyewear in order to extract non-vulnerability iris features. The iris camera interface is used to acquire the sequence of frames of eyes images, among them a suitable frame is extracted for preprocessing. Figure 3 shows a data flow diagram of the enrolment process. Once a basic extraction of suitable frame is completed then the acquired image is passed on to preprocessing and feature extraction

modules. Next, iris code is store appropriately with register user personal details in the database.

Data flow of verification process

Verification process is an important for the security authentication in which users are authenticated by the system or rejected. This process has two different approaches such as recognition or identification and verification. During the recognition process, one user iris feature code will check with the remaining feature code and system itself to be acquainted with such a person is this (Figure 4). However, this process may cause more false positives while iris data store has millions of iris codes. Alternatively, verification process, normally cause minimum false positives even though iris data store has a huge amount of data entries because it requires an indexing key such as user identification to match the registered iris code exactly. This process may cause false alarm due to the nature of meters of distance iris acquired by the system or other sources of external artifacts.

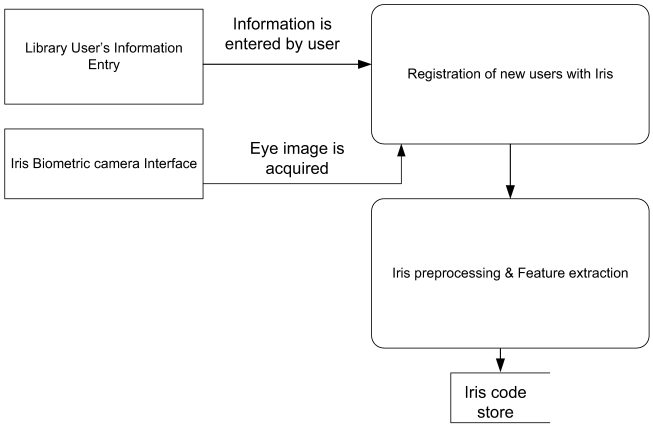


Figure 3: Data flow diagram of the Enrolment process.

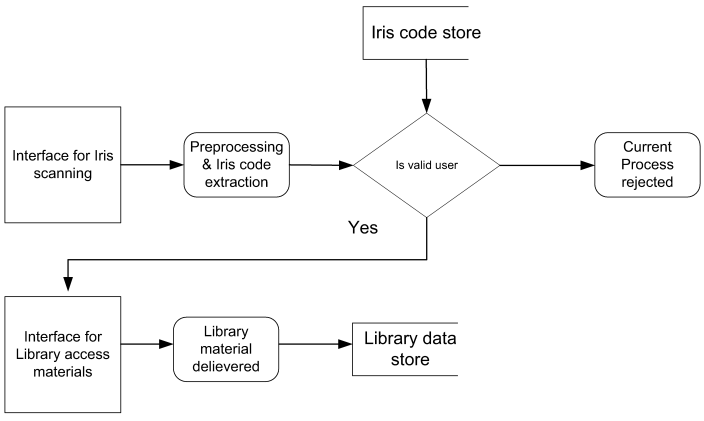


Figure 4: Data flow diagram for the verification process.

SIMULATION RESULTS

Simulation of library security system using iris scanning is implemented in java language. In order to verify the security system we have collected 1250 iris images from 1250 diverse subjects (Bremananth and Chitra 2006, 2008, 2011). The option of giving left and right is also considered while registering the new users. In the iris matching process, bury and non-bury classes of iris features are efficiently estranged and they prevent impostors from entering into the secure system. To authenticate any genuine user in the library, iris feature sets are treated as trained sets and stored in the encrypted database such as Oracle server was utilized. Verification subjects' irises are represented as test sets. The same subject iris feature codes may vary due to external noises, lighting, illuminations and other factors such as closed eyelashes or eyelids. This could lead to different iris template for an eye, even though iris is unique in nature. In order to study the variation of external artifacts, we have done the simulation of the eyelashes and eyelids pixel supremacy in the iris portion as shown in Figure 5. It depicts that eyelashes and eyelid pixels are portrayed about 20-35% of portion on the irises.

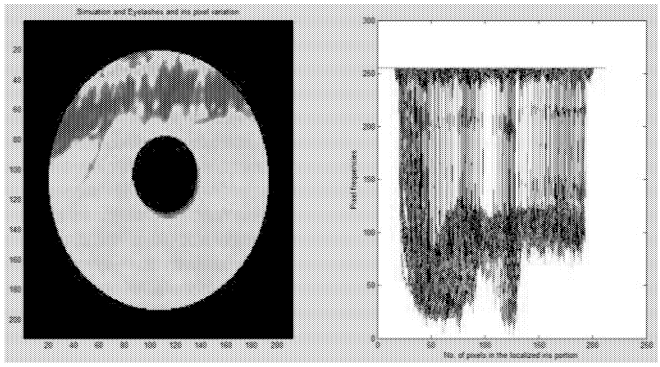


Figure 5: Supremacy of eyelashes and eyelids on iris.

In order to distinguish candidates' iris classes, an iris feature is represented as a 1D feature vector of size 'n' for each candidate. Iris feature code is used for recognition process, where denotes each signal of the feature which is convolved by the Gabor filter bank. These sensitive features are the mean value of resultant Gabor kernel convolution operation. The feature vector was determined by the number of decomposition of sub images and number of Gabor kernel computed in the system. Euclidean norm distance measurement discriminator is employed to separate class variation of iris patterns. It can be computed by taking the difference between enrolled iris code and unknown iris codes, divided it by standard deviation of enrolled iris code. During the testing process standard deviation values of unknown pattern are not necessary for calculating ENDM since it only computes based on mean values of the unknown. Therefore, in the testing process a classifier requires only n/2 features to recognize different sets of iris codes. To authenticate any genuine user, iris feature sets are treated as known sets that are already stored in the encrypted database and verification candidates' iris are represented as an unknown set. The distribution of iris

features was measured and illustrated in based on the authentication library database as shown in Figure 6.

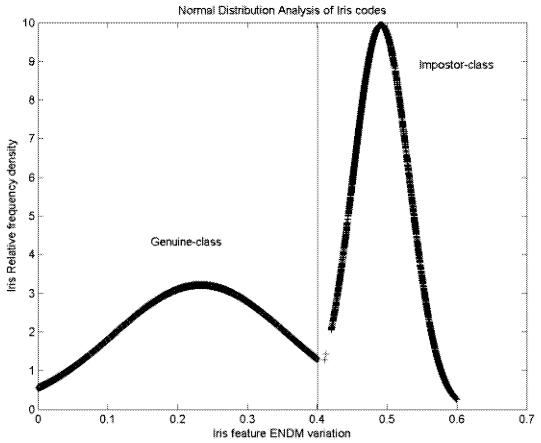


Figure 6: Binominal Normal distribution of Iris codes for Genuine and Imposter classes.

In general, the binomial distribution models the total number of successes in repeated trials from an infinite population in the iris database. Distribution of iris statistical distances depicted that each trial has only two exhaustive and mutually exclusive outcomes, i.e., 'success' and 'failure'. As per statistical theory, the system needs an idealized situation where the iris statistical threshold is measured more and more accurately for an increasingly large database. A normal distribution (ND) provides distribution analysis between the relative frequency density and iris code weighted variations. In the proposed approach threshold value of the decision boundary of ENDM was set from 0.0 to 0.4. If boundary was raised up to 0.4, there is a probability of a hit, i.e., genuine candidates were recognized, else false alarm is fired. This statistical threshold was measured by the distribution mean and standard deviation iris codes and their degree-of-freedom upheld in the system.

Receiver Operator characteristics

The following metrics of iris verifications were performed: Genuine accept rate (GAR): If the system was validated with the genuine subjects and the results were also positive, then it is referred as GAR. False rejection rate (FRR): If the system was confirmed with genuine subjects but the responses were negative, then it is called as FRR. True rejection rate (TRR): If the system was verified with false subjects and outcomes were also negative, then it is known as TRR. False positives rate (FPR): If the false subjects were verified with the system and reactions were positive, then it is referred as FPR.

CONCLUSION

In this research work, we have studied the essential security issues in the library environment. We have contributed mainly two solutions to the problem such as iris recognition with certain distance and a frame work for library security system. The system is simulated in the platform independent environment and it can be deployed in both on-line and off-line applications. Furthermore, the proposed iris methodology is aptly suitable for any real-time applications such as biometric-voting, employee management system, revolutionary identification, passport authentication and other applications. This paper unties a new avenue in iris recognition security and application with diverse areas. In addition, this system can be used to identify a person at a few meters distance. A new approach will be recommended to delete eyewear from acquired images and lessen artifacts occurring in the eye area.

ACKNOWLEDGEMENT

Authors would like to thank the Management of Sur University College to provide financial assistant to do this research a successful.

REFERENCES

Bremananth R and Chitra A 2006. "New methodology for a person identification system". *Springer - S¯adhan¯a*, 31, 3, 259–276.

Bremananth R, 2011. "Transformation Invariance and Luster Variability in the Real-Life Acquisition of Biometric Patterns". *International Journal of Computer Science and Information Security*, 9, 11, 8-15.

Bremananth R. and Chitra A June 2008. "Rotation Invariant Recognition of Iris". *Journal of Systems Science and Engineering, SSI*, 17,.1, 69-78.

Chitra A and Bremananth R. 2004. "Efficient identification based on human iris patterns". *Proc. of Fourth Indian Conf. on Computer Vision, Graphics and Image Processing (ICVGIP 2004)*, 177–183.

Daugman J. 2004. "How iris recognition works". *IEEE Trans. Circuits Syst. Video Technol*. 14, 21–30.

Hussein H. Fakhry and Benedict B C. 2006. "Research and Development of an iris based recognition system for identification and secure authentication". *Information & Security: An International Journal*, 19, 39-57.

Kumar A. 2008. "Computer-vision-based fabric defect detection: A survey". *IEEE Trans. Ind. Elec.*, 55, 1, 348–363.

Picon A, O. Ghita, P. Whelan, and P. Iriondo. 2009. "Fuzzy spectral and spatial feature integration for classification of nonferrous materials in hyperspectral data" *IEEE Trans. Ind. Inform.,* 5, 4, 483–494.

Tsai D and J. Luo. 2011. "Mean shift-based defect detection in multicrystalline solar wafer surfaces". *IEEE Trans. Ind. Inform.*, 7, 1, 125–135.

Uludag U, Pankanti S, Prabhakar S, Jain A K 2004. "Biometric cryptosystems: Issues and challenges". *Proc. IEEE*, 948–960.

Zhang D, G. Lu, W. Li, L. Zhang, and N. Luo. 2009. "Palmprint recognition using 3-D information". *IEEE Trans. Syst., Man, and Cybern.*, Part C: Appl. Rev., 39, 5, 505–519.

Zhang L, D. Zhang, and H. Zhu. 2010. "Online finger-knuckleprint verification for personal authentication". *Pattern Recog.*, 43, 7, 2560–2571.

Zhao Q., D. Zhang, L. Zhang, and N. Luo. 2010. "Adaptive fingerprint pore modeling and extraction". *Pattern Recogn.*, 43, 8, 2833–2844.

WEB REFERENCES

Mark McComb. 2004. "Library Security. 2004," http://www.librisdesign.org/docs/LibrarySecurity.pdf. 1-26.

BIOGRAPHIES

BREMANANTH RAMACHANDRAN was born in India. He had Ph.D. Computer Science in 2008 from Department of CSE, PSG College of Tech., India. He had Post-doctoral Research Fellowship (PDF) from School of EEE, Information Engineering, at Nanyang Technological University (NTU), Singapore, in 2011. He was a Professor and Head in Sri Ramakrishna Engineering college/India Currently, He is an Asst. Professor in the Information Systems and Technology department, Sur University College, Sur, Oman. He received the M N Saha Memorial award for the best application oriented paper in 2006 by Institute of Electronics and Telecommunication Engineers.

AHMAD SHARIEH was born in Jordan. He had PhD in Computer and Information Sciences from Florida State University 1991. He worked as Assistant Professor in Fort Valley College / USA, The University of Jordan (UJ) / Jordan. He worked as Associate Professor with Amman Arab University for Graduate Studies (AAUGS) / Jordan and Dean of KASIT in the University of Jordan. Currently, he is a professor of computer Sciences and Dean at Sur University College (SUC), Oman.

MULTI VARIANT ANALYSIS AND VALIDATION STUDY ON MASTOID TRIANGLE PROCESS FOR SEX DETERMINATION

Janaki Sivakumar,
Research Scholar,
PRIST University,
Tanjore, TamilNadu.
pjanaki78@gmail.com

Prof.K.Thangavelu,
Head of ComputerScience
Periyar University,
Salem, TamilNadu
drktvelu@yahoo.com

Ammar Yassir,
Lecturer Computing Department,
Muscat College,
Sultanate of Oman
al-fayumi@hotmail.com

KEYWORDS
SkullBiometrics, Anthropometry, Computational forensics, Mastoid Process, Discriminant function analysis.

ABSTRACT

The major objectives of this paper are (1) address the factors influencing sexing the human skull through a review of the literature (2) list the general techniques and methods used for sex determination (3) address the mastoid process method in determining sex (4) address the relevance of this information to computational forensic science research and applications. This paper is a summary of findings of Mastoid process sex determination that will give the knowledge that how far skull biometric is useful in determination of sex for unique identification and verification of human. From the available literature in this topic, analysis was done and most relevant accurate method for sex determination was identified. The existing techniques in sex determination are comprehensively reviewed and discussed. Validation of mastoid process approach is done with Matlab (2007) and sex determination was performed with sample database. Experimental results show the effectiveness of this approach in performance.

INTRODUCTION

Historically, human identification is one of the most challenging subjects that man has confronted. The concept of identity is a set of physical characteristics, functional or psychic, normal or pathological, that defines an individual. No two individuals are exactly alike in all their measurable traits, even genetically identical twins (monozygotic) differ in some respects. The determination of the sex of skeletons represents an important stage in the execution of the forensic anthropological examination (Katrin and Sargur 2007). The studies for sex determination are based on the dimorphism between the sexes that is present in the majority of human bones. Computer based methods have become important in forensic science for crime investigation, prosecution and the law. Biometrics is playing a vital role in this area.

COMPUTATIONAL FORENSIC

Computational Forensics (CF) is an emerging research domain. It concerns the investigation of forensic problems using computational methods. The primary goal is the discovery and the advancement of forensic knowledge. CF involves modeling, computer simulation, computer-based analysis and recognition in studying and solving forensic problems (Katrin and Sargur 2007). Computer forensics (also referred to as digital forensics or forensic information technology) is one specific discipline that could use computational science to study digital evidence.
Computational methods find a place in the forensic sciences in several ways, as for example:

- rigorous quantification of individuality
- definition and establishment of likelihood ratio

In recent years, computational forensic has a rapid growth .In this field; algorithms implemented are from the fields of signal and image processing, computer vision, computer graphics, data visualization, statistical pattern recognition, data mining, machine learning, and robotics.

ANTHROPOMETRY

Anthropometry is the study of measurements or proportions of the human body such as dimensions of bones, muscles, and adipose (fat) tissues according to sex, age, etc. for identification purposes.

Cranial Anthropometry also known, as craniometry, is the measurement of the skull and face.3 ways to categorize the skull are

- **Dolichocephalic:** long and thin
- **Brach cephalic:** short and broad
- **Monocephalic:** intermediate length and breadth

Forensic anthropology is the application of the science of physical anthropology and human osteology, the study of the human skeleton (Zeno Geradts and Jurrien). Forensic anthropological techniques can be used to assist in the recovery of assess age, sex and race.

SKULL BIOMETRIC

The skull appears to be the main reliable bone apart from the pelvis exhibility sexually dimorphic features. The human skull also includes 14 facial bones that form the

lower front of the skull and provide the framework for most of the face.

SEX DETERMINATION

Sex can be determined by studying the size and shape of the skull bones by comparing with already established male and female skull databases. Temporal bone plays a vital role in sex determination. For example, the posterior part of the right and left temporal skull bones is large in males than the males.

The published literature on adult head and face, sexing is diverse, spanning topics related to anthropology and biology, medicine and pathology, and computer and forensic science.

Our aim is to integrate key findings from this broad review in as much as the studies reviewed were relatively narrow in focus. It is imperative to understand the measurements, of sexing.

MATERIALS AND METHODS STUDIED

Masato, Yasuhiro, Toshiyuki and Kaoru (1986)

They were studied the shape of skull from roentgen cephalogram based on line drawing technique. They used cephalogram of 50 male skulls and 50 female skulls to analyse. Eminence of Glabella and Nasal root (Figure 1) were picked up as reference points to analyse. The mean radius of circular arc corresponding to the eminence of glabella and nasal root were calculated. They found that the difference was demonstrated to be significant at both the places ($p<0.01$) with smaller radius in male than in female (Hsaio TH and Chang HP 1996)

Similarly the SD of radii at the eminence of glabella in female is 3 times larger than in males. The difference was especially clear at the eminence of glabella where sex could be determined with an accuracy of more than 90%.

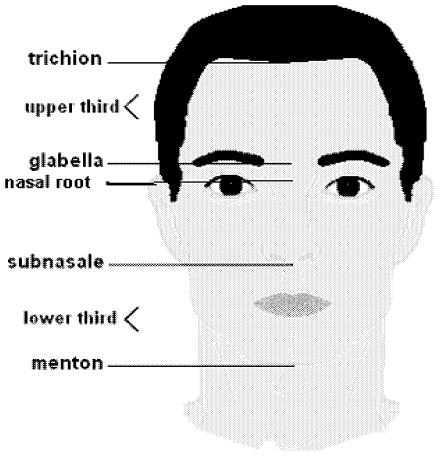


Figure 1: Facial view

Day and Tschabitscher (1998)

They analysed 197 skulls and reported that the mastoid triangle area was a complimentary indicator for the sex determination (Giles E and Elliott 1996).By the use of Discriminant function analysis, they concluded that Mastoid triangle was superior indicator in skull for sex determination.

De Paiva and Sagre (2003)

De Paiva & Sagre analysed 60 skulls (30 male & 30 female) and they used xerographic copy and they marked three landmark points namely Porion, Asterion and Mastoidale.

.

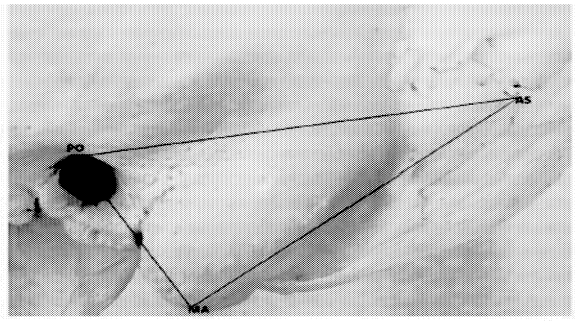


Figure 2:Mastoid Triangle

By using the three points in Figure 2, the demarcated triangle was formed.

Po --- the uppermost lateral point of the external auditory meatus

Ms --- The lowest point Mastoidale

As --- the meeting point of the lamboid, occipitomastoid and pareitomastoid

The area of demarcated triangle for each side of the skull (Left and right) was determined and the total value of these measures was calculated. This study demonstrated the result in three areas that are left area, right area and total area. They concluded that the total area (left & right) could be used of sexing the human skull better than the isolated left area and right area. This result has given 95% accuracy in both the sexes (Sven DE Greef and Guy Williams 2005).

In the same study, they presented the examined results from available literature as

- Temporal bone is the important in sex determination
- Multiple measurements from the mastoid process is superior than an isolated measurements
- Correlation between the surfaces of the mastoid process has given more accurate results.

S.S.Adebesi (2003)

Same like Mastoid triangle some other measurements were also used for determining the sex. S.S.Adebesi studied 350plain radiographs in which 185 are male skulls and 165 are female skulls (Ferembach D, Schwidetzky I,et.al 1999). He measured the following dimensions

- Cranial length &cranial breadth
- Nasal height &Nasal breadth
- Orbital height & Orbital breadth
- Mandibular length & Mandibular breadth
- Facial height.

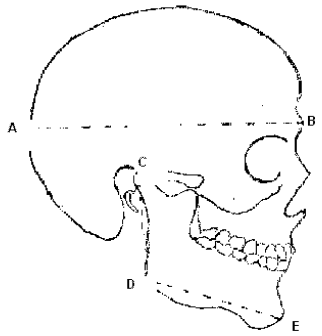


Figure 3: Lateral view of the human skull

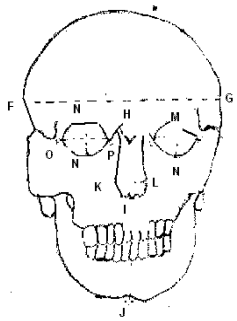


Figure 4: Frontal view of the human skull

Range and mean values were worked out for each parameter in Figure 3 and Figure 4 in both the sexes. T-test ($p<0.05$) was used to compare the values. Indices according to Najjar and McWilliams were calculated (Sant SM Chinmmalgi and M, KulKarni Y, 2007). This study became the extended study of Najjar and McWilliams. In this study almost all dimensions recorded higher values in the male skull than in the female skull.

Suazo, Zavando & smith (2008)

Based on the study of Pavia &Segre (2003), mastoid triangle and lineal dimensions with Fischer's lineal discriminant functions were analysed. They used 81 human skulls (50 male & 31 female). Among the lineal dimensions, higher values were obtained only in Po-Ma in female groups (Hsaio TH and Chang HP 1996).

In the first work of De Paiva and Sagre, they used 2D plane image by means of Xerographs to measure the dimensions. But Suazo, Zayando & smith used direct measurements by using callipers from the skull. So accuracy in lineal dimensions is more in the later.

EXPERIMENTAL RESULTS

Based on Paiva and Sagre, to evaluate the effectiveness, database of 15 landmark data from 23 skulls (10 Female, 13 Male) were collected from the Department of Forensic science, Trichirapalli, TN. From the database, only three landmarks were chosen to investigate and evaluate the performance of Mastoid triangle.

The chosen landmarks are

- Porion (Po)
- Mastoid ale (Ms)
- Asterion (As
-

In order to identify statistical measurements, Matlab 2007, was used. According to Paiva & Sagre, three distances Po-Ms, Ms-As and As-Po is calculated. The triangle area is calculated using the formula with the sides of length l, b, and h:

$$\text{Area} = \sqrt{(S\,(S-l)\,(S-b)\,(S-h))} \qquad (1)$$

$$\text{Where} \quad S = (l + b + h)/2$$

Mean and Standard Deviation were calculated for the 23 samples and summary statistics are given in Table 1.

According to Paiva & Sagre, The Total area of left and Right Mastoid Triangle >= 1447mm^2 and <= 1260 mm^2 characteristics male and female sex respectively.

Table 1 - Statistical Measurements from Male and Female Samples

Measurements from Forensic Samples (Right side)	Mean		Standard Deviation	
	Male	Female	Male	Female
Po-Ms	31.2	28.65	3.4	3.1
Ms-As	50.0	47.6	5.0	5.1
As-Po	48.15	45.7	3.6	3.3
Area	718.25	632.7	106.4	107.4

From the laboratory results, Mastoid process triangle approach has significant, positive performance results in both Male samples (93% confidence) and female samples (95 % confidence). The results show that our approach has an encouraging performance.

Also Table 2 summarizes T-test evaluation performance to add strength to the mastoid process

Measurements from Forensic Samples (Right side)	T-test	
	Male	Female
Po-Ms	0.291	0.693
Ms-As	-0.613	-0.113
As-Po	-0.558	-0.152
Area	-0.111	0.295

Table 2 – T - test: Side Differences by Male and Female

Overall performance of this approach has given improvements in performance and useful for further research in the same.

DISCUSSION

The objective of this study is to demonstrate through practical method that triangle area (based on anthropometric techniques) of Mastoid process is useful in gender identification. From these practical measurements and several literatures reviewed, we can conclude the following

- ✓ Skull is the major biometric to determine the gender
- ✓ Importance of Mastoid process for the purpose of gender identification
- ✓ Mastoid triangle area is significantly more in males than in females
- ✓ The related statistical values to the mastoid process obtained by anthropometric techniques that better demonstrate dimorphism between the genders.
- ✓ Multiple measurements will give effective evolution rather than isolated measurements
- ✓ Accuracy of gender classification obtained from mastoid process is better than previous works
- ✓ Easy to execute, gives fast results and confidence up to 90%
- ✓ Meets the needs of forensic science department especially in Computational Forensic Science.

