IJACSA

WHERE WISDOM SHARES

International Journal of Advanced Computer Science and Applications

Special Issue

Special Issue on
Extended Papers from Science and Information Conference 2013

SAI

www.ijacsa.thesai.org

IJACSA

WHERE WISDOM SHARES

INTERNATIONAL JOURNAL OF
ADVANCED COMPUTER SCIENCE AND APPLICATIONS

Science and Information Conference 2014

Conference Overview

Science and Information (SAI) Conference is a premier venue for researchers and industry practitioners to share their new ideas, original research results and practical development experiences from Computer Science, Electronics and Communication related areas.

Science and Information Conference 2014 to be held from August 27-29, 2014, will feature presentations of contributed papers and posters, project demonstrations as well as tutorial sessions and super charging Plenary Talks.

The conference is Technically Co-Sponsored by
> IEEE Computer Chapter, UKRI Section
> IEEE Computational Intelligence Society, UKRI Section

Past History

SAI Conference 2013 witnessed 160 researchers, students and scientists from more than 55 countries attending the conference! The conference was structured with paper and poster presentations from the international community of authors, including presentations from keynote speakers and state-of-the-art lectures It was full of engaging speakers and lively discussion. The conference was technically sponsored by 4 units of IEEE and supported by Springer. Conference Proceedings were included in IEEE Xplore.

Important Dates

Early Bird Submission
Submission Deadline : January 01, 2014
Acceptance Notification : February 01, 2014
Registration Deadline : March 01, 2014
Camera Ready Submission : April 15, 2014
Conference Dates : August 27-29, 2014

Regular Submission
Submission Deadline : March 01, 2014
Acceptance Notification : April 01, 2014
Registration Deadline : May 01, 2014
Camera Ready Submission : May 15, 2014
Conference Dates : August 27-29, 2014

All SAI Conference 2014 registered papers will be included in the conference proceedings and submitted to IEEE Xplore and various international databases. Authors of selected outstanding papers will be invited to submit extended versions of their papers for consideration of publication as chapters in Springer Book Series.

Supported By:

Science and Information Conference 2014
August 27-29, 2014 | London, UK

facebook.com/SAIConference

twitter #SAIConference

Web. www.conference.thesai.org
E-Mail. conference@thesai.org

Science and Information Conference 2014

Authors are kindly invited to submit their formatted full papers including results, tables, figures and references. All paper submissions will be blind peer reviewed and evaluated based on originality, research content, correctness, relevance to conference and readability.

Submit your Paper/Poster at http://thesai.org/SAIConference2014/Submit
Submission of Manuscripts can be also done by email at conference@thesai.org

Conference Topics

Technology Trends
  Cloud Computing
  Distributed Systems
  High Performance Computing
  Human Computer Interaction
  Ambient Intelligence
  Big Data
  Social Computing
  Software Engineering and Quality

Communication
  Wireless/ Mobile Applications
  Quantum Computing
  Satellite Communication Systems
  3G/4G Network Evolutions
  Mobile Adhoc Networks
  Open Spectrum Solutions
  Communication Protocols
  Cognitive Radio

Intelligent Systems
  Artificial Intelligence
  Neural Networks
  Fuzzy Logic
  Health care Systems
  Agents and Multi-agent Systems
  Sentiment Analysis
  Data Mining
  Robotics

e-Learning
  Collaborative Learning
  Curriculum Content Design
  Delivery Systems and Environments
  Educational Systems Design
  e-Learning Organisational Issues
  Virtual Learning Environments
  Web-based Learning Communities
  e-Learning Tools

Electronics
  Analog Circuits and Signal Processing
  Design Automation
  Computer Aided Network Design
  Assembly and Packaging
  Digital Circuits
  Sensing and Sensor Networks
  E-Waste
  Systems Architectures

Security
  Biometrics
  Internet Security
  Electronic Data Interchange (EDI)
  Web Services and Performance
  Secure Transactions
  Cryptography
  Secure Protocols
  Cyber Security

Machine Vision
  Augmented Reality
  Virtual Reality
  Image Processing
  Unmanned Systems
  Distributed Interactive Simulation
  Video Analysis
  Medical Diagnosis
  Segmentation Techniques

e-Business
  Enterprise Resource Planning
  Business Intelligence
  Digital Goods and Services
  e-Commerce Application Fields
  e-Commerce Economics
  e-Commerce Services
  Digital Marketing

facebook.com/SAIConference

twitter #SAIConference

Web. www.conference.thesai.org
E-Mail. conference@thesai.org

Science and Information Conference 2014
August 27-29, 2014 | London, UK

Editorial Board

Reviewer Board Members

- **Constantin Filote**
  Stefan cel Mare University of Suceava
- **Constantin Popescu**
  Department of Mathematics and Computer Science, University of Oradea
- **Chandrashekhar Meshram**
  Chhattisgarh Swami Vivekananda Technical University
- **Chao Wang**
- **Chi-Hua Chen**
  National Chiao-Tung University
- **Ciprian Dobre**
  University Politehnica of Bucharest
- **Chien-Pheg Ho**
  Information and Communications Research Laboratories, Industrial Technology Research Institute of Taiwan
- **Prof. D. S. R. Murthy**
  Sreeneedhi
- **Dana PETCU**
  West University of Timisoara
- **Duck Hee Lee**
  Medical Engineering R&D Center/Asan Institute for Life Sciences/Asan Medical Center
- **Deepak Garg**
  Thapar University.
- **Dong-Han Ham**
  Chonnam National University
- **Dr. Gunaseelan Devraj**
  Jazan University, Kingdom of Saudi Arabia
- **Dr. Bright Keswani**
  Associate Professor and Head, Department of Computer Applications, Suresh Gyan Vihar University, Jaipur (Rajasthan) INDIA
- **Dr. S Kumar**
  Anna University
- **Dragana Becejski-Vujaklija**
  University of Belgrade, Faculty of organizational sciences
- **Driss EL OUADGHIRI**
- **Dr. Omaima Al-Allaf**
  Asesstant Professor
- **Elena Camossi**
  Joint Research Centre
- **Eui Lee**
- **Firkhan Ali Hamid Ali**
  UTHM
- **Fokrul Alom Mazarbhuiya**
  King Khalid University
- **Frank Ibikunle**
  Covenant University