FUTURE WORK

Based on the success in this practical implementation, in future, we aim to determine the gender from the human skull image (CR X-Ray) for the forensic science applications. In our research work, four main areas need to be analyzed to reach the final goal. Firstly, the image needs to be enhanced to get a better visual quality for further work. Secondly, from the enhanced image, feature extraction needs to be done. Thirdly, identifying landmark locations from the extracted features of skull image needs to be achieved. By using the hybrid method of Mastoid Process, gender identification is the final goal of our research work

CONCLUSION

In this paper, the impact of Mastoid Process Triangle approach for sex determination is analyzed and discussed. Among various techniques, which were developed for sex determination, Mastoid process shows effective results. With my experimental results I conclude that mastoid process will give improvements in performance. Therefore, In future, Digitalization of 3D- skull and creating the database for different orientations like scaling, rotating could be done to achieve greater results in sex determination.

The objective of this study was to demonstrate the different methodologies through a complete readily available literature review. Therefore, compared with the most important historical studies dealing with sex determination of skulls, the present literature review shows that mastoid process with lineal dimensions is an improved method. This paper (1) provided an up-to-date summary of examples from the literature on sex determination with a discussion detailing particular methods, and (2) obviated the benefits of an interdisciplinary approach to literature reviews; specifically in sex determination is meaningful to computer science research and forensic science applications.

ACKNOWLEDGEMENT

I would like to thank Mr. Manickam (ADSP) for providing the guidance in the preliminary discussions and to collect database, in order to proceed further in my area of research.

REFERENCES

Amirhosein Nabatchain,"3D Face Recognition Approaches and Challenges",Research Center for Integrated MicroSystems-University of Windsor,August,2008.

Ariane Kemkesand Tanja Gobel. " Metric Assessment of the "Mastoid Traingle " for sex Determiantion: A Validation Study,J Forensic Science ,2006

Das Gupta A, Aridom Banerjee Anil kumar Sambasiva Rao & Josna Jose .Discriminant Function Analysis of Mastoid Measurements in Sex Determiantion, J Life Sci ,2012

Day, J. D.; Tschabitscher, M. Anatomic position of the asterion. *Neurosurgery, 42*:198-9, 1998.

Ferembach D,Schwidetzky I,Stloukal M. Recommendations for age and sex diagnosis of skeleton. J Hum Evol 1999.

Giles E,Elliott O.Sex determination by discriminant function analysis of crania.Am J Phys Anthrop 1998

Hsaio TH,Chang Hp,Liu KM.Sex determination by discriminant function analysis of lateral radiographic cephalometry.J Forensic Sci 1996.

K.selvam,Dr.B.Poorna,"Recognition of Human Face using Mandible-Ramus-NasalBone",IJCSNS,VOL 9 No 4,April 2009.

Katrin Franke, Sargur N.Srihari, "Computational Forensics:Towards Hybrid Intelligent crime Investigation", *Manchester,UK,2007*.

M.Funayama,Yasuhiro Aoki,Toshiyuki Kudo and Karou Sagisaka,Sex Determination Of the human Skull Based upon Line Drawing from Roentgen Cephalograms,1986.

Maria VaneZis ," Fornsic Facial Reconstruction using 3-D computer Graphics: Evaluation and improvement of its reliability in Identification",University of Glasgow,October 2007.

Paiva LAS *et al* .Sexing the human skull through the mastoid process.REV. HOSP. CLÍN. FAC. MED. S. PAULO 58(1):15-20, 2003.

Peter Claes ,Dirk Vandermeulen,Sven De greef,Guy Willems and Paul Suetens,"Craniofacial Reconstruction using a combined statistical model of face shape and soft Tissue depths: Methodology and Validation",Forensic Science Journal .159S,March,2006.

S.A.Hameed,B.B.Zaidan,A.a.Zaidan,A.W.Naji,Omar Farooq,"Novel Simulation Framework of 3Dimentional Skull Bio-Metric Measurement,IJCSE,VOL 1 No 3,2009.

Sant SM Chinmmalgi M,KulKarni Y . Sexing of Skull by new metrical parameters , ,Journal Of Anat.Soc.India 56(1) 28-32,2007

Sven DE Greef and Guy Williams " Three Dimentional Cranio Facial reconstruction in Forensic Identification:Latest Progress and New Tendencies in 21st century",Journal of Forensic Science,Vol 50,No 1,2005.

Zeno Geradts and Jurrien Bijhold,"Data Mining in Forensic Image DataBases",Netherlands Fornsic Institute.

TRANSPORT AND MAC CROSS-LAYER PROTOCOL

FOR VIDEO SURVEILLANCE OVER WIMAX

Suherman
Faculty of Technology
De Montfort University
Leicester, UK
Email: suherman@usu.ac.id

Marwan Al-Akaidi
Arab Open University
P.O.Box: 3322 Safat
13033 Kuwait
Email:m.al-akaidi@arabou.edu.kw

Raouf Hamzaoui
Faculty of Technology
De Montfort University
Leicester, UK
Email: rhamzaoui@dmu.ac.uk

KEYWORDS
Transport layer protocol, Cross-layer protocol, inter-frame retransmission, video surveillance, WiMAX.

ABSTRACT

Video surveillance is an emerging application for activity and security monitoring. Outdoor surveillance applications can take advantage of a WiMAX network to provide installation flexibility and mobility. A WiMAX-based surveillance system can be implemented as a dedicated network which only serves surveillance nodes to ensure high reliability. However, wireless video transmission is prone to interferences which degrade video quality. This paper proposes a novel transport and MAC cross-layer (TMC) protocol which aims at reducing delay and increasing video quality by integrating a transport layer protocol and bandwidth allocation within WiMAX. The simulations show that the proposed protocol outperforms existing protocols.

INTRODUCTION

Most CCTV and IP surveillance systems use coaxial and Ethernet cable networks for indoor surveillance. Outdoor surveillance systems rely on wireless LAN and point to point radio technologies. Although research on the use of cellular networks for surveillance application exists, real implementations are hardly found since the channel bandwidth is limited. WiMAX (Worldwide interoperability for Microwave Access) is a wireless broadband technology that offers higher capacity than Wi-Fi networks and wider coverage than cellular networks. WiMAX has experienced intensive development from fixed wireless applications, mobile WiMAX, up to standard with 4G capabilities. This makes WiMAX a promising technology for video network infrastructures. Surveillance applications such as multi surveillance cameras placed on high rooftop buildings in urban areas and rural surveillance have the potential to be implemented in aWiMAX network.

Since bandwidth allocation in WiMAX is application dependent, various scheduling techniques have been proposed (Dhrona et al. 2008) to improve the application performance. Each application, including video surveillance, requires particular scheduling and bandwidth request methods. Suitable bandwidth allocation leads to high WiMAX link performance. On the other hand, the transport layer protocol determines the end to end performance as it provides packet transmission for host to host applications. A high performance link provided by WiMAX will not be optimal if the chosen transport layer protocol is poor. TCP (Transmission Control Protocol) provides high reliability data transfer which ensures that each packet is received successfully and sequentially. It guarantees the quality of delivered video. However, routine acknowledgements and retransmissions in TCP generate a significant delay which is not suitable for real time applications. Furthermore, interferences and signal disruption in the wireless channel may cause TCP to experience significant delay as it keeps trying to resend the lost packets. In contrast, UDP (User Datagram Protocol) is able to reduce delay in video delivery. The drawback is that UDP does not respond to network conditions as it keeps sending data regardless of network congestion. UDP potentially makes the congested network even worse. Various transport protocols have been proposed to enhance protocol performances. An overview of these protocols is given in the next section.

This paper combines a transport layer protocol and bandwidth allocation in WiMAX to achieve better video surveillance performance. The transport and MAC cross-layer term refers to a method that explores interactions between the MAC layer and the transport layer. Our proposed method is aimed at enabling the MAC layer to support retransmission in the transport layer. We avoid two-way interactions to prevent processing delays. Instead, the proposed method enables the MAC layer to read the transport layer header in order to provide service to the transport layer. Besides performance improvement, the proposed method requires only a minor change in the MAC layer and the WiMAX device is still compatible with other implementations.

The rest of the paper is organized as follows. After discussing related works, we present the proposed method. Then the proposed method and the existing solutions are compared using ns-2 simulations. The performance of the proposed protocol is initially compared to existing transport layer protocols using the same bandwidth allocation. The improvement is then examined for various bandwidth allocations. Finally, the conclusion section summarizes the contribution and suggests future work.

RELATED WORKS

The TCP/IP protocol stack defines four independent abstraction layers for IP based networks. . Data is passed from one layer to the other by using header encapsulation and de-capsulation. The idea of layer separation may work well for wired communication, but not in a wireless environment where device characteristics and channel quality often vary. Several cross-layer solutions have been proposed for wireless communication.

Raisinghani and Iyer (2004) outlined cross-layer possibility in different layers. In WiMAX implementation, cross-layer approaches mostly occur between PHY and MAC layers. For instance, Noordin and Markarian (2007) implemented a cross-layer optimizer between MAC and PHY layers to maximize WiMAX performance. The optimizer collects data from both layers and returns the optimized parameters for bandwidth allocation in MAC layer as well as coding selection in PHY layer. Cross-layer approach can also be performed between the MAC layer and the network layer. This was used in (Mohanty and Akyildiz, 2007) to provide seamless handover. Meddour et al. (2011) implemented a cross-layer approach between MAC and application layer to optimize unicast and multicast video streaming in WiMAX network. Our work completes the cross-layer schemes by proposing the cross-layer MAC and transport layer protocol. The proposed Transport MAC cross-layer protocol provides high performance end to end transport layer connection in WiMAX network that can replace the existing UDP protocol. Unlike the aforementioned cross-layer schemas, our proposed cross-layer schema does not require a new protocol data unit (PDU) or a separate layer entity. The proposed scheme uses existing PDU and entities. Therefore, the cross-layer design is much simpler and fast. However, the protocol does not aim to compete against the existing schemes, as each cross-layer design has a different emphasis. The MAC-PHY, MAC-Application, and the proposed cross-layer design could be combined to achieve the expected performances.

Cross-layer design between MAC and transport layer protocol has been explicitly used in some existing reliable transport layer protocols which employ congestion control. Ye, Wang and Huang (2011) used the cross-layer method to provide fairness for some TCP flows. Work by Zhai et al.(2007) proposed WCCP (Wireless Congestion Control Protocol) which is effective only for static ad hoc network. WCCP adjusts sending rate based on channel utilization. However, reliable based protocols are not suitable for multi sources real-time video transmission over WiMAX as those protocols exert tremendous delay (Larzon et al., 1999). Our proposed cross-layer design does not explore channel quality to support congestion control as reliable protocols did. The protocol is intended to improve existing unreliable protocols; therefore the implemented methods should not change the nature of the unreliable protocol. The congestion avoidance is performed as simple as possible and the retransmission effort is performed only once.

Various works have proposed improvements on unreliable protocol performance. Reliable UDP (RUDP) adds congestion control mechanism, acknowledgement, and retransmission services to accommodate different transport protocol requirements (Bova and Krivoruchka, 1999). This protocol works between UDP and TCP. However, the excessive features make RUDP behave almost like TCP and remove the nature of unreliable protocol. UDP-Lite (Larzon et al., 1999) implements a partial checksum for the sensitive part and ignores errors in the non-sensitive part of the UDP packets. UDP-Lite performs better than UDP in terms of packet loss. However, it disables network supervision in upper layer as it masks error on transmission (Welzl, 2005). UDP-Lite requires additional processing time to determine whether data needs checksum, as well as to process it in the receiver. The ignored packet passed to application layer may not be acceptable. UDT (UDP-based Data Transfer) (Gu and Grossman, 2007) and RBUDP (Reliable Blast UDP) (He et al., 2002) are datagram protocols that work for high speed bulk data transfer link. Both protocols were aimed to solve TCP weakness which underutilize high speed network (Gu and Grossman, 2007). RBUDP employs negative acknowledgement which sends a TCP request-reply to acknowledge lost packets in a UDP based bulk transfer. UDT and RUBP are intended for single high speed link connection, which may perform worse in multiple traffics environment. BTP (Bidirectional Transport Protocol) modified UDP for tele-controlled robot application using inter-packet gap (IPG) congestion control (Wirz et al., 2009). The inter-packet gap determines the speed of data transfer. ERT (Embedded Reliable Transport protocol) added additional header on UDP to provide reliability for embedded application (Wei and Chao, 2010).

Kohler et al. (2006) proposed DCCP (Datagram Congestion Control Protocol) which employs two congestion controls; TCP-Like and TFRC-Like. DCCP is a potential transport protocol to replace UDP. However, DCCP does not retransmit lost packets and relies fully on client monitoring feedback. SCTP (Stream Control Transmission Protocol) (Stewart, 2007) was initially designed for reliable signalling and control transport protocol for telecommunications traffic running over IP networks. SCTP provides multi-homing features which enable alternative transmission path. However, SCTP performance is worse than that of DCCP for real-time video transmission (Chughtai et al., 2009). Ali etal(2011) proposed a semi-reliable transport protocol called Broadband Video Streaming (BVS). The protocol applies retransmission as soon as packet loss is detected. Our previous work (Suherman et al., 2011) has shown that inter-frame retransmission is able to improve the performance of video transmission in WiMAX. Inter-frame retransmission resends the lost prioritized packets at the end of each frame transmission. The transport layer part of our proposed protocol uses inter-frame retransmission.

PROPOSED PROTOCOL

We assume a dedicated surveillance network which operates in non-saturated conditions and every node generates the same video bit rates. The proposed protocol aims at minimizing the delay and maximizing video quality. It consists of two parts, transport layer and MAC layer. The transport layer part uses inter-frame retransmission with congestion delay. The second part enables the MAC layer to assist the transport layer by providing sufficient bandwidth for the retransmitted packets.
The transport layer part aims to improve the reliability of the protocol. Therefore, we employed a transport layer protocol with simple congestion control and retransmission scheme without repetition for the lost retransmitted packet. Additional bandwidth given by the cross-layer scheme in the MAC layer is the primary feature of the proposed protocol.

Transport layer part

Transport layer protocols that employ negative acknowledgement (NACK) use either quick or delayed response. In quick response, the receiver notifies the sender with a NACK packet as soon as packet loss is detected. The sender then retransmits the requested packets. For example, BVS (Ali et al., 2011) is a quick response retransmission protocol, while RBUDP (He et al., 2002) is a delayed response protocol. Inter-frame retransmission uses a delayed NACK

response to acknowledge lost packets. The NACK packet is sent after receiving the last packet within one frame. The objective is to avoid multiple acknowledgements for multiple losses in one video frame. Inter-frame retransmission also aims at smoothing the network load by sending the NACK in idle time (inter-frame gap (IFG)).

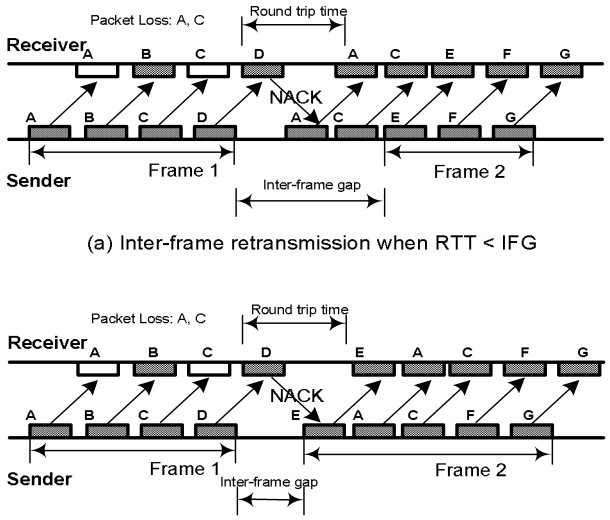


Figure 1: Inter-frame retransmission

Figure 1 shows how inter-frame retransmission works. We assume that packet A and packet C within frame 1 are lost. The receiver requests retransmission to the sender after receiving the last packet within frame 1 (packet D). Soon after receiving the NACK packet, the sender retransmits the requested packets. In case packet D is lost, the NACK packet is sent after receiving the next packet. If the round trip time (RTT) is smaller than IFG, the retransmission occurs in IFG. Otherwise, the retransmitted packets compete with the packets from the next frame. Therefore, inter-frame retransmission is suitable for real time video with a small number of intermediate nodes, as in video surveillance.

MAC layer part

The MAC layer part is responsible to ensure that the retransmitted packets have sufficient bandwidth. The additional MAC functionality detects NACK packet and reads its content. Based on the NACK information, MAC allocates additional bytes in bandwidth request to accommodate the retransmitted packets in the transport layer.

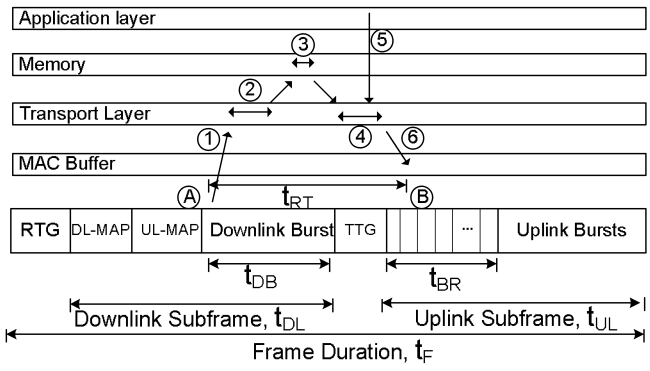


Figure 2: NACK and retransmitted packet flows

Figure 2 illustrates the need of the MAC layer part. We assume the NACK packet is received by subscriber station (SS) in the beginning of a downlink burst in a downlink sub-frame (A). MAC forwards the NACK packet to the transport layer (1). The transport layer protocol processes the packet (2), retrieves the requested packets from the memory (3) and encapsulates the retransmitted packets (4). Since the application layer periodically sends new packets (5), the transport layer may experience congestion that leads to retransmission failure. Packets which reach the MAC layer are queuing in the MAC buffer (6) before being transmitted. Bandwidth request is performed based on the number of bytes in the MAC buffer.
In order to minimize the delay for the retransmitted packets, the access time for the retransmitted packets, t_{RT}, should be as small as possible so that the packets has been in the MAC buffer by the time the nearest bandwidth request is made. Since SS bandwidth request opportunity is randomly chosen within the bandwidth request period t_{BR}, packets should be in the queue at the latest just after TTG (transmit/receive transition gap). In the best case when NACK is received in point A, t_{RT} should be less than $(t_{DB} + \text{TTG})$, where t_{DB} is the downlink burst duration. In the worst case when NACK is received in point B, t_{RT} should be less than TTG. Otherwise, the retransmitted packet will miss the nearest bandwidth request opportunity and must wait for another bandwidth request opportunity which leads to an additional delay of at least one full frame duration, t_F.

By using MAC functionality, the nearest bandwidth request should not wait for the retransmitted packet arriving in the queue to add bandwidth allocation request. Instead, the MAC layer adds additional tasks. First, MAC reads the NACK packet content to determine the number of requested packets. Then, MAC informs the bandwidth request module to add additional bytes in incoming bandwidth request packet. As a result, the requested bandwidth includes the retransmitted packets although they do not appear in the MAC buffer yet.

Since the NACK packet flows through base station (BS) to SS, the MAC functionality for the proposed TMC protocol can be implemented in either BS or SS. The advantage of the SS implementation is that the additional bandwidth is allocated after NACK packet is safely received. On the other hand, BS can allocate additional bandwidth directly without waiting for bandwidth request from SS. However, BS implementation may decrease network performance as BS will have more tasks. Moreover, if NACK packet is lost, then bandwidth is wasted.

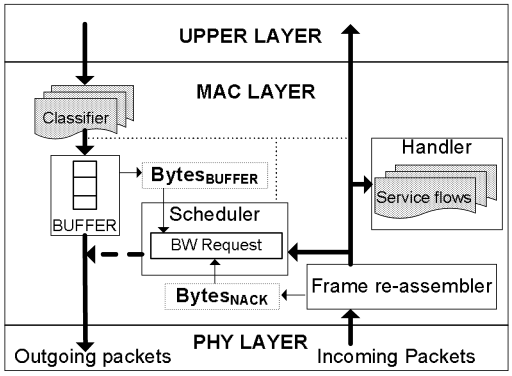
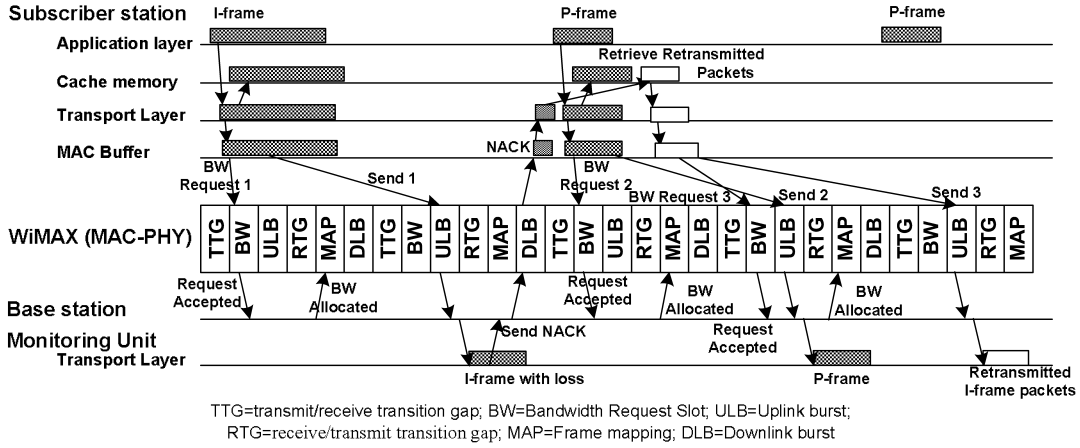


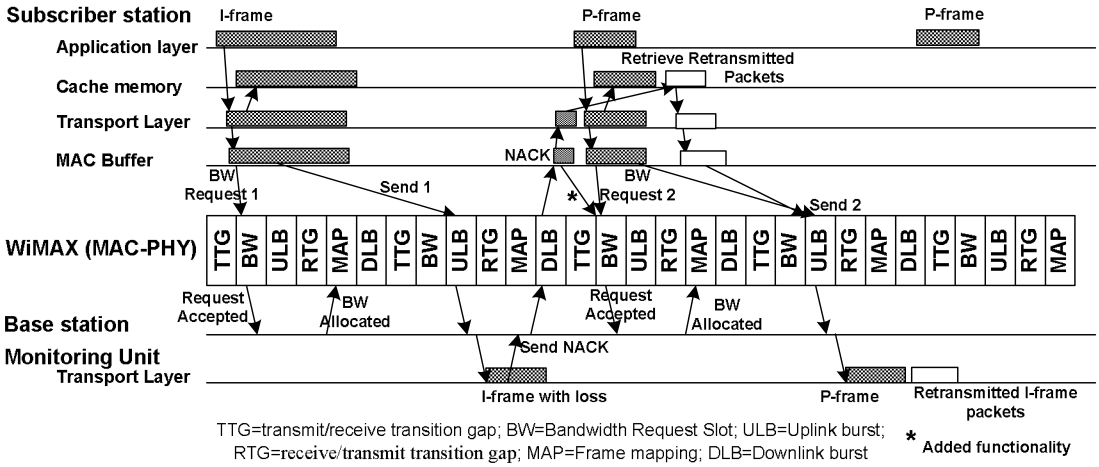
Figure 3: MAC layer implementation

Figure 3 shows the SS implementation of the proposed cross-layer design in the MAC layer part. The scheme is based on the NIST WiMAX module (NIST, 2007). The frame re-assembler in the MAC layer reads the NACK packet and notifies the scheduler to add the number of requested bytes (BytesNACK) in the bandwidth request. In turn, the scheduler sends a bandwidth request based on data size on MAC buffer (BytesBUFFER) and the retransmission bytes (BytesNACK).

Figure 4shows a comparison of simplified layer interactions of the inter-frame retransmission without and with MAC cross-layer functionality. In Figure 4 (a), the bandwidth for the retransmitted packets is separately requested as the packets are not available by the time the SS sends a bandwidth request to BS. Consequently, instead of sending the retransmitted data in the nearest uplink burst, SS will allocate it to the uplink burst after the next burst. This postponement increases the packet delay. On the other hand, the MAC cross-layer protocol accelerates packet retransmission as the earliest bandwidth request accommodates the retransmitted packets. The retransmitted and the current data are sent in the same burst.



(a) Inter-frame retransmission without MAC cross-layer functionality



(b) Inter-frame retransmission with MAC cross-layer functionality

Figure 4: Subscriber station based MAC cross-layer protocol

SIMULATION ENVIRONMENT

In order to evaluate the proposed methods for a dedicated video surveillance network, we conducted simulations using the ns-2 simulator with the WiMAX module taken from NIST (2007). The transmit power and receiver thresholds are set to provide 1000 m coverage radius. The modulation is 64 QAM, with a two-ray ground propagation model. The downlink/uplink ratio is 0.3.The simulated surveillance application has 4 mobile nodes. Node 0 is fixed (0 m/s). Node 1 is at walking speed of 1.39 m/s. Node 2 and Node 3 are assumed to be in a public transportation. Node 2 moves at 4.44 m/s and Node 3 at 6.67 m/s. The network configuration is shown in Figure 5.

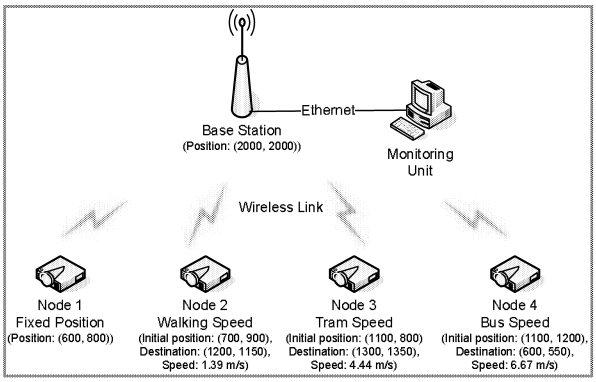
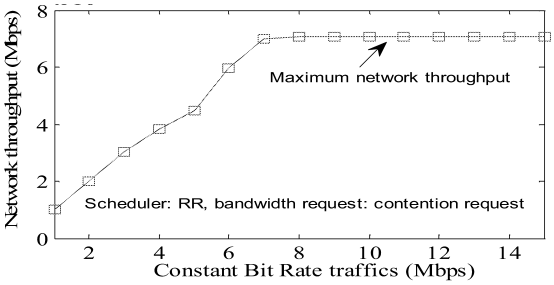
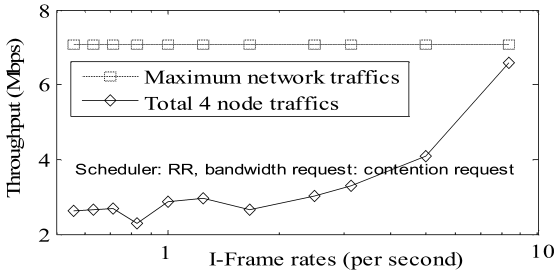


Figure 5: Network configuration

The number of mobile nodes was chosen to simulate a non-saturated network, which means that the traffic load is smaller than the network resources. This is important as the surveillance network should provide sufficient bandwidth in order to maintain video quality. By using constant bit rate (CBR) tests from 1 to 15 Mbps, we obtained a saturated uplink bandwidth of 7 Mbps (Figure 6a). Since the proposed methods deal with packet/frame types, we increase the traffic load based on I-frame rate (GOP) instead of number of SSs. The total video rate for each simulation is depicted in Figure 6b.



(a)



(b)

Figure 6: Traffic rate in simulation

The traffic sources were generated from the akiyo_cif.yuv video. Its video trace was used as simulated traffics in the ns-2 simulations, where the received patterns were reconstructed based on the original video. The traffic generation and reconstruction in the ns-2 simulator were based on the Evalvid video evaluation framework from (Klaue et al., 2003). The prioritized frames were set for I-frames and the transport layer protocol used was UDP. Table 1 shows the traffic parameters.

Table 1: Simulated traffic parameters

Parameter	Value
Video sequence	akiyo_cif.yuv
Frame rate/type	30fps/IPP
Video codec	MPEG4
Video bit rate	559.35 Kbps for GOP of 30 frames
Group of Pictures	3, 5, 8, 10, 15, 20, 25,30, 35, 40, 45

The performance evaluation was conducted by observing sending and receiving ports in each connection. The measurement in the ns-2 simulator refers to those in (Ke et al., 2008).The main performance metrics are the average of delay and PSNR (Peak Signal to Noise Ratio) of the four nodes. Measurement points are in SSs (sender) and in the monitoring unit (receiver). PSNR is obtained by reconstructing video from the received packets and comparing it to the original source.

First, we evaluated the performances of the TMC protocol using round robin scheduler with contention request. We did the same experiment for inter-frame retransmission (IR) without cross-layer (Suherman et al., 2011), BVS(Ali et al., 2011) and UDP. Afterwards, we applied the protocol for various scheduling algorithms to confirm the superiority of the proposed method.

RESULTS AND ANALYSIS

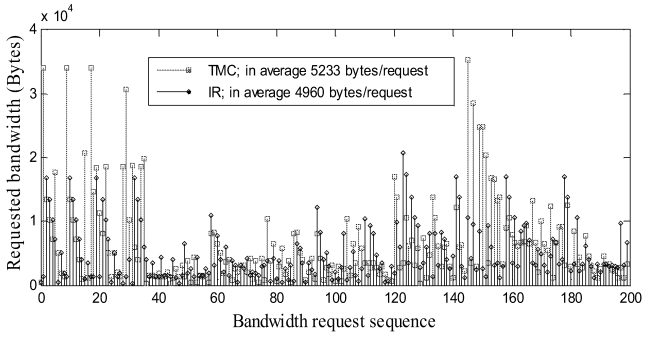
The impact of MAC cross-layer

Transport layer packets queue in the MAC buffer of the SS before being transported by the physical layer. MAC transfers the data to the uplink sub-frame based on the duration allocated by BS in UL-MAP. The duration itself is decided by BS based on SS bandwidth request and the available bandwidth. Since the main feature of the MAC cross-layer is additional bandwidth allocation for the retransmitted packets, the proposed protocol gains higher bandwidth than the basic IR protocol.

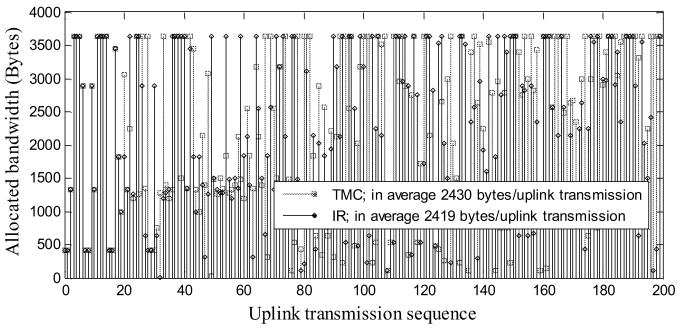
Table 2: Allocated bandwidth comparison (GOP 30)

Protocol	IR	TMC
Number of bandwidth requests	1270	1268
Average requested bandwidth	4960	5233
Number of uplink transmission	1530	1522
Average allocated bandwidth	2419	2430
Network utility	55.29%	55.54%

For the simulated traffics with GOP 30, TMC generates 1268 bandwidth requests, while IR produces 1270 requests (Table 2). In average, TMC requested bandwidth 273 bytes morethan IR. From those requests, BS allocates in average 2430 bytes/uplink transmission for TMC and 2419 bytes/uplink transmission for IR. TMC uses the network better than IR. Since the frame duration is 5ms and the maximum network throughput is 7Mbps, the network utility of the TMC protocol is equal to (2430x8/0.005)/7000000x100%=55.54%. IR utility is 55.29%. Figure 7 shows the requested and the allocated bandwidth for the first 200 bandwidth requests.



(a)



(b)

Figure 7: Bandwidth request comparison

Since the additional bandwidth is requested before the retransmitted packets available in MAC buffer, the allocated bandwidth can be used by regular data, even if the retransmitted packets failed to be retrieved. The higher bandwidth allocation and network utility in the proposed protocol produce lower delay and higher video quality. Figure 8 shows the performance comparisons between IR and TMC. TMC consistently reduces packet delay for all I-frame rates. Although the PSNR decreases when sending data with I-frame rate 1 fps, this is probably caused by the undecodable subsequent error frames.

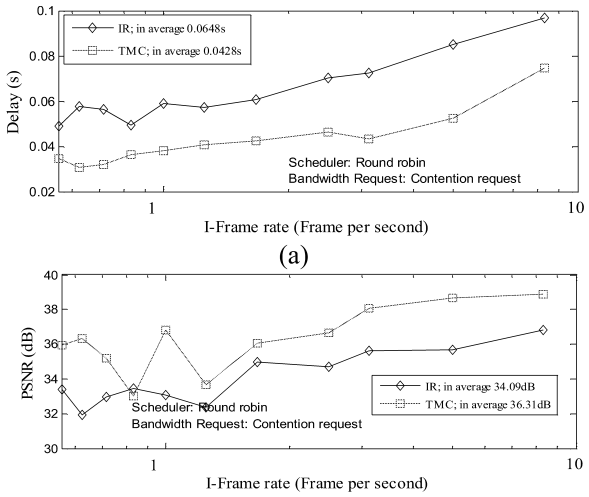


Figure 8: Performance comparison between IR and TMC

Transport layer protocol comparison

Figure 9 compares TMC to existing protocols. TMC was able to reduce UDP delay by 18 to 37%. The PSNR improvements were around 14.3 to 149.5%, 12.6 to 150.2%, 21.3 to 184.3% and 17.9 to 120.2.3% over IR, BVS, UDP and DCCP, respectively. Other existing protocols such as SCTP and RBUDP are not presented as they have been compared in (Chughtai et al., 2009). The result shows that TMC outperforms the existing protocols for surveillance application over WiMAX with uniform traffics.

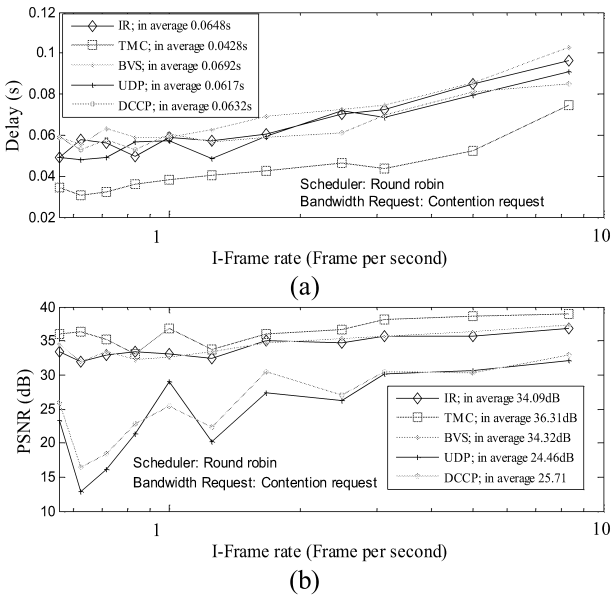


Figure 9: Performance comparison between TMC and other protocols

TMC has lower delay than UDP because it requested more bandwidth when loss occurred. As shown in Figure 10, TMC received more bandwidth than other protocols. TMC experienced lower allocation than BVS for high I-frame rates as the maximum network throughput (Figure 6b) limits the bandwidth for the retransmitted packets. However, the limited bandwidth does not reduce TMC performance as the cross-layer functionality still produces better allocation.

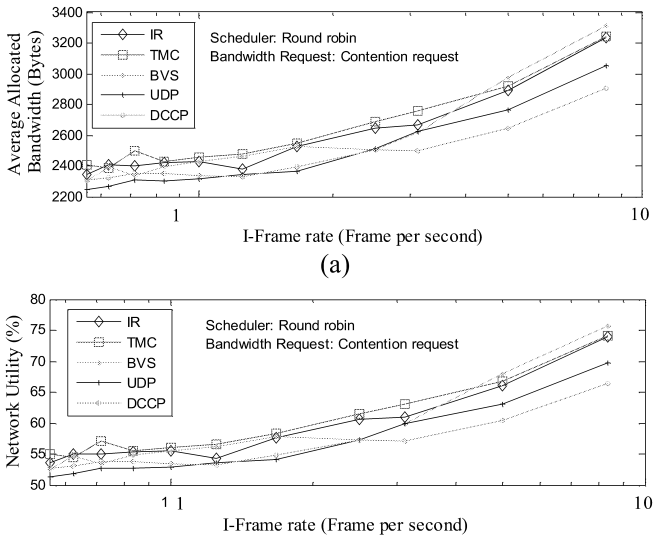


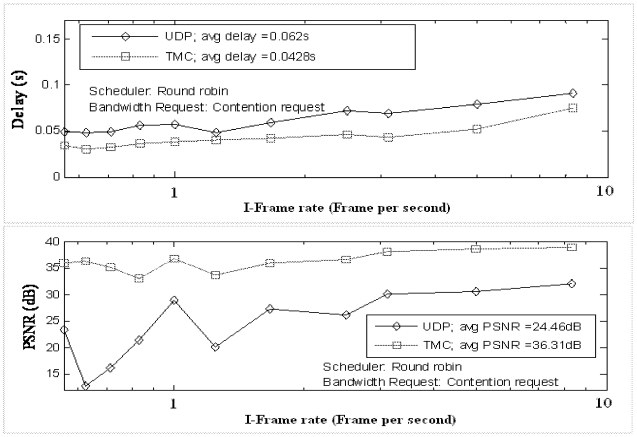
Figure 10: Comparison of the allocated bandwidth and network utility

On the other hand, although BVS received higher bandwidth for higher I-Frame rates, bandwidth may be wasted as multiple NACKs may disturb regular packet transmission. UDP and DCCP suffer low bandwidth allocation as both protocols do not retransmit lost packets. UDP does nothing to increase utility.

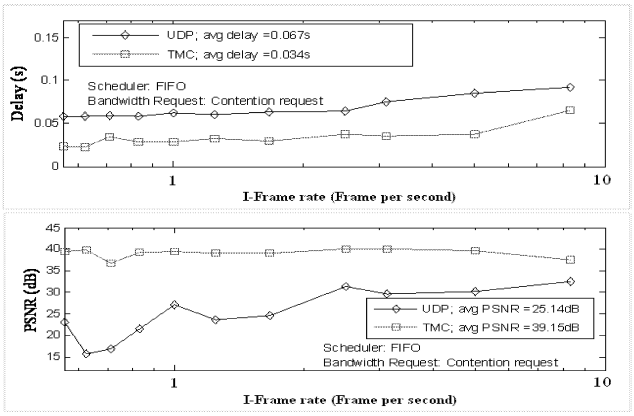
Protocol performance over various schedulers

In order to ensure that the proposed protocol is suitable for various WiMAX schedulers, we evaluated it with Round Robin (RR), First In First Out (FIFO) (Dhrona et al., 2008), Frame based (Kang and Zakhor, 2002), and the Earliest Deadline First (EDF) (Ferrari and Verma, 1990) schedulers for dedicated video surveillance over WiMAX (Figure 11).

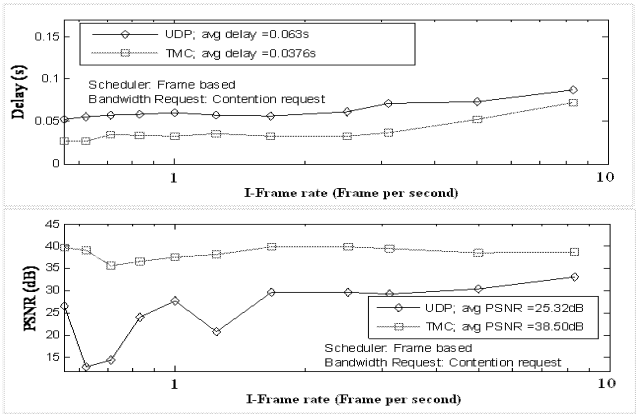
The proposed protocol applied with RR, FIFO and frame based schedulers significantly reduced the delay and increased the PSNR. On the other hand, the implementation of the protocol with the EDF scheduler experienced irregular delays. The reason is that the EDF scheduler is not suitable for applications with uniform traffics as the traffics have similar behaviour and deadlines, while the EDF scheduler classifies the allocated data based on traffic deadlines. As a result, BS performs unnecessary sorting which introduces delay. Although TMC failed to reduce the delay for several I-frame rates, it consistently increased the PSNR values.



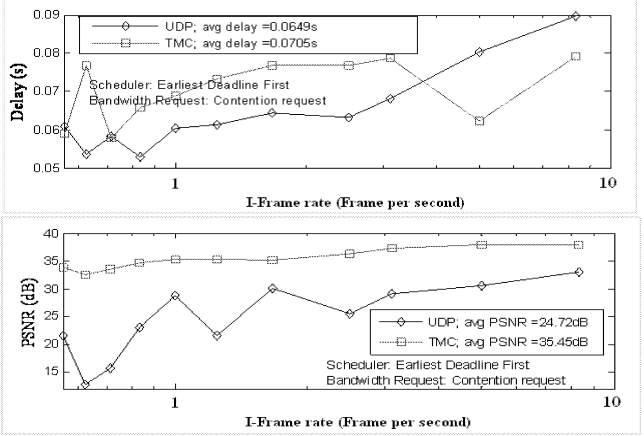
a. Round robin scheduler



b. FIFO scheduler



c. Frame based scheduler



d. EDF scheduler

Figure 11: TMC performance over various schedulers

CONCLUSION AND FUTURE WORK

This paper has proposed a transport and MAC cross-layer (TMC) protocol for a dedicated video surveillance network using WiMAX. The proposed protocol has two components that work separately in two layers. The inter-frame retransmission is used in the transport layer, while the MAC layer adds the capability to read the NACK packet content and uses the information to increase the number of bytes in bandwidth request. The simulations show that the proposed protocol outperforms existing transport layer protocols for WiMAX based video surveillance. It is able to achieve lower delay than UDP and better video quality than other protocols.

The proposed protocol is proven to work well with various scheduling algorithms. The use of the proposed protocol combined with a suitable scheduler improves a WiMAX application for a dedicated surveillance network with real-time video traffic. Further work will be carried out to assess the protocol performance for more general network settings as the simulations in this paper have limited bandwidth.

ACKNOWLEDGEMENTS

This work was supported by Directorate General of Higher Education (DGHE or DIKTI), Ministry of National Education, Indonesia.

REFERENCES

Ali I.; S. Al-Majeed; M. Fleury; and M. Ghanbari. 2011. "Semi-reliable transport protocol for IPTV over mobile WiMAX", in Proceeding of Computer as a Tool Conference (Lisbon, Portugal, April 27-29), IEEE R8, Portugal, 1-4.

Bova T. and T. Krivoruchka. 1999. “Reliable UDP Protocol”, Internet Engineering Task Force (Network Working Group) draft, Cisco Systems, Herdon, VA (February).

Chughtai, H. M. O; S. A. Malik; and M. Yousaf. 2009. "Performance evaluation of transport layer protocols for video traffic over WiMAX", in Proceeding of the 3th Multitopic Conference (Islamabad, Pakistan, Dec 14-15), 1-6.

Dhrona, P.; N. A. Ali; H. Hassanein. 2008. "A performance study of scheduling algorithms in Point-to-Multipoint WiMAX networks", in the Proceedings of Local Computer Networks Conference, (Montreal, Que., Oct.) , 843-850.

Ferrari, D. and D. Verma. 1990. “A scheme for real-time channel establishment in wide-area networks”, IEEE Journal on Selected Areas in Communications, , 368-379.

Gu Y. and R. L. Grossman. 2007. "UDT: UDP-based Data Transfer for High-Speed Wide Area Networks", Journal of Computer Networks, vol. 51, issue 7 (May), 1777-1799.

He E.; J. Leigh; O. Yu; and T. A. Defanti. 2002. "Reliable Blast UDP: predictable high performance bulk data transfer", in Proceeding of Cluster Computing Conference (Chicago, Illinois, Sept.), 317-324.

Kang S.-H.; A. Zakhor. 2002. “Packet scheduling algorithm for wireless video streaming,” in Proceeding of International Packet Video Workshop, (Pittsburgh, PY, April).

Ke C.-H.; C.-K.Shieh; W.-S.Hwang; and A. Ziviani. 2008. “An Evaluation Framework for More Realistic Simulations of MPEG Video Transmission”, Journal of Information Science and Engineering, vol. 24, no. 2, (March), 425-440.

Klaue J.; B. Rathke; and A. Wolisz. 2003. "EvalVid, - A Framework for Video Transmission and Quality Evaluation", in Proceeding of the Modeling Techniques and Tools for Computer Performance Evaluation Conference, (Urbana, Illinois, Sept. 2-5), 255-272.

Kohler E.; M. Handley; S. Floyd; and J. Padhye. 2006. “Datagram congestion control protocol (DCCP),” Request for Comment (RFC) 4340, IETF, (March).

Larzon L; M. Degermark; and S. Pink. 1999. "UDP Lite for Real-Time Multimedia Applications", in Proceeding of the International Conference of Communications (Vancouver, Canada, Jun 6-19), 1999.

Meddour, D-E.; A. Abdallah; T. Ahmed; R. Boutaba. 2011.”A cross layer architecture for multicast and unicast video transmission in mobile broadband networks”, Journal of Network and Computer Applications, Vol. 35, Issue 5 (September 2012), 1377–1391.

Mohanty, S. and I. F. Akyildiz."A Cross-Layer (Layer 2 + 3) Handoff Management Protocol for Next-Generation Wireless Systems", Transactions on Mobile Computing, IEEE, vol.5, no.10, 1347-1360.

NIST (National Institute of Standards and Technology). 2007. “The Network Simulator ns-2 add-on - IEEE 802.16 model (MAC+PHY)”, Draft 1.2.1, (Jun).

Noordin, K. A. and G. Markarian. 2007. "Cross-Layer Optimization Architecture for WiMAX Systems", in the Proceedings of the 18th International Symposium on Personal, Indoor and Mobile Radio Communications, (Athens, Sept.), 1-4.

Raisinghani, V. T.and S. Iyer. 2004. “Cross-Layer Design Optimizations in Wireless Protocol Stacks”, Computer Communication Journal, vol. 27, 720–24.

Stewart, R., "RFC 4960: Stream Control Transmission Protocol", Protocol Specification, Sept. 2007.

Suherman; M. Al-Akaidi, R. Hamzaoui. 2011. “Inter-frame Retransmission for Video Surveillance over WiMAX”, in the proceeding of the 12th Middle Eastern Simulation and Modeling Conference (Amman, Jordan, Nov.), , 42-47.

Wang Q. L.; G. Z. Liu. 2011. "A priority-based EDF scheduling algorithm for H.264 video transmission over WiMAX network," in Proceeding of the International Conference on Multimedia and Expo (Barcelona, July), 1-6.

Wei Z. H. and D. W. Chao. 2010. "Research of Embedded Reliable Transmission Protocol", in Proceeding of Information Engineering Conference (Heibei, China, Aug 14-15), 272-275.

Welzl M. 2005. “Passing Corrupt Data Across Network Layers: An Overview of Recent Developments and Issues,” in EURASIP Journal on Applied Signal Processing, (Jan), 242–247.

Wirz R.; R. Marin; M. Ferre; J. Barrio; J. M. Claver; and J. Ortego. 2009. "Bidirectional Transport Protocol for Teleoperated Robots", in IEEE Transactions on Industrial Electronics, vol. 56, no. 9, (Sept.), 3772-3781.

Ye, J.; J.-X Wang and J.-W.Huang. 2011. “A cross-layer TCP for providing fairness in wireless mesh networks”, International Journal of Communication Systems, vol. 24, issue 12, 1611–1626.

Zhai, H.; X. Chen; and Y. Fang.2007., "Improving Transport Layer Performance in Multihop Ad Hoc Networks by Exploiting MAC Layer Information", IEEE Transactions on Wireless Communications”, vol.6, no.5, 1692-1701.

INFORMATION PROCESSING

MODELLING BEHAVIORAL FLEXIBILIY: A DUAL CONTROLLER FOR A CONTEXT-DEPENDENT REINFORCEMENT LEARNING PARADIGM

Oussama H. Hamid
Faculty of Computer Studies
Arab Open University (AOU) – Kuwait Branch
P.O. Box 830 Al-Ardia, 92400 Kuwait
E-mail: oh.hamid@gmail.com

KEYWORDS
Models and Simulations, Reinforcement Learning, Temporal Sequence Information, Associative Learning.