- **Fu-Chien Kao**
  Da-Y eh University
- **G. Sreedhar**
  Rashtriya Sanskrit University
- **Ganesh Sahoo**
  RMRIMS
- **Gaurav Kumar**
  Manav Bharti University, Solan Himachal Pradesh
- **Ghalem Belalem**
  University of Oran (Es Senia)
- **Gufran Ahmad Ansari**
  Qassim University
- **Giri Babu**
  Indian Space Research Organisation
- **Giacomo Veneri**
  University of Siena
- **Gerard Dumancas**
  Oklahoma Medical Research Foundation
- **Georgios Galatas**
- **George Mastorakis**
  Technological Educational Institute of Crete
- **Gavril Grebenisan**
  University of Oradea
- **Hadj Hamma Tadjine**
  IAV GmbH
- **Hanumanthappaj**
  UNIVERSITY OF MYSORE
- **Hamid Alinejad-Rokny**
  University of Newcastle
- **Harco Leslie Hendric Spits Warnars**
  Budi LUhur University
- **Hardeep**
  Ferozaepur College of Engineering & Technology, India
- **Hamez I. El Shekh Ahmed**
  Pure mathematics
- **Hesham Ibrahim**
  Chemical Engineering Department, Faculty of Engineering, Al-Mergheb University
- **Dr. Himanshu Aggarwal**
  Punjabi University, India
- **Huda K. AL-Jobori**
  Ahlia University
- **Iwan Setyawan**
  Satya Wacana Christian University
- **Dr. Jamaiah Haji Yahaya**
  Northern University of Malaysia (UUM), Malaysia
- **Jasvir Singh**
  Communication Signal Processing Research Lab

- **James Coleman**
  Edge Hill University
- **Jim Wang**
  The State University of New York at Buffalo, Buffalo, NY
- **John Salin**
  George Washington University
- **Jyoti Chaudary**
  high performance computing research lab
- **Jatinderkumar R. Saini**
  S.P.College of Engineering, Gujarat
- **K Ramani**
  K.S.Rangasamy College of Technology, Tiruchengode
- **K V.L.N.Acharyulu**
  Bapatla Engineering college
- **Kanak Saxena**
  S.A.TECHNOLOGICAL INSTITUTE
- **Ka Lok Man**
  Xi'an Jiaotong-Liverpool University (XJTLU)
- **Kushal Doshi**
  IEEE Gujarat Section
- **Kashif Nisar**
  Universiti Utara Malaysia
- **Kavya Naveen**
- **Kayhan Zrar Ghafoor**
  University Technology Malaysia
- **Kitimaporn Choochote**
  Prince of Songkla University, Phuket Campus
- **Kohei Arai**
  Saga University
- **Kunal Patel**
  Ingenuity Systems, USA
- **Krasimir Yordzhev**
  South-West University, Faculty of Mathematics and Natural Sciences, Blagoevgrad, Bulgaria
- **Labib Francis Gergis**
  Misr Academy for Engineering and Technology
- **Lai Khin Wee**
  Biomedical Engineering Department, University Malaya
- **Latha Parthiban**
  SSN College of Engineering, Kalavakkam
- **Lazar Stosic**
  Collegefor professional studies educators Aleksinac, Serbia
- **Lijian Sun**
  Chinese Academy of Surveying and Mapping, China
- **Leandors Maglaras**
- **Leon Abdillah**
  Bina Darma University

- **Ljubomir Jerinic**
  University of Novi Sad, Faculty of Sciences, Department of Mathematics and Computer Science
- **Lokesh Sharma**
  Indian Council of Medical Research
- **Long Chen**
  Qualcomm Incorporated
- **M. Reza Mashinchi**
- **M. Tariq Banday**
  University of Kashmir
- **Mazin Al-Hakeem**
  Research and Development Directorate - Iraqi Ministry of Higher Education and Research
- **Md Rana**
  University of Sydney
- **Miriampally Venkata Raghavendera**
  Adama Science & Technology University, Ethiopia
- **Mirjana Popvic**
  School of Electrical Engineering, Belgrade University
- **Manas deep**
  Masters in Cyber Law & Information Security
- **Manpreet Singh Manna**
  SLIET University, Govt. of India
- **Manuj Darbari**
  BBD University
- **Md. Zia Ur Rahman**
  Narasaraopeta Engg. College, Narasaraopeta
- **Messaouda AZZOUZI**
  Ziane AChour University of Djelfa
- **Dr. Michael Watts**
  University of Adelaide
- **Milena Bogdanovic**
  University of Nis, Teacher Training Faculty in Vranje
- **Miroslav Baca**
  University of Zagreb, Faculty of organization and informatics / Center for biomet
- **Mohamed Ali Mahjoub**
  Preparatory Institute of Engineer of Monastir
- **Mohamed El-Sayed**
  Faculty of Science, Fayoum University, Egypt.
- **Mohammad Yamin**
- **Mohammad Ali Badamchizadeh**
  University of Tabriz
- **Mohamed Najeh Lakhoua**
  ESTI, University of Carthage

(iv)

- **Mohammad Alomari**
  Applied Science University
- **Mohammad Kaiser**
  Institute of Information Technology
- **Mohammed Al-Shabi**
  Assisstant Prof.
- **Mohammed Sadgal**
- **Mourad Amad**
  Laboratory LAMOS, Bejaia University
- **Mohammed Ali Hussain**
  Sri Sai Madhavi Institute of Science & Technology
- **Mohd Helmy Abd Wahab**
  Universiti Tun Hussein Onn Malaysia
- **Monji Kherallah**
  University of Sfax
- **Mostafa Ezziyyani**
  FSTT
- **Mueen Uddin**
  Universiti Teknologi Malaysia UTM
- **Mona Elshinawy**
  Howard University
- **N Ch.Sriman Narayana Iyengar**
  VIT University
- **Natarajan Subramanyam**
  PES Institute of Technology
- **Neeraj Bhargava**
  MDS University
- **Noura Aknin**
  University Abdelamlek Essaadi
- **Nidhi Arora**
  M.C.A. Institute, Ganpat University
- **Nazeeruddin Mohammad**
  Prince Mohammad Bin Fahd University
- **Najib Kofahi**
  Yarmouk University
- **Na Na**
  NA
- **Om Sangwan**
- **Oliviu Matel**
  Technical University of Cluj-Napoca
- **Osama Omer**
  Aswan University
- **Ousmane Thiare**
  Associate Professor University Gaston Berger of Saint-Louis SENEGAL
- **Pankaj Gupta**
  Microsoft Corporation
- **Paresh V Virparia**
  Sardar Patel University
- **Dr. Poonam Garg**

Institute of Management Technology, Ghaziabad
- **Prabhat K Mahanti**
  UNIVERSITY OF NEW BRUNSWICK
- **Qufeng Qiao**
  University of Virginia
- **Rachid Saadane**
  EE departement EHTP
- **Raghuraj Singh**
- **Raj Gaurang Tiwari**
  AZAD Institute of Engineering and Technology
- **Rajesh Kumar**
  National University of Singapore
- **Rakesh Balabantaray**
  IIIT Bhubaneswar
- **RashadAl-Jawfi**
  Ibb university
- **Rashid Sheikh**
  Shri Venkteshwar Institute of Technology , Indore
- **Ravi Prakash**
  University of Mumbai
- **Rawya Rizk**
  Port Said University
- **Reshmy Krishnan**
  Muscat College affiliated to stirling University.U
- **Ricardo Vardasca**
  Faculty of Engineering of University of Porto
- **Ritaban Dutta**
  ISSL, CSIRO, Tasmaniia, Australia
- **Rowayda Sadek**
- **Ruchika Malhotra**
  Delhi Technoogical University
- **Saadi Slami**
  University of Djelfa
- **Sachin Kumar Agrawal**
  University of Limerick
- **Dr.Sagarmay Deb**
  University Lecturer, Central Queensland University, Australia
- **Said Ghoniemy**
  Taif University
- **Samarjeet Borah**
  Dept. of CSE, Sikkim Manipal University
  University College of Applied Sciences UCAS-Palestine
- **Santosh Kumar**
  Graphic Era University, India
- **Sasan Adibi**
  Research In Motion (RIM)
- **Saurabh Pal**
  VBS Purvanchal University, Jaunpur

- **Saurabh Dutta**
  Dr. B. C. Roy Engineering College, Durgapur
- **Sebastian Marius Rosu**
  Special Telecommunications Service
- **Selem charfi**
  University of Valenciennes and Hainaut Cambresis, France.
- **Sengottuvelan P**
  Anna University, Chennai
- **Senol Piskin**
  Istanbul Technical University, Informatics Institute
- **Seyed Hamidreza Mohades Kasaei**
  University of Isfahan
- **Shafiqul Abidin**
  G GS I P University
- **Shahanawaj Ahamad**
  The University of Al-Kharj
- **Shawkl Al-Dubaee**
  Assistant Professor
- **Shriram Vasudevan**
  Amrita University
- **Sherif Hussain**
  Mansoura University
- **Siddhartha Jonnalagadda**
  Mayo Clinic
- **Sivakumar Poruran**
  SKP ENGINEERING COLLEGE
- **Shikha Bagui**
  University of West Florida
- **Sim-Hui Tee**
  Multimedia University
- **Simon Ewedafe**
  Baze University
- **SUKUMAR SENTHILKUMAR**
  Universiti Sains Malaysia
- **Slim Ben Saoud**
- **Sumit Goyal**
- **Sumazly Sulaiman**
  Institute of Space Science (ANGKASA), Universiti Kebangsaan Malaysia
- **Sohail Jabb**
  Bahria University
- **Suhas Manangi**
  Microsoft
- **Suresh Sankaranarayanan**
  Institut Teknologi Brunei
- **Susarla Sastry**
  J.N.T.U., Kakinada
- **Syed Ali**
  SMI University Karachi Pakistan
- **T C. Manjunath**

HKBK College of Engg
- **T V Narayana Rao**
  Hyderabad Institute of Technology and Management
- **T. V. Prasad**
  Lingaya's University
- **Taiwo Ayodele**
  Infonetmedia/University of Portsmouth
- **Tarek Gharib**
- **THABET SLIMANI**
  College of Computer Science and Information Technology
- **Totok R. Biyanto**
  Engineering Physics, ITS Surabaya
- **TOUATI YOUCEF**
  Computer sce Lab LIASD - University of Paris 8
- **Venkatesh Jaganathan**
  ANNA UNIVERSITY
- **VIJAY H MANKAR**
- **VINAYAK BAIRAGI**
  Sinhgad Academy of engineering, Pune
- **Vishal Bhatnagar**
  AIACT&R, Govt. of NCT of Delhi
- **VISHNU MISHRA**
  SVNIT, Surat
- **Vitus S.W. Lam**
  The University of Hong Kong
- **Vuda Sreenivasarao**
  St. Mary's College of Engineering & Technology
- **Wei Wei**
- **Wichian Sittiprapaporn**
  Mahasarakham University
- **Xiaojing Xiang**
  AT&T Labs
- **Y Srinivas**
  GITAM University
- **YASSER ATTIA ALBAGORY**
  College of Computers and Information Technology, Taif University, Saudi Arabia
- **YI FEI WANG**
  The University of British Columbia
- **Yilun Shang**
  University of Texas at San Antonio
- **YU QI**
  Mesh Capital LLC
- **ZAIRI ISMAEL RIZMAN**
  UiTM (Terengganu) Dungun Campus
- **ZENZO POLITE NCUBE**
  North West University
- **ZHAO ZHANG**
  Deptment of EE, City University of Hong Kong

CONTENTS

Approximate SER for M-PSK using MRC and STTD Techniques over Fading Channels

Mahmoud A. Khodeir

Department of Electrical Engineering
Jordan University of Science and Technology
Irbid, Jordan. P.O. Box 3030

Muteeah A. Jawarneh

Department of Electrical Engineering
Jordan University of Science and Technology
Irbid, Jordan. P.O. Box 3030

Abstract—**In this study, approximate *symbol error rate* (SER) expressions for *M-ary phase shift keying* (M-PSK) modulation scheme over *independent and identically distributed* (i.i.d) slow-flat Rician and Rayleigh fading channels are derived. Simulation results show the superior impact of using the *maximum ratio combining* (MRC) space diversity technique on the overall performance. In particular, the communication reliability (i.e., capacity and coverage) will increase by increasing the diversity order (i.e., the number of the combiner's branches), where less power is needed to achieve the same probability of error. Then, a comparison between the approximate and exact probability of symbol error is performed and the results are shown to be comparable (1–2 dB). Next, approximate SER expression is derived over i.i.d slow-flat Nakagami-*m* fading channels. In particular, *space time transmit diversity* (STTD) technique is used to enhance the reliability of the proposed model using two transmit antennas and one receive antenna. The simulation results show the effect of the Nakagami-*m* parameter, *m*, on the SER where the performance will improve by increasing the value of *m* where fading is less severe in this case. Furthermore, the performance of the SER is lower for higher values of SNR and is worse for high order PSK modulation schemes.**

Keywords—*Rician fading channel; Rayleigh fading channel; Nakagami-m fading channel; maximum ratio combining; space diversity; space time transmit diversity; symbol error rate.*

I. INTRODUCTION

In general, the overall performance of any wireless communication system is severely degraded by the time-varying multi-path fading which make the analysis relatively difficult. In particular, fading results from the interference between two or more versions of the transmitted signals that arrive at the receiver with slightly different delay times. Time diversity, frequency diversity, and space diversity are the main diversity techniques that can efficiently be deployed to overcome the destructive effects of fading.