ABSTRACT

A widely discussed problem among researchers of reinforcement learning is how to generalize and transfer experience from a familiar context, where it was acquired to an unfamiliar context, where it might prove helpful. We propose a dual controller for modeling such a behavioral flexibility in a context-dependent reinforcement learning paradigm. Our controller combines two alternatives of ideal learner algorithms. In the first alternative, rewards are predicted by individual objects presented in a temporal sequence (one per trial). In the second, rewards are predicted on the basis of successive pairs of objects. Simulations run on both deterministic and random temporal sequences show that only in case of deterministic sequences, a previously acquired context could be retrieved. This suggests a role of temporal sequence information in the generalization and transfer of experience.

INTRODUCTION

Reinforcement learning (RL) is an algorithmic theory for learning by experience optimal action control (Bertsekas and Tsitsiklis 1996, Kaelbling et al. 1996, Sutton and Barto 1998). Historically, it was born out of mathematical psychology and operations research. Over time, it grew up into a well-established field of machine learning (Niv and Montague 2008).

The problem of RL can be described as follows: a goal-directed agent, which might be a natural (biological) or an artificial system, is interacting with an environment via sensory inputs and subjective actions. The inputs provide the agent with some information about the state of the environment. The agent responds with an action, changing the current state, before it then receives a numerical signal telling how close it moved towards its goal or further away from it (Maia 2009).

Researchers often use RL for tasks that require a sequence of actions (Dayan and Abbott 2005). Performance of the agent depends on the explicit learning of the temporal structure of the sequence (Daw et al. 2005). Sometimes, however, the setting does not involve sequential information as part of the explicit task. In this case, any information provided by the temporal statistics of the operating environment would be task-irrelevant. Yet the incident learning of task-irrelevant temporal sequence information could be, under certain circumstances, vitally helpful. In particular, when reward contingencies change in a way that makes learning 'context-dependent'. Here, past experience becomes crucially important for enhanced performance. But how can experience be generalized and transferred from a familiar context, where it was acquired, to an unfamiliar context, where it may, nevertheless, prove helpful? This question is one of the difficult issues within cognitive systems and learning theory (Gallistel and King 2009, Koechlin and Hyafil 2007, Niv and Montague 2008, Phelps 2007).

We hypothesize that agents adjust rapidly to the changed reward situation (by activating previously memorized associations) when the change is flagged by consistent temporal sequence information, but not otherwise.

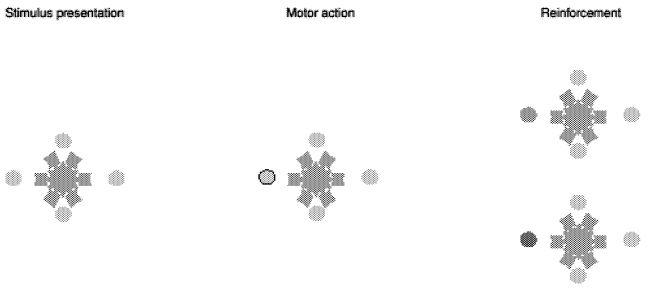
To corroborate this hypothesis we present simulation results of a dual controller that learns (and acts) within a 'context-dependent' RL environment. The simulated paradigm is an extended version of the behavioral paradigm developed by Hamid et al. (2010). It differs from Hamid et al.'s work in that context in the present study is manipulated through multiple reversals. Our dual controller combines two alternatives of ideal learners. The first learner predicts reward by individual objects, whereas reward predictions in the second learner are based on successive pairs of objects within a temporal sequence. Simulations run on both deterministic and random temporal sequences show that the devised controller memorizes a previously learned context in case of deterministic, but not random, sequences. Obviously, it does so because only deterministic sequences have a temporal structure that allows transferring acquired stimulus-response associations forward in time. This suggests that temporal sequence information may well contribute to the generalization and transfer of experience, as it gives appropriate weight to irrelevant information for the case that such information becomes vitally important.

The next section is dedicated for previous related work. Two further sections introduce the applied learning paradigm as well as models and simulations. Simulation results are discussed in a fourth section. The last section summarizes and sets out for future work.

CONDITIONAL ASSOCIATIVE LEARNING

Studying the role of temporal context with humans, Hamid et al. (2010) developed a behavioral paradigm for learning arbitrary visuomotor associations. In this paradigm,

observers viewed sequences of highly distinguishable fractal objects (one per trial) and learned to associate each object with one of four possible motor actions.
Each trial comprised three uninterrupted phases: stimulus presentation, motor action and reinforcement. First, a visual fractal object appears at the center of a computer screen, surrounded by four action options (gray discs). Second, the observer reacts by pressing the key that corresponds to one action option (outlined disk). Third, a color change of the chosen option provides reinforcement (green if correct, red if incorrect matching). Figure 1 shows a time lapse of a single trial.



Figures 1: The behavioral task as devised by Hamid et al. (2010)

Observers were instructed to respond 'correctly' to each fractal object. It was explained that, for each fractal object, one of the four possible responses was 'correct', while the other three responses were 'incorrect'. Moreover, observers were told that they had to become familiar with and learn to recognize each fractal object and that they had to learn the 'correct' response of each object by trial and error. They were further told that there was no pattern or system that would enable them to predict which response a particular fractal object requires. No mention of or reference to the sequence of trials and fractal objects was made. The sequences contained eight objects and were either maximally deterministic or maximally random.
The authors found that human observers could acquire the correct stimulus-response associations in both types of sequences. Yet learning in case of deterministic sequences was *faster* than in case of random sequences. These findings are particularly intriguing, for the temporal structure of the sequences was task-irrelevant. Hence, it could be ignored. However, observers seem to have allocated attentional and memory resources to such "redundant" information. Presumably, in order to better navigate the learning environment. If so, then temporal sequence information should play a fundamental role in context-dependent learning, even if task-irrelevant. Indeed, other studies (e.g. electrophysiological studies) with behaving non-human primates have also shown evidence of incidental learning of task-irrelevant sequence information (Miyashita 1988, Yakovlev et al. 1998).

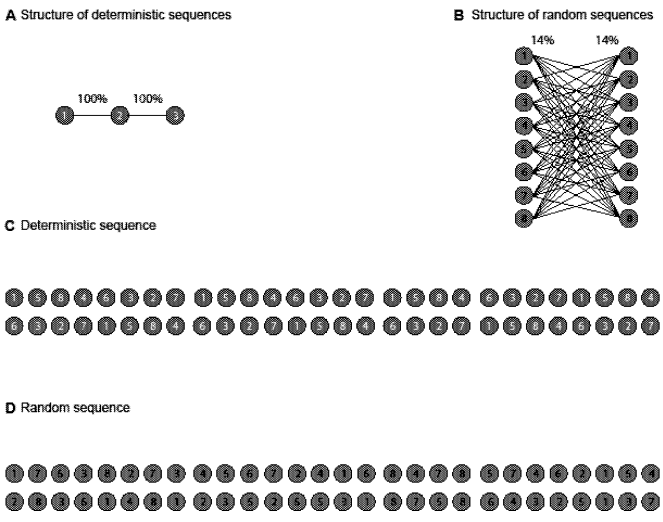
THE LEARNING PARADIGM

The Main Hypothesis of the Current Work

Drawing on Hamid et al. (2010), the present work introduces a novel context-dependent RL paradigm, in which the learning context is switched twice over three phases. In the first phase, a certain context, context 1, defined through a pair of objects sequence and object-action mapping is given. In the second phase, both the sequence and the mapping are replaced by new ones, resulting in a new context: context 2. The third phase is characterized by switching back the currently active learning context (context 2) to context 1. We hypothesize that within the first two phases, learning of the correct object-action associations will proceed gradually, regardless the type of the presented temporal sequence. But when context 1 prevails again, within the third phase, only deterministic sequences will provide a memory mechanism for retrieving the previously learned associations of context 1. As a result, there will be an abrupt jump in performance, rather than a gradual progress.

Primary Sequences

Eight visual objects were used to build primary sequences of 56 trials long. Primary sequences were either maximally deterministic or maximally random (Fig. 2). A sequence is maximally deterministic if each object can be fully predicted by its predecessor object (100% probability, Fig. 2A). It is maximally random if the probability of each object to be predicted by its predecessor is reduced to its minimal value. With eight objects, such a probability is 1/7 (≈14%, Fig. 2B). Deterministic primary sequences were defined by repeating an arbitrary permutation of the eight objects, e.g. 1, 5, 8, 4, 6, 3, 2, 7, seven times (Fig. 2C). Random primary sequences were obtained by making each object precede (follow) every other object exactly once (Fig. 2D). Consequently, objects in both types of sequences had the same number of appearances (seven times each). Yet they differed in the number of appearances of pairs connecting them to their predecessors. While pairs of successive objects in a deterministic primary sequence appeared seven times as the objects themselves, pairs in random primary sequences appeared only once.



Figures 2: Deterministic and random sequences

Composed Sequences and Context Reversal

In every simulation run, three primary sequences of the same type were presented without interruption, summing up to

composed sequences of 168 (56 x 3) trials long. Composed sequences described three learning contexts: context 1 (trials 1 to 56), context 2 (trials 57 to 112) and context 3 (trials 113 to 168).
A context reversal took place if at least one object-action pairing was changed. Such a constraint may seem irrelevant for artificial learning systems; however, it is vitally critical for human learners, as the human's brain is thought to rely on intuitive learning strategies under certain circumstances (Daw et al. 2005).
Importantly, successive contexts were pairwise different, except contexts 1 and 3, which were identical. This was aimed to test whether the designed controller uses contextual temporal information in the retrieval of the previously acquired experience (context 1) after context reversals. Should this happen, it would corroborate the main hypothesis described above.

MODELS AND SIMULATIONS

The Basic Model

A simple model for the given task is that response probabilities are modified directly so as to maximize expected reward.
For each object n, four response probabilities $p_j^{(n)}$ must be learned. Here, $j \in \{1,\ldots,4\}$ and $\sum_j p_j^{(n)} = 1$. When object n is observed, action k is selected, and reward $r_k \in \{0, 1\}$ is received, a suitable rule for updating response probabilities is

$$p_j^{(n)} \leftarrow \begin{cases} p_j^{(n)} + \lambda \left(\delta_{jk} - p_j^{(n)}\right) & : r_k = 1 \\ p_j^{(n)} - \mu \left(\delta_{jk} - p_k^{(n)}\right) \frac{p_j^{(n)}}{\sum_{j \neq k} p_j^{(n)}} & : r_k = 0 \end{cases} \quad (1)$$

where λ and μ are learning rates in the range of [0, 1] and δ_{jk} is the Kronecker delta (which equals 1 if j=k and 0 if $j \neq k$). This rule ensures $0 \leq p_j^{(n)} \leq 1$ and $\sum_j p_j^{(n)} = 1$.

The Ideal Learner

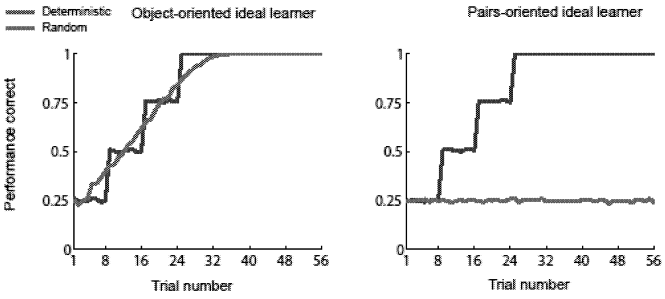
The model in Equation (1) defines a family of learning strategies. For example, choosing $\lambda > \mu$ makes learning faster in rewarded than in unrewarded trials. Choosing the maximal rates $\lambda = \mu = 1$ implements an 'ideal learner'.
An ideal learner is someone who has full knowledge of the structure of reward contingencies and who narrows the remaining possibilities down as quickly as possible. In this sense, even an ideal learner makes mistakes when learning by trial and error. But it never makes the same mistake twice.
Because each visual object deterministically predicts the rewarded action, an ideal learner identifies the correct action for an object in *no more* than three appearances of the object. This applies for each object regardless the temporal order, in which an object occurs, within a sequence of objects.

The Dual Controller Algorithm

Rather than focusing *solely* on the explicit task, our dual controller learns also task-irrelevant temporal sequence information. Specifically, it implements two ideal learner algorithms that run simultaneously. One learns associations between individual objects and corresponding rewarded actions in the current trials. The other learns associations between successive pairs of objects from the current and previous trials (trials t and t-1) along with the currently rewarded action. We call the first ideal learner 'object-oriented ideal learner', whereas the second is termed 'pairs-oriented ideal learner'.
On every trial, the controller updates and carries forward in memory reward probability values, delivered by the two ideal learners. Reward probabilities belong to the four possible actions. Each of the two learners suggests an action on the basis of accumulating reward-related information over trials. A decision, about which action should be eventually taken, is made by comparing the current probability values of the suggested actions. In case that the learners suggest different actions, the controller selects the one with the higher value. In section RESULTS AND DISCUSSION, we will discuss the correctness of this algorithm.

RESULTS AND DISCUSSION

Figure 3 shows simulation results run on both deterministic and random sequences of visual objects. A total of 3416 simulation runs was performed on both object-oriented and pairs-oriented ideal learners. Simulations were done in Matlab.
The object-oriented learner acquired correct associations in either type of sequences (Fig. 3, left subplot). The pairs-oriented ideal learner, however, learned correct associations only when run on deterministic sequences. Applying it to random sequences caused performance to stagnate around the chance level (Fig. 3, right subplot). These results qualified our dual controller to correctly detect previously acquired context and perform flexibly in the present paradigm (Fig. 4). But how can we account for these simulation results?

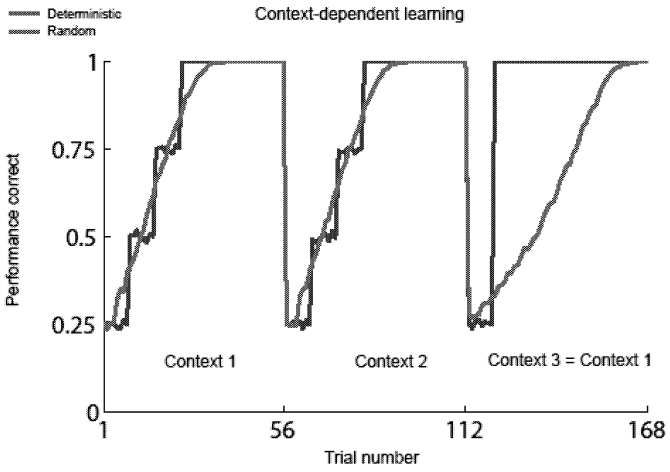


Figures 3: Simulated performance of the two ideal learners

To explain, consider two deterministic sequences of visual objects: sequence 1 and sequence 2. Let the objects x, y, and z be such that object z is always preceded by object x in sequence 1 and by object y in sequence 2. Moreover, let a_1 be the rewarded action for object z within sequence 1, whereas a_2 the rewarded action for object z within sequence 2. Together with the described object-action associations, sequences 1 and 2 define two distinct contexts: context 1 and context 2, respectively. What the object-oriented ideal learner does in this case is that it rewards associations

between object z and action a_1 in context 1 and it rewards associations between object z and action a_2 in context 2. Similarly, the pairs-oriented ideal learner rewards associations between the pair (x, z) and action a_1 in context 1, and it rewards associations between the pair (y, z) and action a_2 in context 2.

At the beginning of learning (phase 1), both ideal learners set reward probabilities for the four actions, corresponding to object z, to the chance level (25%). However, from the fourth appearance of object z on, reward probabilities for action a_1 reach their ceiling value (100%) by both learners. All other actions which are not equal to a_1 end up with a probability of null. This holds, in particular, for the actions, corresponding to the pair (x, z). The pair (y, z) experiences no updates by the pairs-oriented learner within this context.



Figures 4: Simulated performance of the dual controller

When the context changes (phase 2) and object y becomes the new predecessor of object z, the updating procedure of the pairs-oriented ideal learner activates the corresponding actions of the pair (y, z). The learner starts to accumulate reward probabilities for the newly active pair, reinforcing, this time, action a_2. Importantly, the learner preserves reward probabilities for the previously active context in the pair (x, z). By contrast, the object-oriented ideal learner preserves no information across contexts. Instead, reward probabilities are overwritten every time a context reversal takes place. This is given by the basic model (Equation. (1), case $r_k = 0$).

When context 1 prevails again (phase 3), the object-oriented learner selects action a_2 for object z in its first appearance in phase 3. The pairs-oriented learner re-activates the pair (x, z), as object z is now preceded by object x. This leads to suggesting action a_1 by the pairs-oriented learner. The controller selects action a_1, because the object-oriented learner has priority over the pairs-oriented learner when different suggested actions have the same probability values (at this stage, both values are 100%). Here, the controller receives no reward, and performance falls abruptly to the chance level (as expected). As a result, the object-oriented learner updates reward probabilities according to Equation (1), case $r_k = 0$ with values < 1 for all actions. Nothing changes on the side of the pairs-oriented learner. But the next time object z appears within this context, reward probability of action a_1 preserved in the pair (x, z) will still be 100%, which is greater than the value of any other action suggested by the object-oriented learner. Consequently, the controller selects action a_1, which is the correct action. This causes an abrupt jump of performance to the ceiling level.

The lack of consistent predecessor-successor relationship between objects in random sequences prevents the pairs-oriented learner from preserving previously acquired associations in memory. As a result, it learns object-action associations in the phase 3 (context 1) as if they belonged to a new context.

SUMMARY AND FUTURE WORK

We have presented a dual controller for a context-dependent reinforcement learning paradigm. Our paradigm is an extended version of the behavioral paradigm that was developed by Hamid et al. (2010). Yet it differs from previous work in manipulating the learning context. The controller combines two alternatives of ideal learner algorithms that allow to implicitly learn (and deploy) temporal sequence information. As a result, the controller predicts that experience during the learning process may be transferred across contexts by means of temporal sequence information.

Next, we shall investigate these predictions by running behavioral experiments with human subjects.

AUTHOR BIOGRAPHY

OUSSAMA H. HAMID received in 2011 a Ph.D. degree in Natural Sciences (Neuroscience) from the University of Magdeburg, Germany. The title of his doctoral thesis was: “On the Role of Temporal Context in Human Reinforcement Learning”. Before, he was awarded a Master's degree in Computational Visualistics (Germany, 2002) and Bachelor degrees in Computer Science (Germany, 2000) and Mathematics (Germany, 1998). In the years between 2003 and 2012, Dr. Hamid has been investigating both theoretical and practical aspects of Human-Machine-Interaction, Neural Network Functions and Cognitive Systems. Dr. Hamid has published his research findings in several European and international refereed journals and conferences.

REFERENCES

Bertsekas D.P. and Tsitsiklis J.N. 1996. *Neuro-Dynamic Programming.* Athena Scientific.

Daw N.; Niv Y.; and P. Dayan. 2005. “Uncertainty-Based Competition Between Prefrontal and Dorsolateral Striatal Systems for Behavioral Control”. *Nature Neuroscience* 8, 1704-1711.

Dayan P. and L.F. Abbott. 2005. *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*. The MIT Press.

Gallistel C.R. and A.P King. 2009. *Memory and the Computational Brain*. Wiley-Blackwell, West Sussex, United Kingdom, first ed.

Hamid O.H.; Wendemuth A.; and J. Braun. 2010. “Temporal Context and Conditional Associative Learning”. *BMC Neuroscience*, 11:45, No. 1, 1-15.

Kaelbling L.P.; Littman M.L.; and A.W. Moore. 1996. “Reinforcement Learning: A Survey”. *Journal of Artificial Intelligence Research*, 4, 237-285.

Koechlin E.; and A. Hyafil. 2007. “Anterior Prefrontal Function and the Limits of Human Decision-Makinig”. *Science*, 318, 594-598.

Maia T.V. 2009. “Reinforcement Learning, Conditioning, and the Brain: Success and Challenges”. *Cognitive Affective & Behavioral Neuroscience*, 9, 343-364.

Miyashita Y, 1988. “Neuronal Correlate of Visual Associative Long-Term Memory in the Primate Temporal Cortex”. *Nature*, 335, 817-820.

Niv Y. and P.R. Montague. 2008. “Theoretical and Empirical Studies of Learning”. In *Neuroeconomics: Decision Making and The Brain 2008*, P.W. Glimcher; C. Camerer; E. Fehr; and R. Poldrack (Eds.). NY: Academic Press, New York. 329-349.

Phelps E.A. 2007. “Learning: Challenges in the Merging of Levels”. In *Science of Memory: Concepts 2007*, H.L.R. III; Y. Dudai; and S.M. Fitzpatrick (Eds.). Oxford, University Press Inc., New York. 45-48.

Sutton R.S. and A.G. Barto. 1998. *Reinforcement Learning: An Introduction.*. The MIT Press, Cambridge, Massachusetts.

Yakovlev V.; Fusi S.; Berman E.; and E. Zohary. 1998. “Inter-Trial Neuronal Activity in Inferior Temporal Cortex: A Putative Vehichle to Generate Long-Term Visual Associations”. *Nature Neuroscience* 1, 310-317.

A Proposal of an e-Monitoring system for the Online e-TMA (Tutor Marking Assignment) for Quality Assurance Purposes: Arab Open University's experience

Haifaa O. Elayyan
ITC department lecturer
Arab Open University, Saudi Arabia
helayyan@arabou.edu.sa, haifaa_ashraf@yahoo.com

Keywords: Open learning, LMS, quality assurance, online quiz theme, fully online eTMA, e-Monitoring systems

Abstract
The Arab Open university adopts an open learning methodology that focuses on the main characteristics of the pedagogical model which is called the learning management system (LMS). Every educational progress has any kind of course work assessments in an attempt to evaluate and enhance student's performance and required quality assurance processes in general and more crucial for open learning institutions. Tutor marking Assignments and Quizzes are two of the old assessments known and being used in AOU. A previous work and experience in AOU presented the Tutor Marking Assignment (TMA) in an online electronic quiz theme to be e-TMA [2]. The previous work focused how to present e-TMA as a fully online electronic TMA in a friendly and well organized framework that acquire students to adopt fast, learn and apply it before the cutoff date. This paper discusses the possibilities of emerging the e-TMA with an online e-Monitoring system proposed in [1] for quality assurance purposes to monitor the tutors' performance while grading this eTMA given that AOU is committed in a partnership with Open University , England.

1 Introduction: Current Tutor Marking Assignment (TMA) using the Assignment activity module:

The Arab Open University (AOU) uses two forms of student performance assessment in its pedagogical model. These forms are *summative assessment* and *formative assessment*. The former uses electronic and written tests for quizzes and final exams; while the latter form of assessment uses Tutor-Marker Assignments (TMA) as a continuous assessment cycle of forming a solution to real-world problems, assessing the solution, and then improving the solution for further rounds of improvement and re-assessment. TMA is presented by the virtual learning environment or (LMS) the AOU uses. TMA is created by using an assignment activity module in LMS. The following steps should clarify how assignments are created in LMS.
TMAs are usually files (MS word, PDF, Excel, etc) with different kinds of questions but mainly calculations and essay ones. The staff tutor should declare the material this assignment covers and specify a cutoff date with other instructions such as plagiarism warning inside the assignment file itself.

These assignment files are created by uploading these files to the course page in LMS in very obvious spot on LMS so students can notice it. For example, a flashing title such as the following:

Figure 1: Animated notification for the available TMA

Students can download the file by clicking on the flash animated text which is linked to the TMA file.

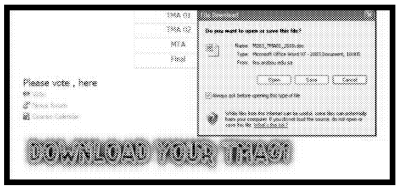


Figure 2: Students download the TMA by clicking on the animated notification

Staff Tutor creates by using assignment activity module to data structure contains a field (a link) called submitted TMA for students to upload one single file or advanced single files. This activity module is well supported with many setting help the staff tutor to control the assignment and monitor it automatically. The following screen shot has the main settings for the assignment activity module : Grade for the maximum grade of this TMA, Available and due dates , prevent late submissions , size of submitted file, notification sent to tutors every time a student submits his/ her TMA. Separate modes to separately upload files, is not common as it is an assignment and should be submitted separately.