Generally, space diversity deploys multiple antennas either at the transmitting end and/or at the receiving end to get advantage of the diversity gain in order to mitigate the multi-path fading in an effective way. Here, multiple antennas provide the receiver with several observations of the same signal. Hence, each antenna will experience a different interference environment, and therefore, if one antenna is experiencing a deep fading most probable the other antennas have adequate signals. Namely, MRC is a classical combining technique in which the signals from the received antenna elements are weighted such that the signal to noise ratio (SNR)

achieved due to the total sum of the received signals is maximized. Moreover, the distance between the multiple antennas must be selected to be sufficiently far apart to guarantee that signals from different antennas face uncorrelated fading process.

First of all, the *moment generating function* (MGF) of the received SNR was used in [1] to analyze the error rate performance for M-PSK modulation scheme, where independent phase estimation was assumed over Rician, Rayleigh, and Nakagami-*m* fading channels. Next, the authors in [2] have derived the exact *bit error rate* (BER) for a *binary phase shift keying* (BPSK) modulation scheme in a *synchronous cochannel interference* (CCI) under Nakagami-*m* flat fading channel using both of *equal gain combining* (EGC) technique with *characteristic function* (CF) method and *selection combining* (SC) technique with Fourier series method. Furthermore, the authors in [3] studied the exact BER performance for BPSK and *binary frequency shift keying* (BFSK) modulation schemes over MIMO system were derived for i.i.d slow-flat Nakagami-*m* fading channel. Here, the MGF based analytical method was used to analyze this model using both *transmit and receive antenna selection* (TAS/RAS).

Moreover, the outage and BER analysis of the MRC for band-limited BPSK system with micro-cellular system over i.i.d Rayleigh and Nakagami-*m* slow-flat fading channels were presented in [4]. Here, the authors considered an independent fading gain and asynchronous timing as well as co-channel interference in the derivation to obtain the exact closed-form outage probability expression with both of equal and unequal interferer's power. Next, both of the optimum and sub-optimum diversity combining were derived in [5] for differential and coherent signals with M-PSK modulation scheme over Rician fading channels plus the impulsive class-A noise. Here, the tight performance is upper-bounded for both of the combining schemes and the derivations were done for both MRC and EGC diversity techniques. In particular, a relatively large performance gain was shown for both of the *suboptimum noncoherent combining* (SNS) and the *sub-optimum coherent combining* (SCC) schemes.

Next, the approximate BER for *M-ary quadrature shift keying* (M-QAM) have been derived in [6] over i.i.d slow-flat Rician fading channel, where a single exponential function was used to approximate the BER for M-QAM modulation scheme over a Gaussian channel. Furthermore, the average BER for band-limited BPSK over i.i.d Nakagami-*m* slow-flat was derived in [7] where the authors assumed a synchronous

timing and independent fading gain, equal gain, and maximum desired user power selection combining. Moreover, the co-channel interference was considered, and the characteristic function and Fourier series methods for different values of the diversity order and fading parameter were used in the derivation. Therefore, the performance and the superiority were increased when the diversity order was increased. Finally, the authors in [8] derived the BER for *non-coherent fast-frequency-hopped M-ary orthogonal frequency shift keying* (FFH/MFSK) for non-selective Nakagami-*m* slow-flat fading channel, where the CF based approach was used. Furthermore, the authors in [9] derived an exact expression for the BER for M-PSK and M-QAM under Nakagami-*m* fading channels with STTD. This expression was obtained numerically and the results showed a better BER performance when both of the SNR and the parameter *m* were increased.

In this paper, an approximate expression of the SER for M-PSK modulation over a Gaussian channel is used to derive the approximate expressions for the SER for M-PSK over i.i.d slow-flat Rician and Rayleigh fading channels using MRC space diversity technique. In particular, the MRC space diversity is used to enhance the overall coverage and the capacity of the proposed model. Here, the SER for optimum MRC receives diversity with *channel side information known at the receiver* (CSIR) is acquired. Moreover, the proposed approximation will be bounded within 1 dB for signal alphabet $M \geq 4$ and for $0 \leq$ SNR ≤ 30 dB. Next, the approximate expression for the SER for M-PSK over i.i.d slow-flat Nakagami-*m* fading channels is derived using STTD technique to improve the overall performance. In general, STTD technique is applied using two transmit antennas and one receive antenna. Overall, the rest of this paper is organized as follows. In section II, approximate SER expressions for M-PSK over i.i.d Rician and Rayleigh fading channels are derived. Then approximate SER expressions for the M-PSK using MRC diversity are obtained in Section III. Next, in Section IV an approximate SER for M-PSK using STTD over i.i.d Nakagami-*m* fading channels is derived. Finally, Sections V and VI present the simulation results analysis and the conclusion.

II. APPROXIMATE SER FOR M-PSK OVER RICIAN FADING CHANNELS

A communication system with M-PSK signaling over AWGN plus i.i.d slow-flat Rician fading channel is considered. Here, the transmitted signal is perturbed by AWGN with two-sided power spectral density $N_0/2$. Moreover, the effect of the Rician fading channel on the signal is multiplicatively distorted by the Rician parameter *k*. The SER of the proposed model is given by [10]:

$$P_s = 2Q\left(\sqrt{2\gamma_S}\sin\left(\pi/M\right)\right) \qquad (1)$$

where $\gamma_S = E_S/N_0$ represents the SNR per symbol, E_S is the overall symbol energy, *M* is the number of modulation levels, and Q (.) is the *Q*-function which is approximated by [11]:

$$Q(x) \approx 0.5\,exp\left(-x^2/2\right) \qquad (2)$$

Therefore, the probability of symbol error for M-PSK over a Gaussian channel can be rewritten as follows:

$$P_s = exp(-sin^2(\pi/M)\gamma log_2(M)\,) \qquad (3)$$

where $\gamma_s = \gamma_b log_2(M) = \gamma\, log_2(M)$. Here, γ_b represents the SNR per bit. Moreover, the SER expression, which is caused by stationary AWGN terms, depends only on the instantaneous SNR associated with each symbol. Furthermore, the instantaneous SNR per symbol becomes a random variable due to fading effects. Therefore, by using the parameter *k* and applying change of variables, the pdf of the instantaneous SNR, γ, is given by [6]:

$$p_\gamma(\gamma) = (1 + k)\frac{exp\,[-k]}{\overline{\gamma}}exp\left[\frac{-(1+k)\gamma}{\overline{\gamma}}\right]$$
$$\times\, I_o\left(2\sqrt{\frac{k(1+k)\gamma}{\overline{\gamma}}}\right) \qquad (4)$$

where $\overline{\gamma}$ represents $E(\gamma)$. Furthermore, the probability of symbol error, $P_1\,(E)$, can be found by taking the average of the conditional probability of error, $P\,(E/\gamma)$, over the pdf of γ as follows:

$$P_1\,(E) = \int_0^\infty P(E/\gamma)\,p_\gamma(\gamma)\,d\gamma \qquad (5)$$

The above equation can be rewritten after some manipulations as follows [13]:

$$P_1\,(E) \approx \frac{(1+k)\,exp(-k)}{(1+k)+(sin^2(\pi/M)log_2(M)\,\overline{\gamma})}$$
$$\times\, exp\left(\frac{k(1+k)}{(1+k)+(sin^2(\pi/M)\,log_2(M)\overline{\gamma})}\right) \qquad (6)$$

Finally, by substituting $k = 0$ in (8), one gets the approximate SER expression for M-PSK over Rayleigh fading channel.

III. ANALYSIS FOR M-PSK RICIAN LINEAR FADING CHANNEL WITH MRC SPACE DIVERSITY

The instantaneous SNR at the output of the MRC combiner is the sum of the instantaneous SNR at each individual branch (i.e., $\gamma = \sum_{i=1}^N \gamma_i$) where *N* is the number of the combiner's branches. Then, the new pdf of γ after applying diversity combining is given by [6]:

$$p_{\Upsilon(\gamma)} = \left(\frac{N+k}{\overline{\gamma}}\right)\left[\frac{(N+k)\gamma}{k\overline{\gamma}}\right]^{\frac{N-1}{2}}exp\left(\frac{-(N+k)\gamma+k\overline{\gamma}}{\overline{\gamma}}\right)$$
$$\times$$
$$I_{N-1}\left(2\sqrt{\frac{k(N+k)\gamma}{\overline{\gamma}}}\right) \qquad (7)$$

where $k = \sum_{i=1}^N k_i$ and I_N (.) is the *n*-th order modified Bessel function of the first kind. Here, the error probability is a function of *k*, $\overline{\gamma}$, and *N*. Next, by averaging the conditional probability of error over the pdf of γ, one gets [13]:

$$P_2(E) \approx \int_0^\infty \left(\frac{N+k}{\bar{\gamma}}\right)\left[\frac{(N+k)}{k\bar{\gamma}}\right]^{\frac{N-1}{2}} \gamma^{\left(\frac{N-1}{2}\right)} exp\left(\frac{-(N+k)\gamma}{\bar{\gamma}}\right)$$
$$\times exp(-k) I_{N-1}\left(2\sqrt{\frac{k(N+k)\gamma}{\bar{\gamma}}}\right)$$
$$\times exp(-sin^2(\pi/M)\gamma log_2(M)) \, d\gamma \qquad (8)$$

After some manipulations and simplifications, one can get the following expression for the probability of error [13]:

$$P_2(E) \approx \left[\frac{(N+k)+(sin^2(\pi/M)\bar{\gamma}log_2(M))}{\bar{\gamma}}\right]^{-N}$$
$$\times exp\left(\frac{k(N+k)}{(N+K)+sin^2(\pi/M)\bar{\gamma}log_2(M))}\right)$$
$$\times$$
$$exp(-k)\left(\frac{N+k}{\bar{\gamma}}\right)^N \qquad (9)$$

Here, when the value of k is set to zero, one can get the approximate probability of error for M-PSK over i.i.d Rayleigh fading channel using N branches which is given by:

$$P_2(E) = \left[\frac{N+sin^2(\pi/M)\bar{\gamma}log_2(M)}{\bar{\gamma}}\right]^{-N}\left(\frac{N}{\bar{\gamma}}\right)^N \qquad (10)$$

IV. Approximate SER For M-PSK With STTD Over Nakagami-M Fading Channel

In general, STTD is an open loop technique which is adopted by the 3GPP to increases the diversity gain. Considering a downlink wireless channel with two transmit antennas and one receive antenna as shown in Fig. 1 [9]:
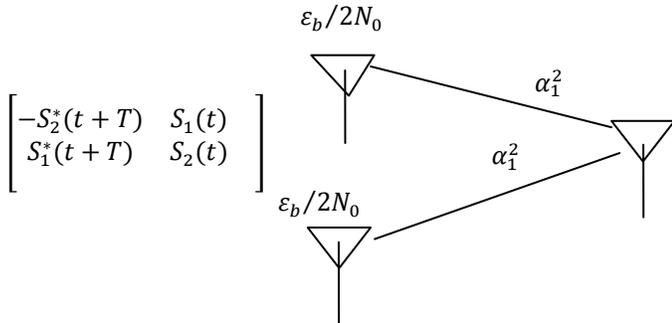
Fig. 1. System model for the basic scheme of STTD

The average transmits SNR per bit for each antenna in STTD technique model is denoted by $\varepsilon_b/2N_O$. This assumption is valid to get the same total radiated power similar to the case with one transmit antenna. Here, $\alpha_1{}^2$ and $\alpha_2{}^2$ denote the independent channel gains (i.e., square of the fading channel factor), $S_1(t)$ and $S_2(t)$ are the symbols transmitted by both antennas at time t, $-S_2{}^*(t+T)$ and $S_1{}^*(t+T)$ at time $t+T$, where * denotes the complex conjugate operation. The Nakagami-m parameter, m, characterizes the severity of the signal fading. In particular, $m = 0.5$ represents the one-sided Gaussian fading channel and $m = 1$ describes the Rayleigh fading channel. Furthermore, the lognormal and Rician distributions can be approximated by the Nakagami-m distribution when $m > 1$. Now, the pdf of the Nakagami-m distribution is given by [9]:

$$p_\alpha(\alpha) = \frac{2}{\Gamma(m)}\left(\frac{m}{\Omega}\right)^m \alpha^{2m-1}e^{-m\alpha^2/\Omega} \qquad (11)$$

where $\Omega = E(\alpha^2)$ represents the average channel gain, $\Gamma(.)$ is the gamma function, $m = \frac{\Omega^2}{E[(R^2-\Omega^2)]}$ and $E(R^n)$ is given by:

$$E(R^n) = \frac{\Gamma\left(m+\frac{1}{2}n\right)}{\Gamma(m)}\left(\frac{\Omega}{m}\right)^{n/2} \quad, m \geq 0.5 \qquad (12)$$