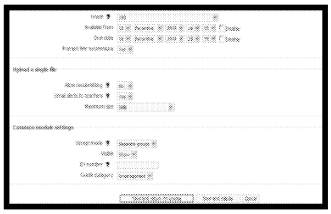


Figure 3: Assignment activity module Settings

Once the settings are saved, the assignment data structure can be described as:

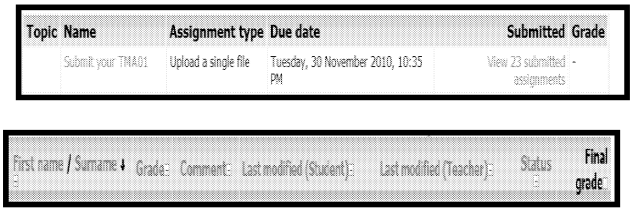
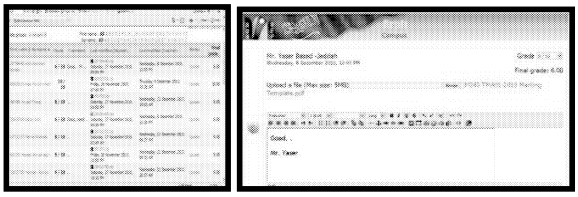


Figure 4: sample from assignment activity module when it is built .

Students and tutors are then notified by sending them private messages that the Assignment file is available on LMS.

Students are supposed to download the assignment file from LMS by clicking on assignment file as the follow screen shots show, and start working on these files using facilities Microsoft Office offers
Before the cutoff date, students are requested to submit their TMAs to the LMS in a requested document format. These TMAs are manually downloaded by the tutor and then by using the pre–received answer key file from the staff tutor, TMAs are graded and the informative feedback is inserted into the assignment marking template for each student.
Grades are submitted along with informative feedback files to the LMS using the above described assignment activity module students used to upload their assignment files.
Students log into the LMS to check their grades and download their feedback files.



(a) Assignment structure . (b) Tutor has submitted grade 6/10
Figure 5. (a) shows the submitted assignments
(b) shows the uploaded grades and feedback

The large number of TMA submissions and the involved TMA handling process adds tremendous amount of work on the tutors and the staff tutors as the cycle described has many offline mechanical steps.

2 Monitoring System proposed by [1]:

The need for an electronic system for monitoring the tutor-marked assignments which was presented above stems from the significance of this assessment part to the success of open learning. Adding to it the tremendous amount of "mechanical" work associated with the monitoring and feedback and quality assurance procedures for the TMA submissions, the urgency of such a system becomes quite apparent and this is what the reference "Essential of e-TMA system [1]" had presented.

As we have mentioned earlier, the AOU decided to use Moodle as one of the popular learning management systems in its online educational delivery. Moodle supports a module called Assignment which has most of the needed functionality for the targeted the monitoring system. The current Assignment module as described above supports submission, deadline enforcement, grading and feedback templates as well as automatic notification tool to notify the tutor when the submission of a TMA is taking place , and notify the student when the grade and feedback file are uploaded to the LMS

2. 1 How the monitoring system works [1] with the Assignment Activity Module:

There are 4 different types of user accounts in Moodle (student, tutor, course creator, and administrator). The tutor account (teacher account in Moodle terminology) allows the tutor to access group activities such as TMA grading. Some tutor accounts are allowed to edit the course activities and hence we call these accounts staff tutor accounts. These accounts enable the staff tutors to edit all groups' activities and most importantly the TMAs in all groups. In order to facilitate the monitoring capabilities in the Assignment module of Moodle. [1] to add one more attribute to the staff tutor account: the monitoring level was proposed. This attribute will be used later on as a trigger for automatic monitoring system of the submitted TMAs in each group for each tutor separately Figure 6 and Figure 7 below show how this attributed is defined.

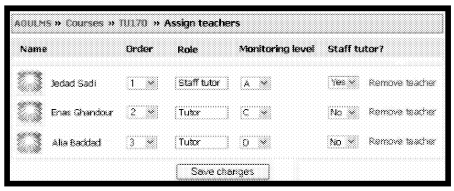


Figure 6: Defining the monitoring level for staff tutors.

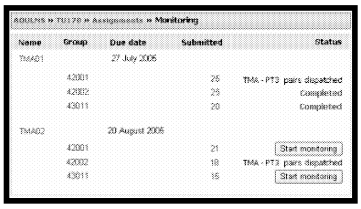


Figure 7: Triggering the monitoring procedure in the modified Assignment module

Figure 6 and 7 above, overview the modifications needed to be done to the Assignment module of Moodle to enhance it with monitoring and archiving capabilities. Figure 8 made a closer look on how the module will function to complete the picture of our targeted TMA monitoring system.

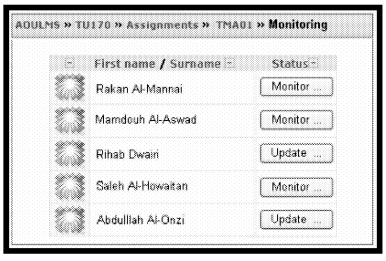


Figure 8: List of TMA scripts ready for monitoring in one group

Upon clicking on the button Monitor a simple form appears (say regarding form) on which the monitor will fill the basic monitoring information. This form should contain the following essential fields:

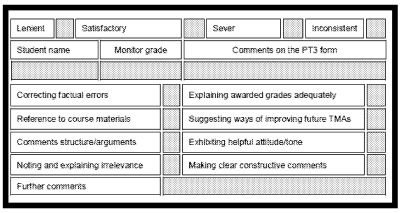


Figure 9: The view form

Figure 10 and Figure 11, below show the view form after applying the Monitoring:

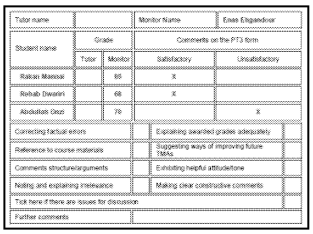


Figure 10: The monitoring form as displayed for the monitor after monitoring the TMAs.

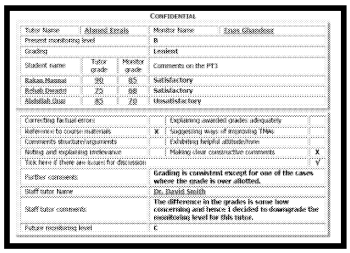


Figure 11: Completed monitoring form.

2.2 Results of applying the monitoring system on TMA activity module:

In the monitoring system proposed by [1] and explained briefly before, we outlined the three main tasks related to handling TMAs (assignments) in the AOU educational model. We then explained how the Assignment module in Moodle can be extended to include monitoring and archiving functionalities. In collaboration with other five partners, the AOU embarked on this extension in a framework of joint project (*EC funded project jointly implemented with AOU, JU, and PSUT - Jordan, AU – Spain, IDEC - Greece, ORT – France, AIESEC – The Netherlands, and PLOCDIV – Bulgaria)*

The system proposed in [1] has many advantages of automating the Monitoring system mainly quality assurance purposes. Unfortnely the system is applied without enhancing the procedure of the TMA itself. Therefore, the TMA as described in section 1 , has lots of mechanical steps including uploading , downloading and grading is still out of LMS concept even after adding the new attributes about automated monitoring system . In the rest of the paper is organized as follows: section 3 will outline the fully online e-TMA experience we had in AOU and compare the results with the original TMA steps. Section 4 gives a proposal of adopting the monitoring system detailed in [1] and briefed in the previous section 2. Conclusions are given in the last section

3 The Tutor – Marked Assignment (TMA) in an online quiz theme (Fully Online e-TMA):

Derived by the success of open learning models; [2] presented a new experience of creating and handling the TMA inside the LMS by using the quiz activity module instead of the regular TMA activity module since the last has lots of manual steps and grading mechanism is done out of LMS environment. This electronic e-TMA system as applied to the Arab Open University is meant to regulate grading, monitoring, archiving, and maintain a complete record for quality assurance purposes. We will brief the previous work in e-TMA as follows.

3.1 Customization of Quiz activity module to generate the eTMA:

a. eTMA Question Bank:

Aimed eTMA is an integrated collection of questions in one eTMA quiz container. Customization of the question bank is simply done by adding a new category called with a proper descriptive name for the eTMA to hold the questions of this Assignment. Many questions types can be used in the same way as a quiz. In our experience presented in [2], we used essay and short-essay question types for their flexibility and common use that helped a lot to break down the TMA into small problems (questions) that can be handled and answered separately. Grades are assigned to each problem which makes it more flexible for the tutor to estimate and evaluate the performance per question. Short answers are graded by the computer which helps to get the results quickly. Figure 12, shows breaking the TMA into problems by using the essay and short essay questions. Note that all the e-TMA screenshots were derived from [2] as the experience was detailed.

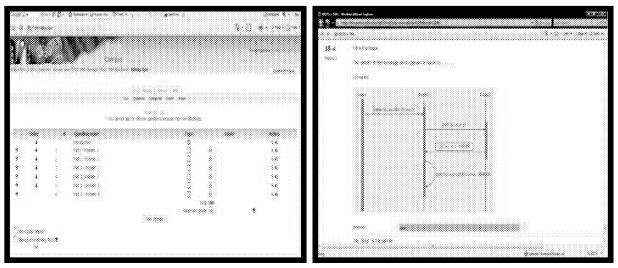


Figure 12:breaking TMA into small problems [2]

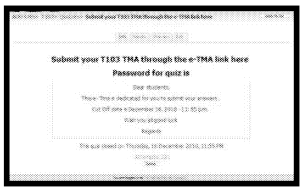
b. eTMA Body

1 Create the eTMA Body and name it. The eTMA body is created by adding a new quiz activity. This body should be named with a meaningful descriptive name such as T**103 eTMA:** T103 is the associative course name while eTMA mean the tutor marking assignment for this course .

2 An introduction to the eTMA: the body text which contains any special instructions for this eTMA as cutoff date, plagiarism

and cheating instructions and consequences. Also, valuable information about how to use the eTMA, number of attempts every student can take and / or scoring rules.

(a) eTMA in quiz theme with introduction



(b) How student will attempt to eTMA.

Figure 13: (a) & (b) Creating an TMA using a quiz activity module. [2]

3 Settings:

1. Select the timing options: Start date this eTMA is available for students and date this eTMA is not available any more (cutoff Date)
2. Time limit: Disabled as the timing options are enough to specify available and cut off dates.
3. Time delay between attempts: Disabled as this setting is not important here.
4. Select the display options: This sets the number of questions the students will see at once. Usually eTMA consists of many questions. This option will organize displaying the questions by limiting the number of questions per page, the students will see a navigation button at the bottom of the page where they can view the questions on other pages.
5. Shuffle questions: This option can be disabled as eTMA should have a unified displaying and arrangements in terms of questions to all students.
6. Shuffle within questions: This option also can be disabled.
7. Attempts options: One attempt is requested to the eTMA.
8. Each attempt builds on the last: This option should be set to yes as we need a continuous work that depends and records the answer from the last login .
9. Select the security options: No need as the eTMA is an assignment that students need to use other facilities such as Microsoft office or browsing the web for more information.
10. Require password: This option can be disabled as the eTMA should be open for students to attempt.
11. require network address : This option is left to the staff tutor , but recommended to be disabled.
12. Select the common module options: separate group mode is recommended as it is an assessment and need grades to be private for students.

4. eTMA Grading options:

Highest grade is selected with other setting id desirable such as a plenty for late submissions.

5 eTMA Feedback:

Feedback has two types: an answer key saved in the feedback text for each question while being created and saved. The other feedback is the comments and informative encouraging feedback from tutor to the student in response for their answers. This feedback is being inserted in the comments text body while manual grading each question. It will be shown after tutor finishes grading. The following are screen shots of how to view the eTMA:

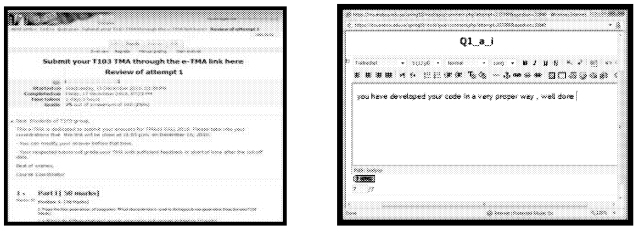


Figure 14: eTMA feedback for each question separately [2]

6 Saving and Archiving:

While students are working on their eTMA , they are requested to click on save without submitting button. This will give him/her another chance to work on the eTMA again for further modification.

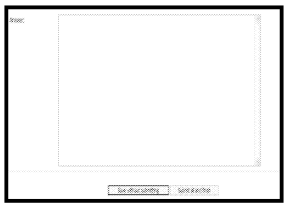


Figure 16: shows the step student needs to do to save his/her work and get another chance to work again.[2]

Clicking on save button will Archive student's work. Archiving is one important point to mention by using the eTMA and quiz activity module in general. The system will record students work every time they log to the eTMA and work on it. Their work is saved and archived with details as the following screen shows that archiving is not only for student work, but it is also for tutor work as well.

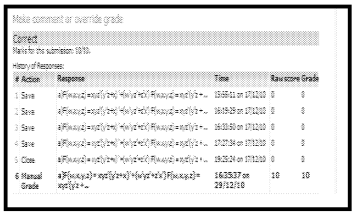


Figure 17: Screen shots from the eTMA [[2]

Saving Student work will avoid being unable to submit their work with the assignment activity module when it closes. Using the quiz theme will save the work even it is partially, better than granting a zero for no submission.

7 Following up and Evaluating students performance:

Using the quiz theme, evaluating student is an easy step as the quiz activity module is supported with analysis and item analysis figure 18 shows :

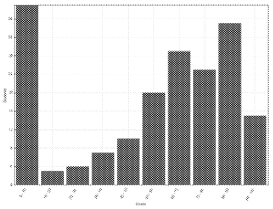


Figure 18. Bar Graph of Number of Students Achieving

4. Advantages and Challenges of the fully online e-TMA:

The experience of eTMA has many practical advantages and challenges as well that we want to list them briefly as they were detailed in [2]:

4.1 Main Advantages

1- Breaking down the TMA into small problem that can be handled separately in a fully online container which came to overcome the manual steps with TMA and grading out of LMS. As the eTMA made it possible to break down one problem into small problems and present it in many different ways using the variety of questions, quiz activity module supports. This is really an important point when it comes to students.
2- Evaluating the performance for tutors and students are more practical as the quiz theme has many additions more than the simple TMA activity module.
3- Being online, increased the students' activities on LMS as they had to be logged online every time they want to work on their eTMA.
4- Upgrading the level of students' computer skills as they had to learn and adopt in a limited time.
5- In general, eTMA saves the cost of managing staff as Students work at their own pace.
6- A wide variety of eTMA reports could be viewed not only focusing on attempt but also the overall performance of students in specific questions can be easily done

4.2 Challenges:

The major challenge is the internet connectivity and the broadband. e-TMA being online , students may face problems when they don't have constant internet connectivity.

5. A proposed system: Emerging the online e- TMA [2] along with the Monitoring system [1] for quality assurance:

The Arab Open University adapts any positive enhancements and any new e-learning technologies that can be applied to the current educational model, the LMS, it uses. A learning revolution is occurring with a technology-driven. The new emerging between online e- TMA [2] and the Monitoring system [1] is democratic, as students and teachers share responsibility for both teaching and learning. The students construct their learning by actively seeking their own information. In fact the philosophy "learn how to learn," is one concept of open learning where students must take charge of their education through direct, expression and experience. Therefore, teachers become more the "guides on the sides" helping the students to access, organize and transfer information to find answers to real life problems.

This emerge will come up with extremely desired results of automated monitoring of the online and archived electronic e-TMA and generate automatic monitoring forms to evaluate the teaching process concentrating on the most crucial course assessments, TMAs. And worth to mention that , as long as quiz (container of e-TMAs) is graded question by question , such monitoring form will help a lot to keep a trace and notify if there is one question not yet graded while others in e-TMA are graded. In the following section, we will propose the possibility of emerging the two systems to come with e-monitoring of the fully online e-TMA system.

6. Adopting Monitoring system in the quiz activity Module

Taken into consideration that the online e-TMA is contained into an online quiz body as described in section 3, the modification will be adopted on the quiz module not on the TMA activity module as [1] proposed. Once the online attempts (students online e-TMAs) are submitted and saved and the e-TMA system notified to the tutor to start grading, the archive system attached to the quiz module should trigger the monitoring form and create automatically the following folder as figure 19 shows:

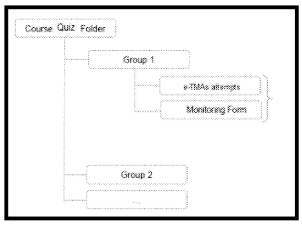


Figure 19: Archived course e-TMA (quiz) folder

In order to facilitate monitoring capabilities in the Quiz activity module (e-TMA container) of Moodle, we propose to add one more attribute to the Staff tutor account: the *monitoring level* as shown in figure 20.

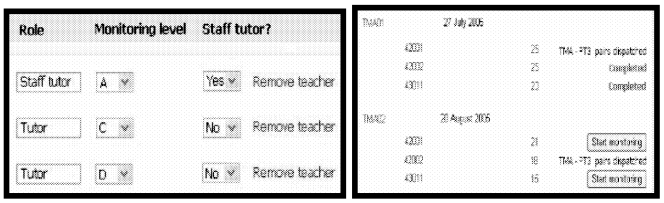
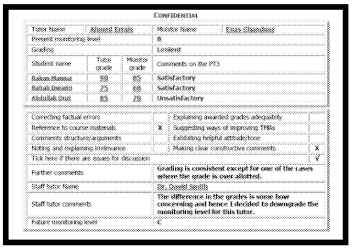


Figure 20: Monitoring and triggering Level added to the Quiz Activity Module.

In the proposed extension for Moodle's Quiz module, the monitoring starts automatically in a certain group once the tutor of that group finishes grading all the submitted e-TMAs. Once all e-TMAs attempts are graded, the monitoring (the groups marked with the link *Completed* which should necessarily point to the completed monitoring form). The following Figure 21,

shows how we start monitoring system. The monitoring form when viewed by the staff tutor will display both grades given by the tutor and the monitor, and of course the tutor name as well as the fields supplied by the monitor in the regarding form. The monitoring form will be final once the staff tutor manually fills the remaining fields. Through a staff tutor account we should be able to view the filled monitoring forms in their various states and for any group. The monitoring form can be obtained by clicking on the designator *Completed form shown in Figure 22 taken from [1]*


CONFIDENTIAL

Tutor Name	Abeer Emails	Monitor Name	Enas Ghandour
Present monitoring level		B	
Grading		Excellent	

Student name	Tutor grade	Monitor grade	Comments on the PT3
Rahaa Mamoul	88	85	Satisfactory
Rehab Dwaidi	75	68	Satisfactory
Abdullah Orud	85	70	Unsatisfactory

Correcting factual errors		Explaining awarded grades adequately	
Reference to course materials	X	Suggesting ways of improving TMAs	
Comments structure/arguments		Exhibiting helpful attitude/tone	
Noting and explaining irrelevance		Making clear constructive comments	X
Tick here if there are issues for discussion			Y

Further comments	Grading is consistent except for one of the cases where the grade is over allotted.
Staff tutor Name	Dr. David Smith
Staff tutor comments	The difference in the grades is some how concerning and hence I decided to downgrade the monitoring level for this tutor.
Future monitoring level	C

Figure 22: Completed monitoring Form integrated with e-monitoring form inside the Quiz Activity Module (e-TMAs container).

7.Conclusion:

The capability of students to acquiring to a new knowledge or technology can be monitored and evaluated with such fully online e-TMA application [2]. Also, teacher performance is monitored to assure e-TMAs as continuous assessment cycle of forming a solution to real-world problems and "learn how to learn" for further enhancements through sufficient feedback that come on time the student gets benefit from . Integrating the online e-Monitoring System and come out with confidential monitoring form such as we have in Figure 22, which is a convenient entry point for the staff tutor and all other quality assurance personnel to perform their tasks easily. The form acts as an archiving tools that simulates the archived course e-TMAs . This form will alleviate the burden on the staff tutors as well as the tutors and the monitors. No need to generate the various CDs for the numerous courses at the end of each semester and academic year.

8.Author Biography

Haifaa Elayyan, is an ITC lecturer and academic coordinator works for Arab Open University, Saudi Arabia Branch . She is a committed researcher with solid background in web systems and has obsessions and high ambitions in new e- technologies and e-solutions that can be an addition to the e-Learning systems.

9 References

Essentials of the e-TMA System for the Arab Open University, *Abdel-Elah Al-Ayyoub , ayyoub@acm.org.*

Tutor Marking Assignment in an electronic quiz theme Arab open university's experience , Haifaa Elayyan , helayyan@arabou.edu.sa , Samar Shilbayeh ICITeS_2011_1569399993

An AI-Inspired Intelligent Agent/Student Architecture to Combine Language Resources Research and Teaching Bayan Abu Shawar, Eric Atwell Information Technology and Computing Dept, Arab Open University, Amman, Jordan and School of Computing, University of Leeds

Quality Assurance Procedures: New Enhancements to the Learning Management System at AOU, Jehad Al-Sadi, Bayan Abu-Shawar, Taleb Sarie

The Arab Open University Professor Taleb Sarie, previous Director, Arab Open University, Jordan

Concept Based Information Retrieval Using Semantic Analysis

Sherimon P.C & Dr.Youssef Saad
Arab Open University
Muscat, Sultanate of Oman
sherimon@aou.edu.om

Dr.Reshmy Krishnan
Muscat College
Muscat, Sultanate of Oman
reshmy_krishnan@yahoo.co.in

Vinu Sherimon
HCT
Muscat, Sultanate of Oman
vinusherimon@yahoo.com

KEYWORDS
Ontology, WordNet, Synonym, Hyponym, Hypernym, semantic analysis, Keyword based retrieval, concept based retrieval

ABSTRACT
To retrieve documents from Internet, keyword based retrieval method was used before. This causes. the problem of polysemy. Concept based retrieval gives the benefit of getting semantic relationships among concepts and eliminating irrelevant documents. Semantic analysis can reveal the conceptual relationships among words in a given document. In this paper the potential of concept-based information access via semantic analysis is explored with the help of a lexical databases called WordNet. The mechanism is applied in the selected text documents and extracting the Synonym, Hyponym, Hypernym of each word from WordNet. The ranking will be calculated after checking the frequency rate of each word in the input document and a hierarchy model will be generated according to the ranking.

1. INTRODUCTION

Although volume of Information available in www has been increasing continuously, most of the information is still unavailable to normal people due to the lack of proper techniques for Information retrieval.85% of the internet users are using Internet for Information retrieval. The Unstructured nature and huge volume of information in www has made it difficult for getting proper result while searching [1].The main issue related to the Information retrieval is poor quality of retrieved results.

The techniques used for Information retrieval was keyword based. This technique uses the keyword list for searching the contents of information. The main concern regarding this approach is the poor quality of the result .One of the reason for this concern is the vocabulary problem facing by the non-expert users. The keywords chosen by the users were often different from those used by the authors of the relevant documents. These problems are referred as synonymy and polysemy.

The information needs of people are in concept space. Keyword based access to information is sometimes unsatisfactory since it works in word space. Words represent concepts in human language but the mapping from words to concepts is many-to-many. That means one concept may be represented with many different words (synonym) and one word may represent many different concepts (polysemy). This mapping problem is known as Word Sense Disambiguation. Secondly, since concepts are abstract entities, representing them is another problem.

Concept-based information retrieval is an alternative IR approach that aims to tackle these problems differently. Concept-based IR represents both documents and queries using semantic concepts, instead of keywords, and performs retrieval in that concept space. This approach holds the promise that representing documents and queries using high-level concepts will result in a retrieval model that is less dependent on the specific terms used [2]. Such a model could yield matches even when the same notion is described by different terms in the query and target documents, thus alleviating the synonymy problem and increasing recall. Similarly, if the correct concepts are chosen for ambiguous words appearing in the query and in the documents, non-relevant documents that were retrieved could be eliminated from the results.