Now, let $\gamma_b = \alpha^2\frac{\varepsilon_b}{N_o}$, then one can write the pdf of the received SNR per bit for i.i.d Nakagami-m flat fading channels as follows:

$$p_{\gamma_b}(\gamma_b) = \frac{m^m}{\bar{\gamma}_b^m\Gamma(m)}\gamma_b^{m-1}e^{-m\gamma_b/\bar{\gamma}_b} \qquad (13)$$

where $\bar{\gamma}_b = E(\alpha^2)\frac{\varepsilon_b}{N_O}$ represents the average received SNR. Now, the channel gain with STTD diversity technique over i.i.d Nakagami-m flat fading channel is given by [9]:

$$\alpha^2 = \frac{1}{2}(\alpha_1^2 + \alpha_2^2) \qquad (14)$$

let

$$y = \alpha^2, \quad x_1 = \alpha_1^2, \quad x_2 = \alpha_2^2 \qquad (15)$$

Therefore, $\frac{\varepsilon_b}{N_0} = \frac{\bar{\gamma}_b}{\Omega}$ and the pdf of y can be rewritten [6]:

$$p_y(y) = 2\sqrt{\pi}\left(\frac{m}{\Omega}\right)^{2m}\frac{y^{2m-1}}{\Gamma(m)\Gamma\left(\frac{1}{2}+m\right)}e^{-2my/\Omega} \qquad (16)$$

Hence, the pdf of the received SNR with STTD technique over i.i.d Nakagami-m flat fading channel is given by [9]:

$$p_\gamma(\gamma) = \frac{1}{\varepsilon_b/N_0}p_\gamma\left(\frac{\gamma_b}{\varepsilon_b/N_0}\right)$$
$$= \frac{2\sqrt{\pi} \, m^{2m} \gamma_b^{2m-1}}{\Gamma(m)\Gamma(1/2+m)\bar{\gamma}_b^{2m}}exp\left(\frac{-2m\gamma_b}{\bar{\gamma}_b}\right) \qquad (17)$$

The pdf of γ at the received SNR per bit for i.i.d Nakagami-m fading channel with STTD technique using two transmit antennas and one receive antenna is given as follows [9]:

$$p_\gamma(\gamma) = \frac{2\sqrt{\pi} \, m^{2m} \gamma^{2m-1}}{\Gamma(m)\Gamma(1/2+m)\bar{\gamma}^{2m}} \times exp\left(\frac{-2m\gamma}{\bar{\gamma}}\right) \qquad (18)$$

Furthermore, by using the approximation of the probability of symbol error for M-PSK over a Gaussian channel which is shown in (3), one can find the probability of symbol error, $P_3(E)$, using (6) as follows:

$$P_3(E) \approx \int_0^\infty exp(-sin^2(\pi/M)\gamma log_2(M))$$
$$\times exp\left(\frac{-2m\gamma}{\bar{\gamma}}\right)\frac{2\sqrt{\pi} \, m^{2m} \gamma^{2m-1}}{\Gamma(m)\Gamma(1/2+m)\bar{\gamma}^{2m}} \, d\gamma \qquad (19)$$

which can be written as follows?

$$P_3\,(E) \approx \frac{2\sqrt{\pi}\; m^{2m}}{\Gamma(m)\Gamma(1/2+m)\;\bar{\gamma}^{2m}} \int_0^\infty \gamma^{2m-1}$$

$$\times exp\left[-(sin^2(\pi/M)\gamma log_2(M)) + \frac{-2m\gamma}{\bar{\gamma}}\right]d\gamma \qquad (20)$$

Now, to simplify the calculations let:

$$C = \frac{2\sqrt{\pi}\; m^{2m}}{\Gamma(m)\Gamma(1/2+m)\;\bar{\gamma}^{2m}}$$

Then (20) can be rewritten as follows:

$$P_3\,(E) =$$
$$C \int_0^\infty exp\left[\frac{-(sin^2(\pi/M)\gamma log_2(M))\bar{\gamma} - 2m\gamma}{\bar{\gamma}}\right]\gamma^{2m-1}\;d\gamma \qquad (21)$$

Next, to solve (21), one can use the following integral [12]:

$$\int_0^\infty x^n\; e^{-xa}dx = \begin{bmatrix} \frac{\Gamma(n+1)}{a^{n+1}} & n > -1,\; a > 0 \\[2mm] \frac{n!}{a^{n+1}} & a > 0,\; n: positive \end{bmatrix} \qquad (22)$$

Now let

$$n = 2m - 1\;,$$

$$a = \frac{sin^2(\pi/M)log_2(M)\bar{\gamma} + 2m}{\bar{\gamma}}$$

Then (21) can be written for the case when $n > -1$, $a > 0$, as follows:

$$P_3\,(E) \approx C_2 \times \frac{\Gamma(n+1)}{a^{n+1}} \approx C_2 \times \frac{\Gamma(2m-1+1)}{\left[\frac{sin^2(\pi/M)log_2(M)\bar{\gamma}+2m}{\bar{\gamma}}\right]^{2m}}$$

$$\approx \frac{2\sqrt{\pi}\; m^{2m}}{\Gamma(m)\Gamma(1/2+m)\;\bar{\gamma}^{2m}} \times \frac{\Gamma(2m)}{\left[\frac{sin^2(\pi/M)log_2(M)\bar{\gamma}+2m}{\bar{\gamma}}\right]^{2m}} \qquad (23)$$

Next, the above equation can be rewritten as follows:

$$\frac{P_3\,(E) \approx}{\frac{2\sqrt{\pi}\;\Gamma(2m)}{\Gamma(m)\Gamma(1/2+m)}} \times \frac{m^{2m}}{[sin^2(\pi/M)log_2(M)\bar{\gamma}+2m]^{2m}} \qquad (24)$$

Furthermore, when $a > 0$ and n is positive, one can write (21) as follows:

$$\frac{P_3\,(E) \approx}{\frac{2\sqrt{\pi}\; m^{2m}}{\Gamma(m)\Gamma(1/2+m)}}\frac{(2m-1)!}{[sin^2(\pi/M)log_2(M)\bar{\gamma}+2m]^{2m}} \qquad (25)$$

V. NUMERICAL RESULTS

Here, the simulation results for the approximate and exact SER for Rician, Rayleigh, and Nakagami-*m* channels with MRC and STTD diversity techniques are presented. First of all, Fig. 2 shows the approximate SER over Rician fading channel with MRC diversity technique for $M = 8$ and 16. Here, when $N = 2$ and for different values of the Rician parameter, k, one can see the direct impact of increasing k on improving the overall performance. However, the worst case (i.e., the Rayleigh fading channel case) is achieved when $k = 0$. Furthermore, when $N = 4$, one can see the great reduction in the average SNR as shown in Fig. 3. The same approach is

applied in Figs 4 and 5 for $M = 16$ and 32. The overall performance is improved by applying diversity. However, when M increases the probability of error increases (i.e., the receiver needs more power to achieve the same probability of error). Next, Figs. 6, 7 and 8 show comparisons between the approximate and the exact SER values over the Rayleigh fading channels using MRC technique for $M = 8$, 16 and 64, respectively. Here, one can observe the overall significant improvement of the SER in the proposed model by using MRC diversity. In particular, when the diversity order increases the overall performance is improved.

Now, in Fig. 9 shows the approximate probabilities of error over Rayleigh fading channel with MRC diversity technique for $M = 8$, 16, and 64 when $N = 1$, 4, and 10. Here, one can note that the overall performance when $M = 8$ or 16 is better than the case when $M = 64$. Next, in Fig. 10, the approximate SER for *M*-PSK with STTD is shown over Nakagami-*m* fading channels when $M = 4$ for different values of the parameter m. The direct impact of increasing m on improving the overall performance is obvious (i.e., for $m = 3$, one can see the great reduction in the average SNR). However, one can get the worst case scenario when $m < 1$. The same approach is applied in Figs. 11 and 12 for $M = 16$ and 32. Again, the probability of error is increased for the higher order of M.

Furthermore, Fig. 13 show comparison between the approximate SERs for M-PSK modulation scheme under Nakagami-*m* fading channel with STTD diversity technique for different values of M. In particular, the comparison is done for $M = 4$, 16, and 32 when $m = 1$ and $m = 6$. From the simulation results, one can observe again the effect of increasing the parameter m on the SER. The overall performance of the model is great when $m = 6$ as compared with the Rayleigh fading channel case (i.e., when $m = 1$). Finally, Fig. 17 shows comparisons between the approximate SER for different values of M (e.g., $M = 4$, 8, and 16) when $m = 6$, 3, and 0.5, respectively. Again, one can note the overall performance under Nakagami-*m* fading channel when $m = 6$ or 3 outperforms the case when $m = 0.5$.
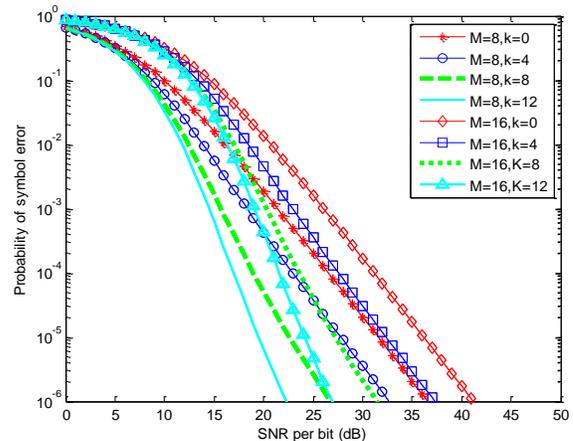
Fig. 2. Approximate SER for M-PSK over Rician fading channels for $N = 2$ and for various values of k when $M = 8$, 16.