To tackle polysemy, the main proposed method was to apply automatic wordsense disambiguation algorithms to documents and query. Disambiguation methods use resources such as the Wordnet thesaurus [3] to find the possible senses of a word and map word occurrences to the correct sense. WordNet is a large lexical database of English. Nouns, verbs, adjectives and adverbs are grouped into sets of cognitive synonyms (synsets), each expressing a distinct concept. Synsets are interlinked by means of conceptual-semantic and lexical relations. WordNet superficially resembles a thesaurus, in that it groups words together based on their meanings. However, there are some important distinctions. First, WordNet interlinks not just word forms—strings of letters—but specific senses of words. As a result, words that are found in close proximity to one another in the network are semantically disambiguated. Second, WordNet labels the semantic relations among words, whereas the grouping of words in a thesaurus does not follow any explicit pattern other than meaning similarity.

Section 2 reviews the state of the art in concepts based extraction from documents and section 3 sketches out our methodology for the generation of ontology from a text document by extracting semantic web concepts with the help of Word.Net in terms of Design, Implementation and results. Section 4 shows the conclusion and future works.

2. INFORMATION RETRIEVAL

The term Information retrieval (IR) refers to the access to Information and its representation. The key role of information retrieval process is to retrieve relevant information to a given request. The efficiency of the process is the retrieval of all relevant information available and rejection of all non relevant ones. Even though in reality, the results will contain both relevant and non relevant information, the aim is to achieve the ideal criteria. One information retrieval system could handle different information simultaneously. Majority of the information retrieval based on the text documents and hence can be named as text retrieval or data retrieval. The text retrieval incorporates all types of texts including complete articles ,books, web pages and minor fragments of texts such as sections, paragraphs ,sentences etc., Instead of retrieving information directly, the documents will be retrieved in IR process, from which information can be obtained. The basic model of Information retrieval can be shown as follows.

Fig 1.Information Retrieval

Queries are the requests to the system to get the results. Any one of the searching strategy can be used for searching from the internet. According to the queries of the user, documents will be retrieved from the storage using any appropriate search techniques. Storage comprises an abstract description of the input document .This description will be unstructured except for the syntax .The similarity between the given query and the stored documents will be checked in the matching process.

2.1 Concept-based Information Retrieval Model

In the cognitive view of the world, there exists the presumption that the meaning of a text (word) depends on conceptual relationships to objects in the world rather than to linguistic or contextual relations found in texts or dictionaries. A new generation information retrieval model is drawn from this view. We call it concept-based information retrieval model. Sets of words, names, noun-phrases, terms, etc. will be mapped to the concepts they encode.

Generally, a content of an information object is described by a set of concepts in this model. Concepts can be extracted from the text by categorisation. Crucial in this model is existence of a conceptual structure for mapping descriptions of information objects to concepts used in a query. If keywords or noun-phrases are used, then they should be mapped to concepts in a conceptual structure. Conceptual structures can be general or domain specific, they can be created manually or automatically, they can differ in the forms of representation and ways of constructing relationships between the concepts. Naturally, the tools considered in this paper differ in this respect.

For establishing definitions of concepts it is necessary first to identify concepts inside the text and then classify found concepts according to the given conceptual structure. There are several ways of identification of concepts present in the text. This process is called categorization. Concepts can be identified also by using fuzzy reasoning about the cues (terms) found in the text for calculating likelihood of a concept present in the text.

After the concept is categorized, it can be given the definition by a classification process. Classification is determining where in the conceptual structure a new concept belongs. For this purpose, either an existing conceptual structure (like dictionary, thesaurus or ontology) or automatically generated one can be used. It is reported in many papers, that pre-existing dictionaries often do not meet the user's needs for interesting concepts, or ontology like WordNet does not include proper nouns.

3. STATE OF THE ART OF SEMANTIC EXTRACTION OF DOCUMENTS

There is a strong requirement in the Information retrieval research area in recent years due to the enormous growth in the number of text databases available on-line and need for better techniques to access this databases[4][13]. Since the future web –semantic web-consists of pages containing texts and semantic mark up, the current IR techniques are unable to exploit the semantic knowledge within the documents and hence cannot give precise answers to precise queries [5].Information retrieval models can be distinguished such as Keyword-based Information Retrieval Model and content based IR model. In the first one, Information retrieval model is based on keyword indexing systems, frequency of occurrence of a keyword is taken into account[6][14]. Using the first one we can do data retrieval and latter gives us Information retrieval. As the name implies, the main task in information retrieval is to find information rather than data .Keyword based access can do the data retrieval which aims to provide data sets which fit the keywords of a query.

During the semantic web period, the meaning of a text or a word is depending on the conceptual relationships to objects in the world rather than to the contextual relations found in texts or dictionaries. The concepts of the words, names and nouns in the documents will be mapped to the concepts in wordNet.

A content of an information object is described by a set of concepts in Content based IR model. Concepts can be extracted from the text by categorization. The main problem facing is the non existence of a conceptual structure for mapping objects to concepts used in the user query. The nouns or names in the input documents should be mapped to concepts in wordNet in a conceptual structure. Since wordNet groups words together

based on their meanings (synsets) [10], the groups can be interlinked using the relationships such as is-a and part-of/member-of. Since concepts are abstract entities, representing them is a big problem. Words represent concepts in human language but the mapping from words to concepts is many-to-many. That means one concept may be represented with many different words (synonym) and one word may represent many different concepts (polysemy) [7]. This mapping problem is known as Word Sense Disambiguation [8].

4. METHODOLOGY

We are presenting a method for semantic concept extraction from the text document with the help of WordNet by reducing the above existing problems in this area. WordNet is such an existing general ontology from which a sub ontology can be generated[10]. Synsets are interlinked by means of conceptual-semantic and lexical relations. WordNet can be queried according to the input text document and create classes of concepts based on the results of the query. Extraction of semantic concepts from the keywords is the initial phase of actual construction of the ontology which will be covered during the next phase of this project. To extract semantic concepts, a word in the text document is taken as input which one wants to improve the knowledge, WordNet is searched about this word and different meanings of words are taken from which initial documents are collected. Terms frequencies are then calculated and compared with each group and concept with highest frequencies will be displayed first. The second phase of the construction of ontology will be done using the result of the first phase.

In part of this study, we used WordNet 2.1 as our knowledge base. WordNet is a large lexical database of English. Nouns, verbs, adjectives and adverbs are grouped into sets of cognitive synonyms (synsets), each expressing a distinct concept.

4.1 High Level Design

4.1.1. Extraction of semantic concepts from documents

To identify suitable concepts from WordNet by analyzing the text document is the main challenge. When retrieving/identifying concepts. it is important to make sure that irrelevant concepts should not be extracted and relevant concepts should not be discarded.Words can represent multiple concepts and different words can represent the same or very similar concepts.The input text documents should be analysed and process to extract relevant information.To retrieve semantic concepts form the document, a four-stage extraction process is invoked[3]. This includes: (1) concept selection, (2) relationship retrieval, (3) constraint discovery.

4.1.2 Term weighting

One of the simple representations of documents in information retrieval is a collection of terms corresponding to all the words contained in the documents. The classical approach for doing this is term weighting. Weights indicate the frequency of words appearing in the document. The frequency (number of occurrences) of each word can be calculated by constant rank frequency law Zipf

$$\text{Frequency.rank} \approx \text{constant} \qquad (4.1)$$

;where rank is obtained, by sorting words by frequency in decreasing order. Hence the frequency of a given word multiplied by the rank of the word is equal to the frequency of another word multiply by its rank. A method to find term weighting is term frequency $\boldsymbol{tf_{i,j}}$ where each word $\boldsymbol{t_i}$ is calculated as per the number of occurrences of the word associated with the term in document d_j. One popular global weight is inverse document frequency which assigns the level of discrimination to each word in collection of terms in a document. A word appearing in most items should have lower global weight than words appearing in few items.

$$idf_i = \log N/n_i \qquad (4.2)$$

here $\boldsymbol{n_i}$ are the No of item in which term $\boldsymbol{t_i}$ appear and N is the total number of documents in collection. The approach which states that a weight to each word in a document depending not only on the local frequency of the word in the item, but also the resolving power of that word in the collection of document is known as ***tf.idf*** (term frequency-inverted document frequency).

5. IMPLEMENTATION AND RESULT

The retrieval of semantic concepts for the given text document have been implemented successfully using Java, the most powerful platform independent language .The retrieved semantic concepts will be used to generate the taxonomy for the ontology generation. JDK and Net Beans IDE 6.7.2 are used to develop the application. WordNet 2.1 is used as the knowledge source. The extraction of the required concepts has been done by using the following steps.

1. Text documents which are to be extracted are stored in a folder called input. Any number of text documents can be stored in the above folder. Fig 2.

Fig 2. Selecting documents from input folder

All the text documents are read from the input folder and adding to the array list. The stream of texts is broken into words, phrases, symbols, articles, pronouns and prepositions. (Tokenization).Unwanted terms like articles, pronouns and prepositions etc. are removed from the array list. Stemming is used to generate a group of words of nouns from the present set of words. At the end of stemming process we get a group of nouns from all input documents. Frequency of the each word in the group is checked in each document and the whole documents using the formulae Ttf_idf= Math.log10 (tdf+1)*Math.log10 (N/NT). (fig.4). The word which gets highest frequency weight will come at the root of the taxonomy. Synonyms, hypernyms and hyponyms are extracted for each word with the help of WordNet by the usage of appropriate functions.

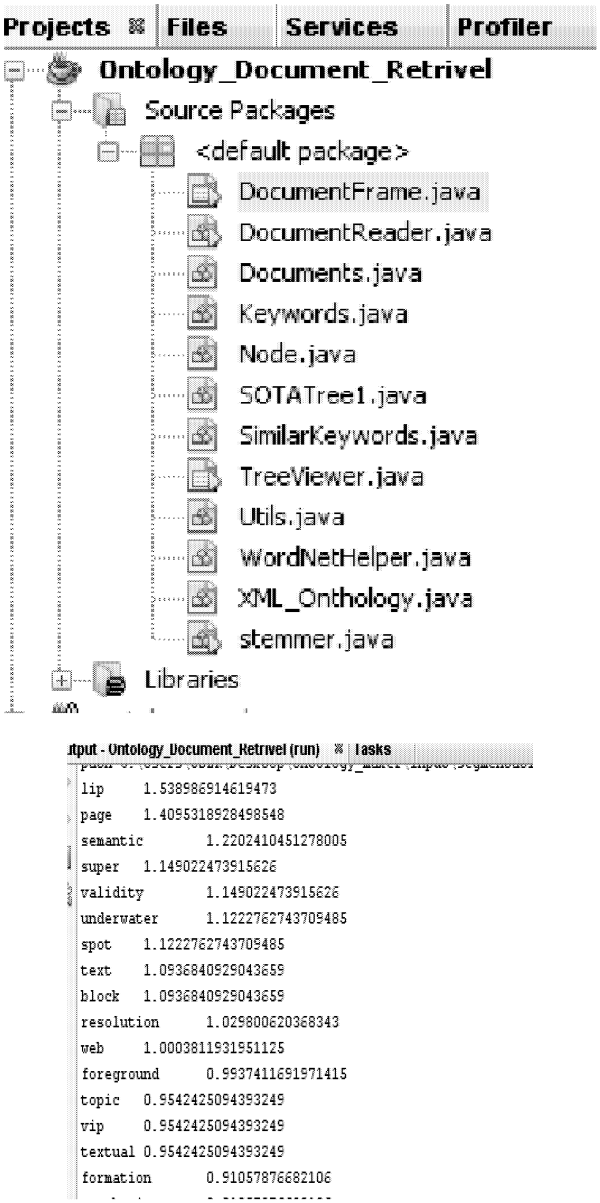


Fig. 3- calculation of frequency weightage

The sample input document is shown below (fig.4) from which the stemming done. The frequency weightage calculated is shown in the ontology creator. Synonym, Hypernym and hyponym are extracted for the word 'interval' is shown below. For the word 'interval', frequency calculated is 1.30102999 and synonyms extracted are time interval,separation,interval etc.. Hypernymas are credibility, credibleness, believability. Hyponym are effect and force.

pixel center for every grid intervals. In order to make super pixels usable, they must fast, easy to use and high quality segmentation should be done. Unfortunately most of the super pixel Methods do not meet all these requirements, as they often suffer from high computational costs. This approach address these issues and produces high quality uniform super pixels efficiently than other methods. This shows the over segmentation process in which the super pixels are grouped to larger sub regions. These sub regions are formed using midlevel

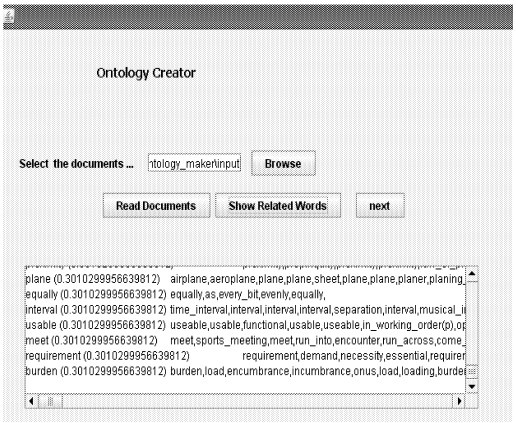


Fig.4.sample input document and extracted words

Synonym represents different words with almost with similar meaning. Hypernyms and hyponyms represent a general category and a specific instance of that category. A hyponym shares a type-of relationship with its hypernym. For example Toyota, Ford, Nissan are all hyponyms of Car (their hypernym) which in turn a hyponym of Vehicle. *Is-a* relationship is generally used to represent the hyponym and hypernym relationships. For example *Car is-a Vehicle* can be used to describe the hyponymic relationship between car and vehicle. WordNet 2.1 browser is used to find the synonyms, hyponym and hypernym of the input document. In the tree view,the frequency weighteage, synonyms, hypernyms and hyponyms are shown in hierarchical way.

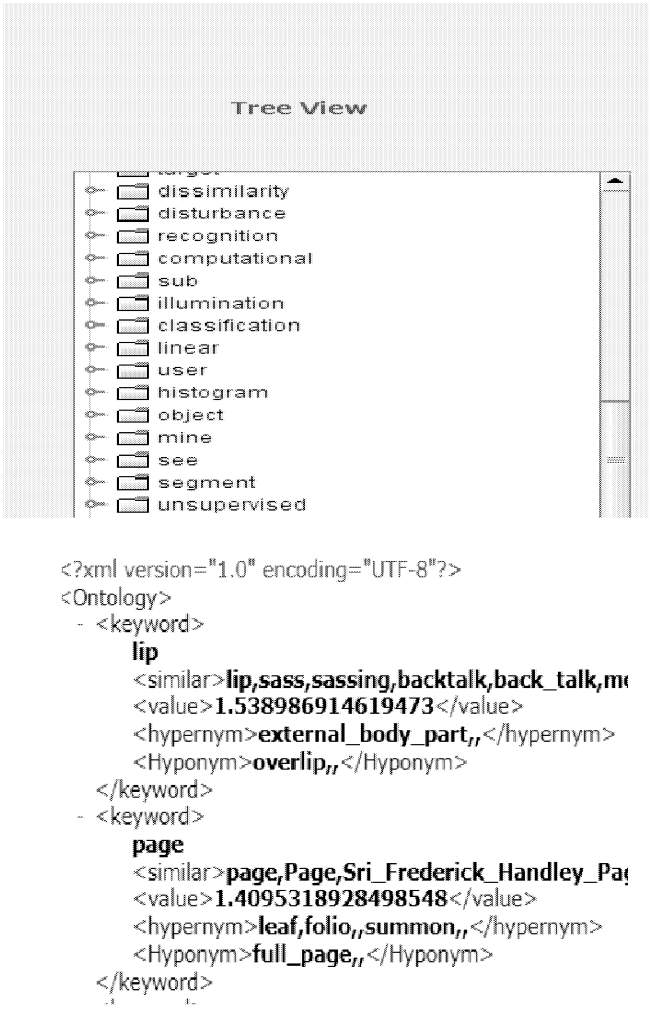


Fig.5 Tree view and Xml file

6. CONCLUSION AND FUTURE WORK

Keyword based Retrieval leads to inaccurate and incomplete results when different keywords are used to describe the same concept in the documents and in the queries. Concept based retrieval methods are the solution for this scenario. This gives the benefit of getting semantic relationships among concepts in finding relevant documents. Also elimination of irrelevant documents by identifying conceptual mismatches is another benefit obtained from this. The Initial step is the concept based extraction from wordNet. Words and phrases are the linguistic representatives of concepts. The extraction of the concepts is achieved by breaking into words, phrases, symbols, articles, pronouns and prepositions. (Tokenization)

Unwanted terms like articles, pronouns and prepositions etc. are removed from the array list. Stemming is used to generate a group of words of nouns from the present set of words. At the end of stemming process we get a group of synonym, hyponym and hypernyms of each word. Frequency of the each word in the group is checked. At the end of this phase, semantically related words and their relationship will be extracted from the input document with the help of knowledge base, WordNet. These concepts and their relationships are the source for automatic construction of ontology. The construction of ontology from the extracted words is identified as the future work of this paper.

REFERENCES

[1] Hele-Mai Haav, An Application of Inductive Concept Analysis to Construction of Domain-specific Ontologies. Akadeemia tee 21, 12618 Tallinn, Estonia ...

[2] W.Bruce Croft, What Do People Want from Information Retrieval? www.dlib.org dlib /november95 /11croft.html.

[3] Urvi Shah, Tim Finin, Anupam Joshi, R. Scott Cost, James Mayfield, Information Retrieval on the Semantic web, http://www.csee.umbc.edu/~finin//papers/cikm02/cikm02.pdf.

[4] Rifat Ozcan, Y. Alp Aslandogan, Concept-based Information Retrieval Using Ontologies and Latent Semantic Analysis, *www.cse.uta.edu/research/pblications/Downloads/CSE-2004-8.pdf*

[5] Hele-Mai Haav,,Tanel-Lauri Lubi, A Survey of Concept-based Information Retrieval Tools on the Web, 5th East-European Conference, ADBIS 2001 Vilnius, Lithuania: (2001) .

[6] Ide, N., J.Véronis. Word Sense Disambiguation: The State of the Art. Special issue of Computational linguistics on Word Sense Disambiguation, 24:1, Pages 1-40, 1998.

[7] Christian Safran, A Concept-Based Information Retrieval Approach for User-oriented Knowledge Transfer, Master's Thesis, 10th December 2005.

[8] http://wordNet.princeton.edu/

[9] Fensel, D]2001], Ontologies: Silver bullet for knowledge management and electronic commerce. Springer-Verlag, Berlin.

[10] Asuncion Gomez Perez and V. Richard Benjamins [1999], Overview of Knowledge Sharing and Reuse Components: Ontologies and Problem-Solving Methods. IJCAI- Workshop on Ontologies and Problem-Solving Methods: Lessons Learned and Future Trends.

[11] R. Bodner and F. Song[1996], "Knowledge-based Approaches to Query Expansion in Information Retrieval," in Proc. of Advances in Artificial Intelligence, pp. 146-158, New York,Springer.

[12] Lopez, M.F.[1999], "Overview of the methodologies for building ontologies". Proceedings of the IJCAI- 99 Workshop on Ontologies and Problem-Solving Methods (KRR5), Stockholm, Sweden, August.

[13] G.Madhu and Dr.A.Govardhan Dr.T.V.Rajinikanth[2011] , Intelligent Semantic Web Search Engines: A Brief Survey.

[14] Henrick Bulskov Styltsvig.Ontology based Information Retrieval , http://coitweb.uncc.edu/~ras/RS/Onto-Retrieval.pdf.

GAMING

GLIDE:
TOWARDS AN INTELLIGENT KINESTHETIC LEARNING SPACE FOR TEACHING DANCE

Penny de Byl and James Birt
Bond University, Australia
E-mail: pdebyl@bond.edu.au, jbirt@bond.edu.au
Muqeem Khan
Northwestern University in Qatar
muqeem@usa.net

KEYWORDS
Learning, Teaching, Motion-Capture, Dance, Artificial Intelligence, XBox Kinect

ABSTRACT

In this paper we present the overview for a proposed artificially intelligent learning environment which instructs dance. The Guided Learning and Immersive Dance Environment (GLIDE) will teach dance via a virtual instructor that senses student movement in the real world via a Microsoft Kinect. As a student attempts to mimic the movements of the instructor the system will extract skeleton movement and joint rotations to evaluate the dance performance. The analysis of the student's movements will be fed into the system's artificial intelligence which will provide real-time feedback and customised targeted instruction to assist in bettering the student's performance. This technology has far reaching application from tradition dance instruction to preservation and dissemination of intangible cultural heritage.

INTRODUCTION

Dance can be learned from verbal description, spatial guides or instructor imitation. Verbal description works well if the dance isn't too complicated, however as the movements become more complex and increase in speed, words are too deliberate, general and linear to embody the complexity of precise movement. The spatial guide method used in old-fashioned dancing, which forms the basis of numerous computer games such as Dance Dance Revolution, instructs players through tasks of hitting a sequence of spatial targets in rhythmic time. This provides a greater kinesthetic learning experience than verbal description alone. In this paper we outline the proposed GLIDE system being developed by Bond University in Australia, Northwestern University in Qatar and Hanze University, The Netherlands. Through it we will evaluate motion-sensing technologies, developed for use in contemporary digital computer games, through the design of an intelligent kinesthetic learning space. This project aims to produce the following outcomes 1) create a proof-of-concept intelligent kinesthetic learning space, 2) evaluate and explore knowledge transfer opportunities afforded by kinesthetic peripheral games technology; and 3) perpetuate the notion of serious games in the realm of kinesthetic learning space.

GLIDE: A Guided Learning and Immersive Dance Environment

In the hierarchy of visual and kinesthetic experiences with GLIDE, the cognitive subprocesses play the leading role. The peripheral and focused visions constantly scan the application on the screen for learning cues and feedbacks. This transection of information from user's interaction interacts simultaneously with short term or working/active memories, depending on the user's choices and sudden emotional or structural changes from digital elements presented by GLIDE application. Research shows that human have limited capacity for coding, storage and retrieval of information from the screen (Lang, 2006). The limited capacity model categorizes television viewers as information processor. Viewers process the information on the screen with parallel cognitive sub-process of coding, storage and retrieval of messages that they have exposed.

The main difference in our approach to designing GLIDE with respect to existing systems is in determining the optimal configuration between the artificial intelligence (AI) techniques used to provide student feedback and the accuracy with which motion is being captured and analysed. In short, we endeavour to determine a point in the design at which learning effectiveness does not improve with more complex processing and feedback.

GLIDE will include the projection of a virtual dance instructor character programmed with real world dance routines captured from expert dancers and a dance area for students (as illustrated in Figure 1). Students entering the space will receive dance instruction from the character. The motion-sensing devices will track the students' movements and the character will give them feedback on their performance.

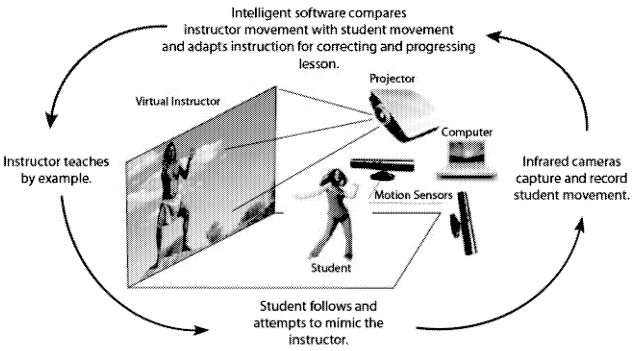


Figure 1: GLIDE Architecture

Pose Evaluation

Pose evaluation identifies in real time or in post processing how a human body and/or individual limbs and joints are configured in a given scene. There are many techniques revealed in contemporary literature for determining the accuracy of human pose capture (Moeslund & Granum, 2001). These techniques can be coarsely grouped into two classes, 2D image sequencing and 3D data point analysis (Raptis, Kirovski, & Hoppe, 2011). Methods using 2D image sequences focus on extracting image and pattern features of the movement in space and time. Examples of these features include contours to identify similar motions between the recorded movement and the user (Aaron, James, & IEEE Computer, 2001), and matching systems against pre-recorded poses or templates of human actions such as walking, waving and running (Schuldt, Laptev, & Caputo, 2004). The use of 3D information provides many benefits over the pure 2D image based approaches. The data does not suffer from sensitivity such as vantage point, scale and light changes (Raptis, et al., 2011). The 3D approach allows tracking of individual parts of the human body enabling temporal analysis of human body dynamics or simply dance movement in time.

The method used for determining pose accuracy in our project will depend on the level of detail required to achieve the optimal learning situation. Using the provided pose evaluation examples as a baseline we propose to experiment using: 1) distinct time intervals and 3D joints, 2) musical beat intervals and 3D joints, 3) realtime 3D joints, 4) approximate 3D joints in real time with simple Gaussian analysis, 5) approximate 3D joints in real time with inverse kinematics and Euclidean distance The goal is to generate a series of algorithms from these pose evaluations that range in complexity and can adapt to the user.

Proposed Efficacy and Affordance

The current prototype, shown in Figure 2, is implemented with the Unity 3D game engine. It has been developed to begin evaluation on the first configuration of visual accuracy feedback and the use of Euclidian distance to determine player pose accuracy. Although the image illustrates and Australian aboriginal dancer the design will change to accommodate our first study which will evaluate the use of GLIDE in teaching Hip Hop. This decision is primarily based on the research team's access to a Hip Hop dance instructor and a large cohort of secondary school children.

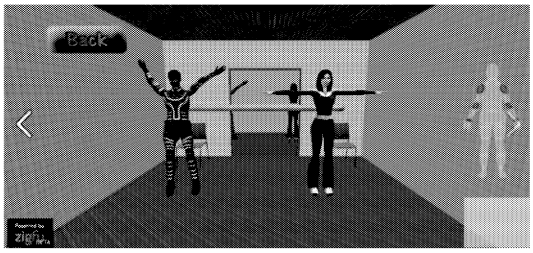


Figure 2: GLIDE illustrating the instructor on the left, the student's avatar and an accuracy display highlighting the joints of the student which are not in line with those of the instructor.

SUMMARY

GLIDE offers a well crafted kinesthetic learning space with playfulness that can easily foster the active, purposeful and entertaining user interaction with observational learning opportunities. With the marketplace entry of the Nintendo Wii, Playstation Move and Microsoft Kinect motion capture technology is now available to everyone. While there are a plethora of serious applications of this technology being examined, there is little research into its efficacy with respect to teaching dance. This project represents a cutting-edge and necessary examination of motion-sensing technologies and their potential for kinesthetic knowledge transfer to a wider international community. While the project outlined herein will examine the design parameters for a truly effective virtual dance tutor, the applications stemming from such are further reaching. For example, the preservation and teaching of cultural dances is an important focal area in ICH research. The use of these technologies not only provides tremendous showcasing opportunities for dance and related kinesthetic domains but also guarantees spontaneous, undirected learning experiences for people of all ages (Tanenbaum & Bizzocchi, 2009).

REFERENCES

Lang, A. (2006). Using the limited capacity model of motivated mediated message processing to design effective cancer communication messages. Journal of Communication, 56, S57–S80.

Aaron, F. B., James, W. D., & IEEE Computer, S. (2001). The Recognition of Human Movement Using Temporal Templates.

Moeslund, T. B., & Granum, E. (2001). A Survey of Computer Vision-Based Human Motion Capture. Computer Vision and Image Understanding, 81(3), 231-268. doi: 10.1006/cviu.2000.0897

Raptis, M., Kirovski, D., & Hoppe, H. (2011). Real-time classification of dance gestures from skeleton animation.

Tanenbaum, J., & Bizzocchi, J. (2009). Rock Band: a case study in the design of embodied interface experience. Proceedings of the 2009 ACM SIGGRAPH Symposium on Video Games (pp. 127–134).

ANALYSIS OF INTELLIGENT TECHNIQUES IN GAMES

Rula Al-Azawi
Gulf College, Muscat, Oman
rulaalazzawi@yahoo.com
rula@gulfcollegeoman.com

Aladdin Ayesh
DMU University, Leicester,UK
aayesh@dmu.ac.uk

Khalfan Abdullah AL-Masruri
Higher College of Technology, Muscat, Oman
khalfan.almasruri@hct.edu.om

Ian Kenny
DMU University, Leicester,UK
ikenny@dmu.ac.uk

ABSTRACT

Many Artificial Intelligence (AI) techniques used in the new generation of games. This paper presents the most popular techniques such as finite state machine, fuzzy logic, neural networks, genetic algorithms, agent and machine learning. It explains the most important AI requirement criterion needs to appear in games and how AI techniques could be used with different type of games.

KEYWORDS

Artificial intelligent, Computer Games, AI techniques, AI criterion, Game Agents.

1 INTRODUCTION

Computer games are an increasingly popular application for AI research and a big software industry. A huge variety of games are currently available, the requirements of such games are very different from those of the games normally studied in AI. Recently AI has been used for solving some issues in classical games. Usually computer games refer to either the designing of behavior, platforms and decision-making process. In the new generations of the computer games, new techniques from the AI fields are used such as autonomous agents, decision making, scheduling, path finding and learning would appear in the new kinds of games to be more important than classical games.
This paper is organized as following: section 2 describes the background of AI techniques in games; section 3 describes an analysis of games, games classification, games requirement and user expectation; section 4 explains analysis of AI techniques criteria; section 5 illustrates AI techniques in games; finally section 6 presents critical review with conclusion.

2 BACKGROUND

Many of AI technique used with games.; the most popular techniques include finite state machine, scripting fuzzy logic agent...etc, and a few game developers use decision trees, neural networks and genetic algorithms.[1]
This section reviews some of AI techniques that used for games development.

- Finite state machine

Finite state machines (FSM) are a set of finite number of states, set for input, output and state transition function. The FSM in games divide game object's behavior into logical states. In this case the object has only one state for each different type of behavior. FSM are realized by simple if-then statements, it used the graphical representations which is part of the nine diagrams defined by the Unified Modeling Language (UML)[2].

- Fuzzy logic

The Boolean logic contains only true of false while Fuzzy Logic allow us to use intermediate values such as 'rather hot' or 'very fast' that are used to describe continuous, overlapping states can be used in mathematical. The fuzzy logic is able to represent a concept using small number of fuzzy values. Fuzzy logic used in decisions that can be made based on incomplete or erroneous data that cannot be used in Boolean logic. It is useful in decision-making, behavior selection and input/output filtering.

- Neural network

Neural networks contain a number of relative components linked in the system that will be able to produce

output based on the identification of patterns in data. Many games that used neural networks could 'learn' through experience or training and has the ability to make decisions.

- Genetic algorithm

Genetic Algorithm (GA) used to find an optimal solution and used in machine learning that need to evaluate and find a solution to a problem. GA started with a small number of initial strategies, and then creates an entire population to find solutions and evaluating each population to solve the problem.
GA is a suitable solution when we have problem or game that contain large enough search domain. It is also useful with nonlinear problems.

- Intelligent agent

In game AI research, terms such as agents are often used. Intelligent agents are software agents that have been seen in environment and act in that environment to reach their goals by using some kind of evaluation function which is called heuristic function to help the agent to decide the best value. Also agents are studies in many problems in AI. The agents could define as follows:"Agents can be defined to be autonomous, problem solving computational entities capable of solving effective operations in dynamic and open environments"[3].
Games are ideal environments for agents because it is use realistic environments but there is limitation in available information where decisions must be made under time and pressure constraints. Generally, agents in games are sets of FSMs or could used any other techniques or any combination of some or all of these techniques that work on their particular problems and send messages to each other.

- Machine learning

The process of learning in games generally implies the adaptation of behavior for opponent players in order to improve performance. Machine learning techniques may be able to cover the search space in computer games and efficiently search for successful combinations of parameters[4]. Machine learning can either take place on line or off line.

- Decision making

The decision making is necessary for games. Intelligent games need to be able to affect the entities using two strategy AI push and entity pull. In AI push we separate the element of the game architecture. While the entity pulls is best work for the games deals with simple entities. Those simple entities call the AI system when need to think or update itself, the shooter games are good example for using this strategy.
The Goal –oriented action plan (GOAP) is a decision making architecture that defines first necessary conditions to satisfy the goal, second steps to satisfy the goal in the real time and the players found set of actions that could leads him to the goal state.

- Fuzzy state machine

A fuzzy state machine (FuSM) is mixing between fuzzy logic and FSM. Instead of only used "on" or "off" state, we adds "almost on" or "little on". FuSM are increase used in game play because it adds interesting and varied response by non playing character (NPC)l. Therefore, the player can interact with NPC that can be various degrees of 'mad', 'wounded' or 'helpful' and the player could test different experience and get different outcomes in similar situation in each time play the game.

- Scripts

Scripts are usually used to control games from outside. It contains the benefits that it is understandable, easy to implement and executable. It is deterministic such as FSM so it needs from the developer to cover all aspects of the character behavior. In most modern games are used about hundreds of different parameters and scenarios.

- Prediction

Prediction is the ability to effectively anticipate an opponent's next movie is crucial in an adaptive system. There are many method used in prediction such as past pattern recognition or random guess to determine the next action to take. The past history can decide the best action against a player team which appears in tactical games.

- Path finding

The main algorithm for Path finding is A* which is used for performing fast search to find the optimal path to reach to the goal between two points in the graph or map. The optimal path in games not means the shortest but we need to take extra factors such as (type of terrain, number of enemies in the path and many other parameters depending on the game itself). In Figure 1. The pathfinder searches the entire possible path in a room then chooses the optimal path. [5]

Path finding is essentially for a solved problem in game development. In many games a large number of entities needs to find the unique path but multiple paths could found simultaneously with different threads. The system needs to keep track for the discover path so the same path no need to discovered more than once.

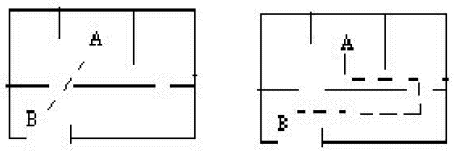


Figure 1: The two possible paths and the first one is the optimal path

3 ANALYSIS OF GAMES

A huge variety of games are currently available. Devising a single game classification is complex because there are many different categories for classification.

- Action games: This referred to the player played through a series of levels (set in a virtual world) containing a variety of enemies. Most of action games are the first person shooter (FPS).

- Sport Games: This referred to the game that deals with different kind of sport games and it creates virtual representations of those sport game.

- Adventure Games: This referred to the game tend to be plot based and involve activities such as exploration, information gathering and problem solving.[6]

- Strategy Games: This referred to the game are depend on create a military-type battle scenario. It is focus on it design not mainly on the characters but also on resource (building of defensive, offensive units and troops) and needs to organize and manage the battles.

- Role Playing Games (RPGs): This referred to the type of games usually assume the roles of characters acting in a fictional setting. The main RPG played with a handful of participants mostly used face-to-face. RPGs generally focus on the player undertaking a quest

- Real time Strategy games (RTS): the players must focus on logistics and resource production, as well as managing combat and war. These games usually involve quick decision making and fast reflexes. Players need to manage commanding armies and gathering resources at the same time. This can prove immensely challenging and most RTS games have developed many in game tools to help players deal with this task. Common examples of real-time strategy games: Age of Empires, Age of Empires 3, Empire Earth.[7]

Table 1 show us in general the most important AI requirement linked with games types followed by example of commercial games name.

Table 1: AI requirements for games.

AI requirement	Game type	Game name
Manage game world	FPS- RTS	Doom,Quake, unreal, CCTP, Enemy nations
Optimal solution	RTS	CDD
Path finding	Sport game racing game RTS	Dirt Track Racing
Decision making	Action game RTS	S.W.A.T2, CCTP
Learning	Racing, Fight, Action, Adventure, Strategy games	Dirt Track Racing , Battlecruiser 3000AD
Developing game strategy	RTS	CDD

In the next section an explanation of the possible link between AI game requirement and AI techniques on Table 1 will presents.

4 ANALYSIS AI CRITERIA TECHNIQUES

There are many AI techniques used in games as declared in the background section.
This section will focus to present the most important AI criteria and how it will be linked with game requirement. In general, game AI focuses on creating the appearance of intelligence in many ways as shown previously in Table 1. while Table 2 show us the main AI requirement criteria for games and how it links to AI techniques.

Table 2: AI requirements criteria .

AI requirement	AI techniques
Manage game world	FSM, FuSM, Script
Optimal solution	GA
Path finding	A*, Streeing
Decision making	NN , Agent, Fuzzy logic
Learning	NN, GA
Developing game strategy	GA

5 AI TECHNIQUES IN GAMES

This section explores the use of AI techniques which have been used with a specific game. According to

Table 1 and Table 2 a discussion of games in which AI techniques are undertaken as following:

- FSM technique in games

For game AI, the possible ways to use an FSM are endless. It could be applied to manage the game world, maintain the status of the game. Also it could module unit behavior in RTS. Also it could be used to parse input from the human player or even to simulate the emotion of a NCP [8]. FSM are a natural choice for game developer when designing NPC. It is used in most commercial computer games or video games such as Age of Empires or in FPS video games such as Quake, Quake 2, Doom, and Half Life or in real-time strategy game such as Enemy Nations.
The games that are using FSM could also use another techniques such as fuzzy state or NN.

- Fuzzy logic technique in games

Fuzzy logic is used to determine how frightened they are of the player, for NPC to decide how much they such as the 'player. The commercial games mostly used this technology included in video game such as Platoon Leader; real-time tactics Action game such as SWAT 2; strategy game such as CCTP. If we have non linear problem or there is no simple solution or where NN are not applicable then the fuzzy logic it will appropriate.

- Neural Network technique in games

The neural network can be improve continuously, which means the player will be challenged to change the style of the strategy of play and mustn't reuse same strategy repeatedly. Neural network has been used in adventure games such as Battle-cruiser 3000AD; racing games such as Dirt Track Racing; strategy games such as Fields of Battle. The most interested application of neural network in AI is 'Battle-cruiser 3000AD', in this game the NPC controlled by a neural network. Also the action games 'Heavy Gear', which the player controls a large mechanized robot.

- Genetic Algorithm technique in games

GA is effective when the domain knowledge of games needs expert knowledge.Also GA used if no mathematical analysis is available or if the traditional search methods fail or if search space is poorly understood.
Genetic algorithm is used in real time strategy game(RTS) to adapt the computer strategy to exploit the human player's weakness and to define the behavior of individual units rather than group. 'Cloak, Dagger, DNA' (CDD) are examples of used GA in RTS games. In CDD the NPC and the players shared in 'DNA' that responsible for monitors and stores performance[3]. GA

Aspects of autonomy

	Reactivity	Supervision by higher level agents	Visible inter-agent communication	Complexity of decision making	Variety of action repertoire	Learning	Situatedness	Temporally continuous	Goal-oriented
World of Warcraft	Low	None	Low	Low	Low	No	Yes	Low	Yes
Half Life 2 (Single player mode	High	None	High	Low	High	No	Yes	Low	Yes
Supreme Commander	Low	High	Low	Low	High	No	Yes	High	Yes
Warcraft 3	Low	High	Low	Low	High	No	Yes	High	Yes
Black & White 2	Low	Low	Low	High	High	Yes	Yes	High	Yes
Tomb Raider Anniversary	Low	None	High	Low	Low	No	Yes	Low	Yes
Unreal Tournament (2004 (bots	High	None	High	High	High	No	Yes	High	Yes

Figure 2: link between popular agent characteristics and popular games.

is used in RPG or FTS to evolve behaviors of characters and events.

- Intelligent agent technique in games

The agent have the ability to take decisions and perform tasks to reach the goal as human done. Every game that includes AI can be said to be using an agent of some form. In FTS or in RPG the monster would be an active agent and more suited to simply reacting to what is currently happening in the game. While in strategy game needs to be careful when we plan the movement and what will happen later in the game. A good architecture for RTS agent is necessary to ensure success. In Empire Earth which is a RTS that contains several components called managers. There are managers for the civilization, building, units, resources, research and combat. The civilization manager is the highest level manager which responsible for the player development and coordination between the other managers. The other managers are lower-level which responsible for send requests and reports to each other.
In reference[9] the author create Figure 2 which is linking between the most popular agent characteristics and the most popular games.

- Machine learning technique in games

The machine learning are used in NPC, also at development time; machine learning techniques can be employed to automate the creation of intelligent NPC behavior. Machine learning is appears in old games such as tic-tac-toe, backgammon, go, Othello, and checkers .Recently; machine-learning techniques have begun to appear in video games as well as in fight in first- and third-person shooter games and also in strategy games.[10]

- Fuzzy state machine technique in games

FuSMs were used in FPS to make the enemy appear intelligent. Based on the element of the battle situation, fuzzy logic is used such that enemy characters that can decide to run away when losing a battle[4]. Also it used in CCTP which is a real time strategy game. FuSM used in RPG or hit points of NPC or agent and it is ideal for controlling the behavior of game characters and gives variation in actions and reactions.[6]
The Table 3 show us the links between the popular AI techniques and popular games

Table 3: AI techniques linked with games.

AI techniques	Game type	Game name
FSM, FuSM	FPS- RTS	Doom,Quake, unreal, CCTP
GA	RTS	CDD
A*	Sport game racing game RTS	
Agent	Action game RTS	S.W.A.T2, CCTP, Empire Earth
NN	Racing, Action, Adventure, Strategy games	Dirt Track Racing , Battlecruiser 3000AD, Heavey Gear
GA	RTS	CDD
Fuzzy logic	Adventure, Strategy games	SWAT, CCTP

6 CRITICAL REVIEW AND CONCLUSION

In previous section an analysis on the use of AI techniques in commercial games and in research games are written. The main goal of used AI in games is to simulate the intelligent behavior and to make the games believable challenge and fun at the same time. It is also helps to find the most optimal action for the available information. In the marketing they focus on features such as graphics and physics and know they start to put some AI features in commercial games such as the on line games world. But there is still less use of learning in commercial games. Some of the AI techniques are mostly use such as FSM, Agent, decision making this is because they are simple while other AI techniques need extra recourses such speed CPU and needs extra run time such ad GA and NN. Also many games could used more than one AI techniques depending on games requirements, furthermore some techniques links together such as NN could used together with fuzzy logic. At the end we need to make academics and commercial expert to use standard design and implementation in game AI.

References

[1] P. Spronck, M. Ponsen, and I. Sprinkhuizen-Kuyper, "Adaptive Game Ai With Dynamic Scripting," *Machine Learning*, vol. 63, no. 3, pp. 1–42, 2006.

[2] B. Yue, "The state of the art in game AI standardisation," *the international conference on Game research and development*, vol. 223, pp. 41–46, 2006.

[3] S. Yildirim and S. B. Stene, "A Survey On The Need And Use Of Ai In Game Agents," *Proceedings of the 2008 Spring simulation multiconference:SpringSim '08 (Ottawa, Canada)*, pp. 124–131, 2008.

[4] M. Ponsen and P. Spronck, "Improving Adaptive Game AI With Evolutionary Learning," *Computer Games: Artificial Intelligence, Design and Education (CGAIDE 2004)*, pp. 389–396 University of Wolverhampton., 2004.

[5] A. El Rhalibi, K. W. Wong, and M. Price, "Artificial Intelligence for Computer Games," *International Journal of Computer Games Technology*, vol. 2009, pp. 1–3, 2009.

[6] D. Johnson and J. Wiles, "Computer Games With Intelligence ," *Australian Journal of Intelligent Information Processing Systems*, vol. 7, pp. 61–68, 2001.

[7] J. Eldridge, "The definition and meaning of real time strategy games," 2012.

[8] P. Sweetser, "Current AI in Games: A review," tech. rep., Technical report, School of ITEE, University of Queensland, 2002.

[9] C. Weddle, "Artificial Intelligence and Computer Games," *Dissertation, Computer Science, Florida State University.*, pp. 1–8, 2008.

[10] E. Anderson, "Playing smart-artificial intelligence in computer games," *Proceedings of zfxCON03 Conference on Game Development, ZFX - 3D Entertainment*, 2003.

LATE PAPER

Ensemble Based Face Recognition

H. A. Khan, Marwan Al-Akaidi and A. S. Tolba
Arab Open University, HQ Kuwait
{h.khan, m.al-akaidi, a.tolba}@arabou.edu.kw

Key words:
Face recognition, wavelet transform, classifier ensemble, majority voting.

Abstract:

Face recognition is important for a variety of applications such as law enforcement, marketing, commerce and social networking. A number of techniques exist for designing and implementing face recognition systems based on features extracted from images of faces available to the system. Wavelet Transform based features and an ensemble of classifiers has been used to develop an efficient face recognition system. Experiments have been performed to test the performance of the proposed ensemble. Promising correct classification results have been achieved. The proposed ensemble is more stable compared to majority of base classifiers.

1. Introduction

The importance of face recognition systems is increasing due to the proliferation of CCTV systems and other digital imaging technologies such as feature and smart phones with powerful cameras. The applications of face recognition systems are important for a number of areas including law enforcement, surveillance, industry and commerce and social networking. The first two areas are well known whereas the last two areas are newer application domains. In commercial applications, customers entering a shopping mall may be recognized and then can be presented with customized offers based on their shopping history. In the social context it may be desirable in big gatherings and meetings to identify faces in scenarios such as if the participants are wearing Google glasses or other image capturing technology. Hence no formal introduction would be necessary in meetings since the person would have already been recognized by the system.

The breakdown of this paper is as follows. Section 2 presents the background about face recognition techniques available till date. In Section 3 we propose the ensemble-based face recognition system. Section 4 presents the experimental set up along with description of the database of faces that has been used. Section 5 presents the Wavelet Transform based feature extraction. Results and discussions are presented in Section 6. Conclusions have been presented in Section 7 and suggestions about future work are provided in Section 8.

2. Background on Face Recognition Techniques

The problem of face recognition has been studied extensively by various researchers. An elaborate survey of face recognition techniques has been presented in (Zhao et al. 2003). Subspace methods of face recognition are quite popular in the research community (Gregory and Baback 2011; Turk and Pentland 1991; Belhumeur et al. 1997). Various techniques associated with subspace analysis of images of faces have been presented in (Gregor and Baback 2011).In subspace methods a lower dimensional face space is assumed as subspace of the higher dimensional image space. The face space is then modeled as a principal manifold and the goal of subspace analysis is to extract the principal modes or basis functions of the manifold as discussed in (Gregory and Babak 2011).The Eigenfaces methods for face recognition in (Turk and Pentland 1991) was the initial work on subspace methods for face image analysis. Principal Component Analysis (PCA), Linear Discriminate Analysis (LDA) and Independent Component Analysis (ICA) have been used to study face recognition.

Elastic Bunch Graph Matching based face recognition system has been presented in (Wiskott et al. 1997). They have used a novel approach of bunch graph in order to achieve image graph extraction in their work on face recognition. Wavelets have been used for Iris and face recognition in (Daugman 2003; Daugman 2004). Wavelet analysis has been used to extract phase descriptions on features for recognition of iris patterns for personal identification. 3-D face recognition systems have also been studied by researchers (Heseltine 2005). 3-D methods promise improved performance compared to 2-D techniques by utilizing geometric information of the face images. Multi-classifiers have been used in (El-Bashir 2012) and a correct classification percentage (CCP) of 96% has been reported. In (Ebrahimpour2007) the Bagging technique has been used and a CCP of 97.5% has been achieved. The proposed ensemble base face recognition system in this paper is based on enforcement of diversity through variation of training data, training algorithms and neural network classifier architectures. Our ensemble has achieved a CCP of 97.5% using the 10-Fold cross-validation methodology and 95.6% for the more challenging 50% sample split case.

3. An Ensemble Based Face Recognition System

This paper presents a powerful ensemble of classifiers for face recognition purposes. We have used a group of nine

Neural Network based classifiers to perform face recognition. The face recognition architecture is shown in Figure 1.

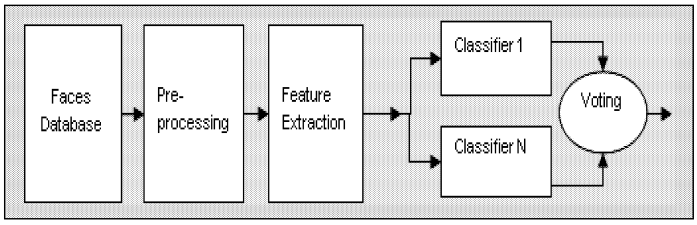


Figure 1: The architecture of the face recognition system

Our system uses the AT&T Database of Faces (formerly known as the Olivetti Research Lab (ORL) Database of Faces) (AT&T Laboratories Cambridge) as the bench mark data for face recognition. Our system performs basic image processing operations on the face images as a pre-processing step. Features are extracted using the 2-D discrete wavelet decomposition. The ensemble consists of 9 classifiers (N=9) which perform classification tasks on the feature set provided to them. The proposed architecture combines the decisions of the various classifiers based on the concept of majority voting.

The ensemble of classifiers based approach is an efficient technique that utilizes a combination of a number of weak classifiers to achieve better classification accuracy. We utilize diversity in our classifier ensemble by using 2 different types of classification algorithms i.e. one based on Feed-Forward Neural Network (FFNN) and the second is based on the Cascade-Forward Neural Network (CFNN). Diversity has also been enforced by using different neural network architectures. The majority voting scheme has been used to achieve final decision amongst the ensemble of classifiers.

4. Experimental Setup of the System

We have implemented the proposed face recognition system shown in Figure 1 using the Matlab programming environment. We use the ORL Database of Faces for face recognition which consists of face images of 40 subjects each having 10 images leading to 400 images in the database. We have used two methodologies to test the performance of our system i.e. the 10-Fold cross-validation methodology and the 50% sample split methodology. In the 10-Fold case, we divide the available images into 10 folds with the training set containing 90% of the images and the test or validation set containing 10% of images. At the end we average the performance of classifiers over the 10 folds. In the 50%case, we have used 50% of the images for training the base classifiers and 50% for testing purposes. This is a more challenging case since the system will be trained on fewer samples and then tested on more unseen number of samples. The Training Set and the Test Set thus consist of 40 classes having 5 images each. We have considered images of 2 spatial resolutions i.e. 64x64 and 112x92 in our work. It is mentioned here that using 10-fold cross validation gives improved classification accuracy since more sample images are used for training purposes. A snapshot of some images from the ORL database is shown in Figure 2.

The preprocessing block in the architecture performs basic image processing operation of histogram equalization on the images to make them suitable for recognition purposes. We perform discrete wavelet decomposition of the images to extract the feature set to be used by the base classifiers. The classifier ensemble block uses 4 feed forward and 5 cascade neural networks to perform classification of images. The final decision of the system is performed by the majority voting block. The Table 1 below presents information about the experimental setup.

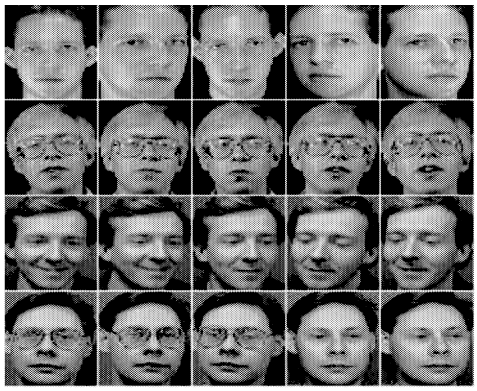


Figure 2: A snapshot of the images from ORL database (AT&T Laboratories Cambridge)

Table 1: Experiment setup

Item/Factor	Implementation in experiments
Database of face images	ORL
Total subjects in Database	40
Number of images of each subject	10
Total Number of Images	400
Testing/Validation methodologies	10-Fold and the 50% sample split case
Image resolutions	Size 64x64 and 112x92
Feature extraction methodology	2-D Discrete Wavelet Transform
Wavelets	Daubechies 1 (db1) and Symlet 4 (sym4)
Wavelet decomposition levels	Level 3 and Level 4

5. Wavelet Transform Based Feature Extraction

The Wavelet Transform is a well established technique for multi-resolution and multi-scale analysis of 1-D and 2-D signals (Daubechies 1992). The 1-D Wavelet Transform of a continuous time signal x(t) is given by (Mertins 1999):

$$W_x(b,a) = |a|^{\frac{-1}{2}} \int_{-\infty}^{+\infty} x(t)\psi^*\left(\frac{t-b}{a}\right) dt. \qquad (1)$$

The above shows that the Wavelet Transform is the inner product of the signal x(t) with a translated and scaled version of the wavelet $\psi(t)$.The parameters 'a' and 'b' are the scaling and translation parameters respectively.

We use the 2-D wavelet decomposition at levels 3 and 4 to extract features from face images. The decomposition gives both the approximation and detail coefficients at various decomposition levels. We have used the approximation coefficients as a feature set. An example of reconstruction using the approximation coefficients from Level 1 to Level 3 with the Daubechies wavelet 'db1' is shown in Figure 3.Reconstruction using detail coefficients is shown in Figure 4 for Level 1 to Level 3 where 'hd', 'vd' and 'dd' indicate horizontal, vertical and diagonal details respectively. It is clear from the figures that the Approximation coefficients carry more useful information about the faces than the detail coefficients. Also, Level 1 coefficients give better reconstruction compared to higher level decompositions but this is at the expense of increased feature vector length. In this work we have used approximation coefficients at Level 3 and 4 as our feature vectors since they give good classification performance with low feature vector lengths.

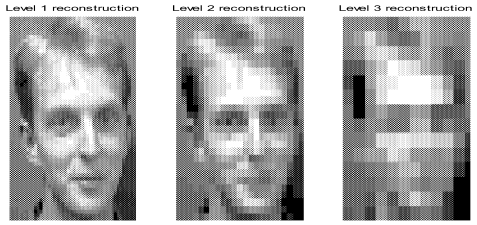


Figure 3: Reconstruction using approximation coefficients

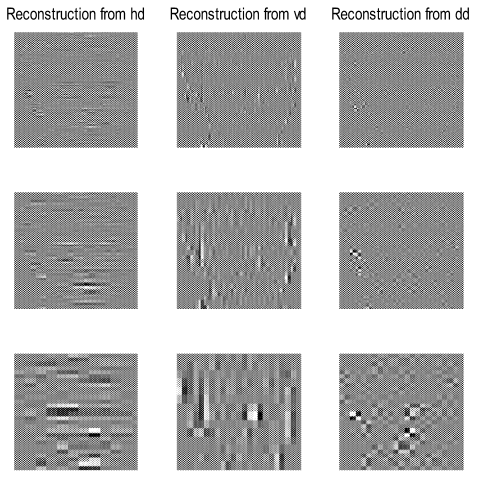


Figure 4: Reconstruction using detail coefficients

6. Results and Discussions

In order to test the face recognition system we have conducted a number of experiments. We have considered two distinct validation methodologies for our experiments i.e. we have conducted experiments using the 10-Fold cross-validation methodology and the 50% sample split methodology. In both cases, we have used the discrete wavelet decomposition for feature extraction and an ensemble of 9 weak classifiers. The effect of 2 wavelets i.e. 'db1' and 'sym4' and images at 2 resolutions i.e. size 64x64 and 112x92 have been considered. The classifiers we have used in our ensemble consist of the feed-forward and the cascade-forward neural networks.

6.1 The 10-Fold Cross-Validation Experiments

In the 10-Fold Cross-Validation experiments we have divided the available face images from the database into 10 sets. The 10-Fold methodology uses 90% images for the training set and 10% for the testing/validation set. The performance of the classification task is then averaged over the 10 folds. Table 2 shows the results of our simulations. It is clear from the table that the best performance of 97.75% correct classification has been achieved with wavelet 'db1' and images of size 112x92 for the 10-Fold case. Performances of the base classifiers for the 10-Fold case have been plotted in Figure 5.

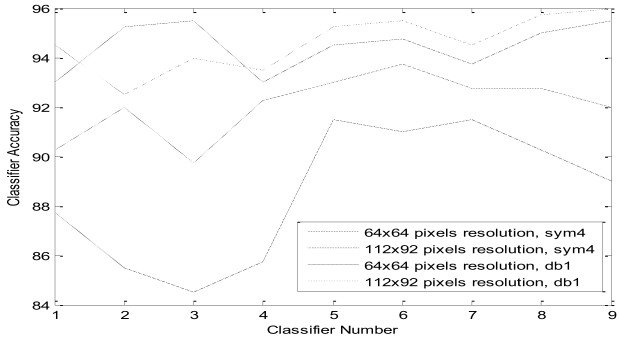


Figure 5: Accuracy of base classifiers for 10-Fold case

6.2 The 50% Sample Split Case

In this case we have split the available images into equal size training and test sets. We have then performed 5 runs of the experiments and have then averaged the results of the individual classifiers and those of our proposed Ensemble structure. Table 3 shows the results of our simulations. It is clear from the table that the best performance of 95.90% correct classification has been achieved with wavelet 'db1' and images of size 112x92 for the 50% case.

Table 2: Classification Results for the 10-Fold case (C: Classifier, E: Ensemble, STD: Standard Deviation)

sym4, 64x64 resolution										
Fold	**C1**	**C2**	**C3**	**C4**	**C5**	**C6**	**C7**	**C8**	**C9**	**E**
1	95.0000	100.0000	92.5000	95.0000	97.5000	100.0000	97.5000	97.5000	97.5000	100.0000
2	92.5000	90.0000	92.5000	92.5000	97.5000	95.0000	95.0000	92.5000	87.5000	97.5000
3	95.0000	85.0000	77.5000	82.5000	92.5000	87.5000	95.0000	92.5000	95.0000	95.0000
4	95.0000	92.5000	87.5000	85.0000	85.0000	92.5000	100.0000	95.0000	97.5000	100.0000
5	90.0000	85.0000	87.5000	85.0000	97.5000	90.0000	95.0000	95.0000	90.0000	97.5000
6	85.0000	82.5000	80.0000	85.0000	87.5000	85.0000	90.0000	82.5000	82.5000	87.5000
7	70.0000	70.0000	82.5000	80.0000	85.0000	85.0000	82.5000	85.0000	80.0000	82.5000
8	90.0000	87.5000	85.0000	87.5000	92.5000	95.0000	82.5000	85.0000	90.0000	97.5000
9	87.5000	87.5000	82.5000	85.0000	87.5000	87.5000	87.5000	90.0000	85.0000	90.0000
10	77.5000	75.0000	77.5000	80.0000	92.5000	92.5000	90.0000	87.5000	85.0000	95.0000
Mean	87.7500	85.5000	84.5000	85.7500	91.5000	91.0000	91.5000	90.2500	89.0000	**94.2500**
STD	8.2874	8.4820	5.5025	4.8663	5.0277	4.8876	6.0323	5.0621	6.1464	**5.7795**
sym4, 112x92 resolution										
1	87.5000	97.5000	95.0000	97.5000	92.5000	97.5000	90.0000	92.5000	92.5000	97.5000
2	95.0000	95.0000	87.5000	92.5000	95.0000	95.0000	100.0000	97.5000	95.0000	100.0000
3	92.5000	92.5000	97.5000	92.5000	100.0000	100.0000	97.5000	97.5000	97.5000	100.0000
4	85.0000	90.0000	87.5000	90.0000	92.5000	92.5000	90.0000	92.5000	85.0000	95.0000
5	87.5000	87.5000	90.0000	87.5000	85.0000	92.5000	90.0000	85.0000	90.0000	95.0000
6	97.5000	97.5000	90.0000	95.0000	97.5000	95.0000	92.5000	92.5000	95.0000	97.5000
7	92.5000	97.5000	92.5000	95.0000	95.0000	95.0000	95.0000	92.5000	95.0000	95.0000
8	90.0000	92.5000	80.0000	95.0000	92.5000	95.0000	95.0000	92.5000	95.0000	97.5000
9	85.0000	90.0000	82.5000	90.0000	87.5000	85.0000	90.0000	92.5000	87.5000	95.0000
10	90.0000	80.0000	95.0000	87.5000	92.5000	90.0000	87.5000	92.5000	87.5000	92.5000
Mean	90.2500	92.0000	89.7500	92.2500	93.0000	93.7500	92.7500	92.7500	92.0000	**96.5000**
STD	4.1583	5.5025	5.5840	3.4258	4.3780	4.1248	3.9878	3.4258	4.2164	**2.4152**
db1, 64x64 resolution										
1	97.5000	97.5000	100.0000	100.0000	97.5000	100.0000	95.0000	97.5000	100.0000	100.0000
2	100.0000	97.5000	100.0000	97.5000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000
3	97.5000	82.5000	92.5000	92.5000	95.0000	97.5000	95.0000	95.0000	92.5000	97.5000
4	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000
5	100.0000	95.0000	97.5000	100.0000	100.0000	97.5000	100.0000	97.5000	100.0000	100.0000
6	90.0000	87.5000	92.5000	87.5000	92.5000	92.5000	92.5000	95.0000	95.0000	95.0000
7	92.5000	92.5000	92.5000	92.5000	92.5000	92.5000	90.0000	95.0000	95.0000	95.0000
8	87.5000	90.0000	87.5000	90.0000	95.0000	95.0000	95.0000	95.0000	95.0000	95.0000
9	87.5000	90.0000	90.0000	85.0000	90.0000	90.0000	87.5000	92.5000	87.5000	92.5000
10	92.5000	92.5000	87.5000	90.0000	90.0000	90.0000	90.0000	90.0000	95.0000	95.0000
Mean	94.5000	92.5000	94.0000	93.5000	95.2500	95.5000	94.5000	95.7500	96.0000	**97**
STD	5.1099	5.2705	5.0277	5.5528	3.9878	4.0483	4.5338	3.1292	4.1164	**2.8382**
db1, 112x92 resolution										
1	92.5000	97.5000	100.0000	87.5000	97.5000	100.0000	92.5000	95.0000	100.0000	100.0000
2	97.5000	97.5000	100.0000	95.0000	95.0000	95.0000	95.0000	97.5000	100.0000	97.5000
3	97.5000	100.0000	97.5000	100.0000	100.0000	95.0000	95.0000	100.0000	97.5000	100.0000
4	82.5000	95.0000	87.5000	85.0000	87.5000	90.0000	85.0000	90.0000	90.0000	92.5000
5	87.5000	90.0000	95.0000	92.5000	87.5000	87.5000	95.0000	87.5000	90.0000	95.0000
6	97.5000	97.5000	92.5000	100.0000	95.0000	100.0000	97.5000	97.5000	95.0000	100.0000
7	97.5000	95.0000	100.0000	95.0000	100.0000	100.0000	100.0000	97.5000	100.0000	100.0000
8	90.0000	90.0000	97.5000	92.5000	97.5000	95.0000	95.0000	92.5000	97.5000	97.5000
9	95.0000	92.5000	90.0000	87.5000	95.0000	90.0000	87.5000	95.0000	92.5000	97.5000
10	92.5000	97.5000	95.0000	95.0000	90.0000	95.0000	95.0000	97.5000	92.5000	97.5000
Mean	93.0000	95.2500	95.5000	93.0000	94.5000	94.7500	93.7500	95.0000	95.5000	**97.7500**
STD	5.1099	3.4258	4.3780	5.1099	4.6845	4.4799	4.4488	3.9087	4.0483	**2.4861**

Table 3: Classification Results for the 50% case

sym4, 64x64 resolution										
Run	**C1**	**C2**	**C3**	**C4**	**C5**	**C6**	**C7**	**C8**	**C9**	**E**
1	75.5000	79.5000	75.0000	73.5000	84.0000	81.0000	81.0000	82.5000	83.5000	87.5000
2	70.5000	77.5000	73.0000	77.0000	81.0000	82.5000	80.0000	78.0000	79.5000	84.5000
3	73.5000	71.5000	72.0000	72.5000	81.5000	82.0000	80.5000	82.0000	81.5000	86.0000
4	76.5000	79.0000	78.5000	74.5000	83.5000	78.5000	84.5000	81.5000	81.5000	88.0000
5	73.0000	74.5000	78.0000	74.5000	83.5000	80.0000	81.0000	82.0000	79.5000	86.0000
Mean	73.8000	76.4000	75.3000	74.4000	82.7000	80.8000	81.4000	81.2000	81.1000	**86.4000**
STD	2.3345	3.3615	2.9069	1.6733	1.3509	1.6047	1.7819	1.8235	1.6733	**1.3874**
sym4, 112x92 resolution										
1	86.5000	88.5000	90.0000	87.5000	91.5000	93.0000	88.0000	92.5000	90.5000	95.5000
2	89.0000	88.0000	88.0000	93.0000	90.5000	89.0000	92.0000	87.5000	91.5000	94.5000
3	83.0000	84.0000	88.5000	89.5000	90.5000	88.5000	93.5000	94.0000	90.0000	95.0000
4	90.0000	89.0000	88.0000	89.5000	93.5000	93.0000	91.5000	89.5000	88.5000	94.5000
5	87.0000	89.0000	89.5000	88.5000	93.0000	91.0000	91.0000	91.0000	90.5000	94.0000
Mean	87.1000	87.7000	88.8000	89.6000	91.8000	90.9000	91.2000	90.9000	90.2000	**94.7000**
STD	2.7019	2.1095	0.9083	2.0736	1.3964	2.1331	2.0187	2.5348	1.0954	**0.5701**
db1, 64x64 resolution										
1	88.5000	86.0000	88.0000	86.5000	85.5000	89.0000	82.0000	86.5000	89.0000	91.0000
2	79.5000	88.0000	82.0000	87.5000	86.5000	87.0000	86.5000	88.5000	87.5000	91.0000
3	87.5000	89.5000	82.5000	87.5000	86.5000	87.5000	87.5000	88.5000	84.5000	90.5000
4	87.5000	88.5000	86.5000	86.0000	88.5000	86.0000	89.0000	86.5000	89.0000	92.5000
5	87.5000	87.5000	86.5000	89.0000	90.0000	85.0000	89.0000	88.0000	85.0000	91.0000
Mean	86.1000	87.9000	85.1000	87.3000	87.4000	86.9000	86.8000	87.6000	87.0000	**91.2000**
STD	3.7148	1.2942	2.6786	1.1511	1.8166	1.5166	2.8853	1.0247	2.1506	**0.7583**
db1, 112x92 resolution										
1	92.5000	94.5000	94.5000	93.5000	94.0000	94.0000	94.0000	92.0000	92.0000	95.5000
2	94.5000	91.5000	93.0000	93.0000	91.5000	93.0000	93.5000	92.5000	92.0000	95.5000
3	95.0000	90.5000	90.5000	93.0000	93.5000	97.0000	91.5000	91.0000	96.5000	96.5000
4	92.0000	92.0000	95.0000	94.0000	95.0000	95.0000	92.5000	97.0000	96.0000	96.5000
5	93.0000	97.0000	92.0000	93.0000	93.5000	92.0000	93.5000	95.0000	94.5000	95.5000
Mean	93.4000	93.1000	93.0000	93.3000	93.5000	94.2000	93.0000	93.5000	94.2000	**95.9000**
STD	1.2942	2.6315	1.8371	0.4472	1.2748	1.9235	1.0000	2.4495	2.1389	**0.5477**

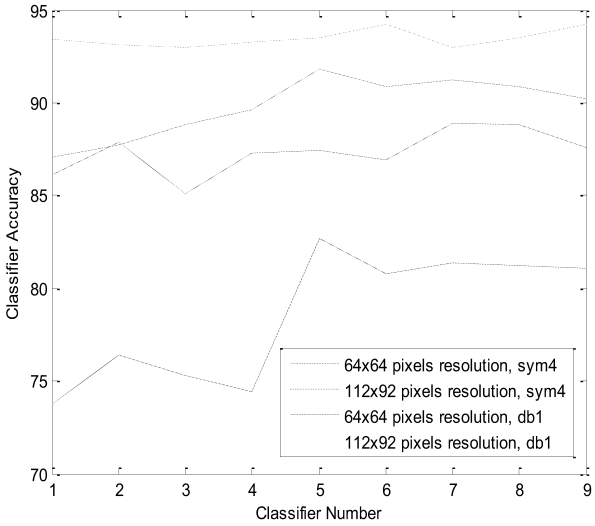


Figure 6: Accuracy of base classifiers for the 50% case

6.3 Stability Analysis and Resolution Effect

The proposed ensemble has been found to be more stable than the individual base classifiers. It is clear from Tables 2 and 3 that the ensemble exhibits less performance variations compared to the base classifiers. Also, Table 4 provides a summary of the results of the experiments. Increasing the facial resolution results in increasing the performance in a majority of cases by a value ranging from 0.75% to 8.3%.

7. Conclusions

Experimental results demonstrate the high accuracy of the proposed face recognition ensemble. Selection of the used wavelet type and the depth of wavelet decomposition affect the features extracted from images and the subsequent classification performance. The presented ensemble has achieved high correct classification percentages of 97.75% for the 10-Fold Cross-Validation case and 95.90% for the

more challenging 50% sampling case. By utilizing the power of combining diverse weak classifiers, high accuracy has been achieved. Furthermore, the ensemble has shown better stability of performance than the majority of base classifiers.

Table 4: Summary of all results (Accuracy)

Experiment	Ensemble
10 Fold, sym4, 64x64 pixels	94.25%
10 Fold, sym4, 112x92 pixels	96.50%
10 Fold, db1, 64x64 pixels	97.00%
10 Fold, db1, 112x92 pixels	**97.75%**
50% Sampling, sym4, 64x64 pixels	86.40%
50% Sampling, sym4, 112x92 pixels	94.70%
50% Sampling, db1, 64x64 pixels	91.20%
50% Sampling, db1, 112x92 pixels	95.90%

8. Future Work

The current work has utilized the 2-D discrete wavelet decomposition for feature extraction and a number of classifiers for decision making purposes. Both the feature vector generation and the classifiers involve a large number of parameters to be selected for optimal performance. In future we plan to investigate the optimal selection of these parameters to achieve highly accurate and robust results.

Acknowledgements:

The authors wish to acknowledge the research support provided by the Arab Open University (AOU).

REFERENCES

Abu-Rezq A. N., A. S. Tolba, G. A. Khuwaja and S. G. Foda. 1999. "Best parameters selection for wavelet packet-based compression of magnetic resonance images". Computers and Biomedical Research., 32(5), pp. 449-69.

Belhumeur P., J. Hespanha and D. Kriegman. 1997. "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 19, No. 7, pp. 711-720.

Daubechies I. 1992. "Ten Lectures on Wavelets", SIAM.

Daugman J. 2003. "Demodulation by complex-valued wavelets for stochastic pattern recognition". International Journal of Wavelets, Multiresolution and Information Processing, Vol. 1, No. 1, pp. 1-17.

Daugman J. 2004. "How Iris Recognition Works". IEEE Transactions on circuits and systems for video technology, VOL. 14, NO. 1.

Ebrahimpour H. and Kouzani A. 2007. "Face Recognition Using Bagging KNN". International Conference onSignal Processing and Communication Systems, Gold Coast, Australia.

El-Bashir M. S. 2012. "Face Recognition Using Multi-Classifier". Applied Mathematical Science, Vol. 6, no. 45, 2235-2244.

Garcia C., G. Zikos and G. Tziritas. 2000. "Wavelet packet analysis for face recognition". Image and Vision Computing, Vol. 18, No. 4, pp.289-297.

Gregory S. and Baback M. 2011. "Face Recognition in Subspaces". Handbook of Face Recognition, pp. 19-49.

Heseltine T. D. 2005. "Face Recognition: Two-Dimensional and Three-Dimensional Techniques". PhD thesis, The University of York, Department of Computer Science.

Jain A. K., A. Ross, and S. Prabhakar. 2004. "An introduction to biometric recognition". IEEE Transactions on circuits and systems for video technology, vol. 14, no. 1, pp. 4–20.

Kirby M and L. Sirovich. 1990. "Application of the Karhunen–Loeve procedure for the characterization of human faces". IEEE Trans Pattern Analysis and Machine Intelligence, 12:103 108.

Mertins A. 1999. "Signal Analysis: Wavelets, Filter banks, Time-Frequency Transforms and Applications", John Wiley.

Turk M. and A. Pentland. 1991. "Eigenfaces for Recognition". Journal of Cognitive Neuroscience, Vol. 3, No. 1, pp. 71-86.

Wiskott L., J. Fellous, N. Kruger and C. Von Malsburg. 1997. "Face Recognition by Elastic Bunch Graph Matching". IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 19, No. 7, pp. 775-779.

Zhao W., R. Chellappa, A. Rosenfeld and P.J. Phillips. 2003. "Face Recognition: A Literature Survey". ACM Computing Surveys, pp. 399-458.

W. Zhao and R. Chellappa. 2002. "Image-based Face Recognition: Issues and Methods". Image Recognition and Classification, Ed. B. Javidi, M. Dekker, pp. 375-402.

WEB REFERENCES

AT&T Laboratories Cambridge. "Database of Faces, (formerly 'The ORL Database of Faces')". http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html.

AUTHOR LISTING

AUTHOR LISTING