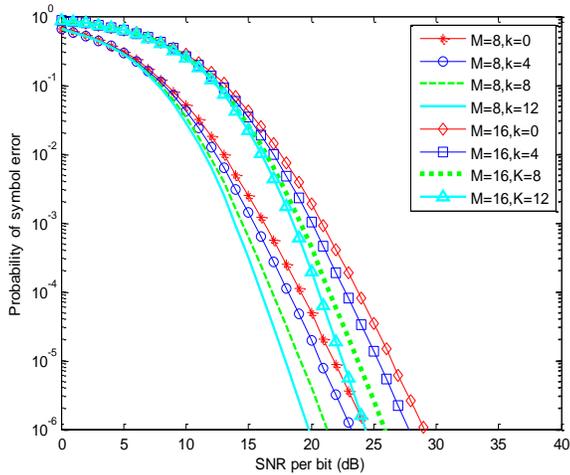
Fig. 3.   Approximate SER for M-PSK over Rician fading channels for $N = 4$ and for various values of k when $M = 8, 16$.
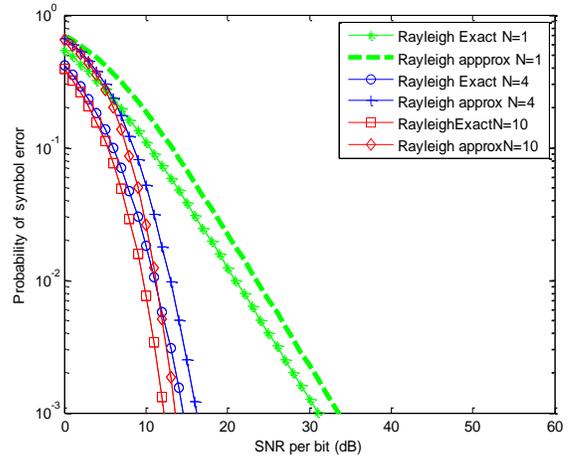
Fig. 6.   Exact and approximate SER for M-PSK with space diversity over Rayleigh fading channel for $M = 8$, $N = 1$, $N = 4$ and $N = 10$.
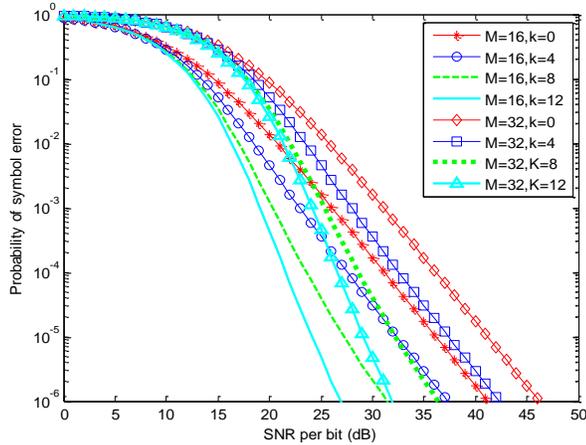
Fig. 4.   Approximate SER for M-PSK over Rician fading channels for $N = 2$ and for various values of k when $M = 16, 32$.
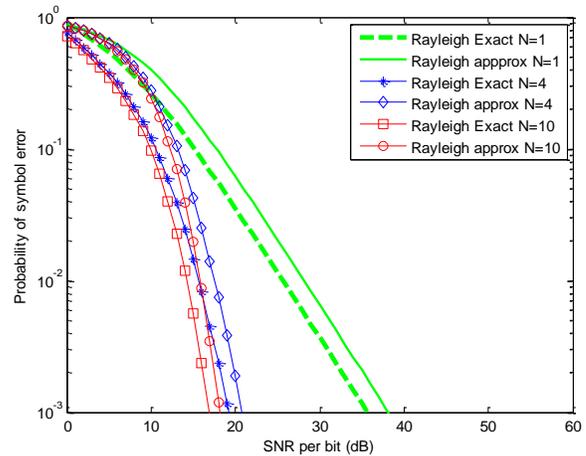
Fig. 7.   Exact and approximate SER for M-PSK with space diversity over Rayleigh fading channels for $M = 16$ and $N = 1$, $N = 4$ and $N = 10$.
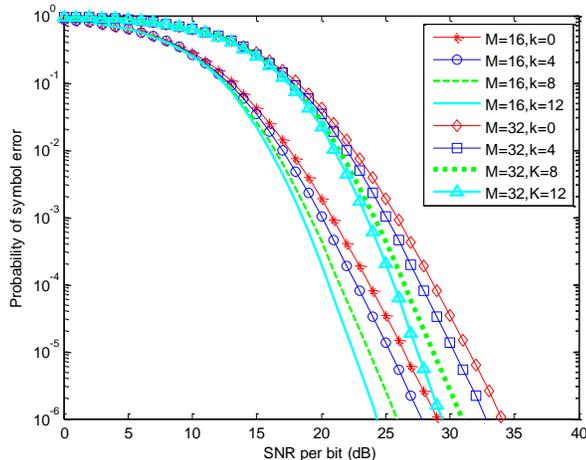
Fig. 5.   Approximate SER for M-PSK over Rician fading channels for $N = 4$ and for various values of k when $M = 16, 32$.
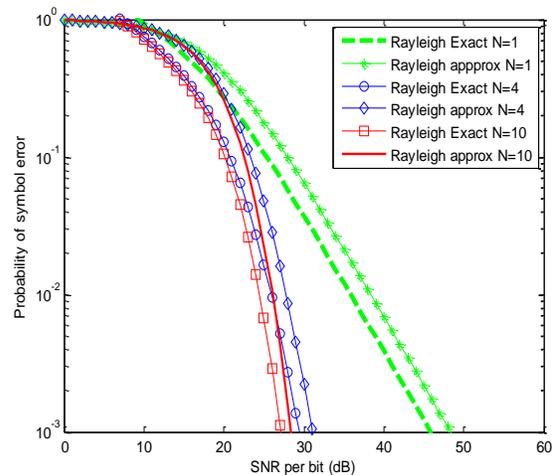
Fig. 8.   Exact and approximate SER for M-PSK with space diversity over Rayleigh fading channels for $M = 64$ and $N = 1$, $N = 4$ and $N = 10$.
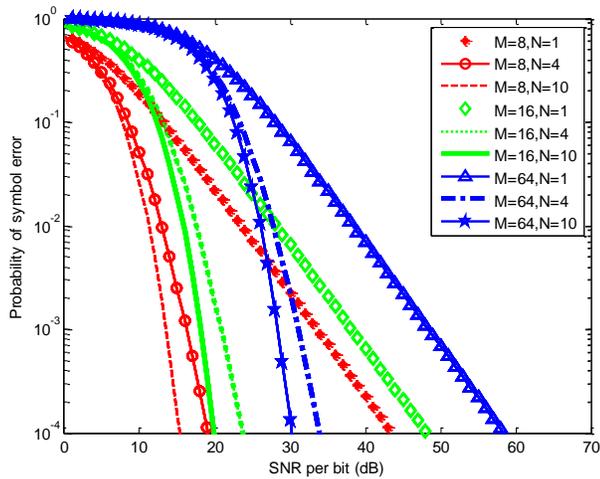
Fig. 9. Approximate SER for M-PSK with space diversity over Rayleigh fading channels for $M = 8$, $M = 16$ and $M = 64$, $N = 1$, $N = 4$ and $N = 10$.
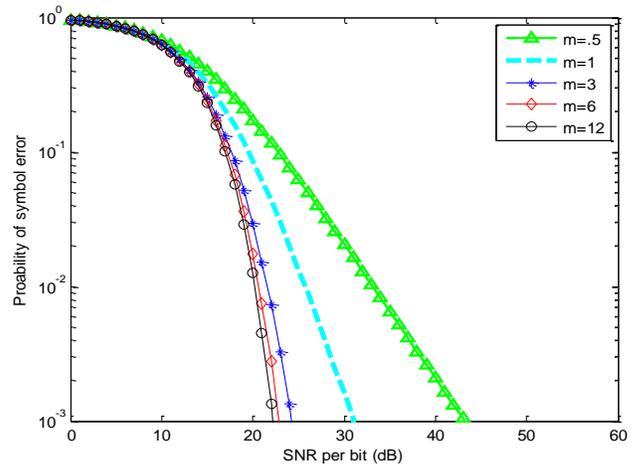
Fig. 12. Approximate symbol error rate for M-PSK with STTD over Nakagami-m fading channels when $M = 32$ for different values of m.
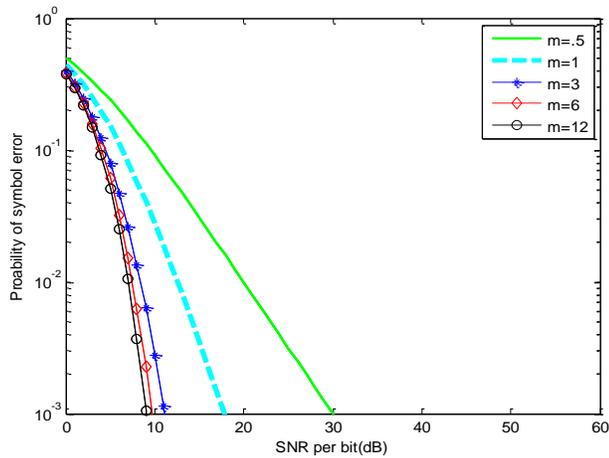
Fig. 10. Approximate symbol error rate for M-PSK with STTD over Nakagami-m fading channels when $M = 4$ for different values of m.
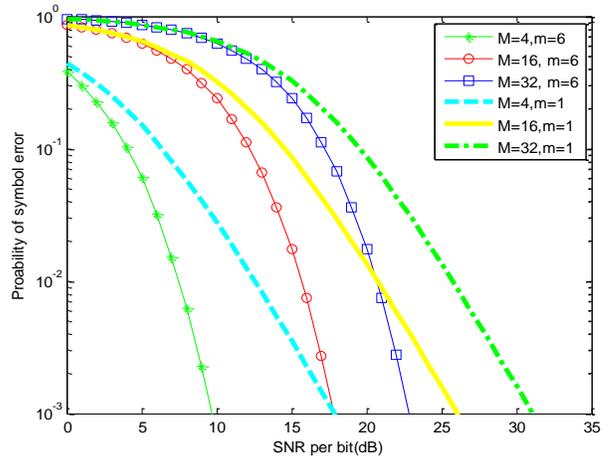
Fig. 13. Approximate symbol error rate for M-PSK with STTD over Nakagami-m fading channels when $m = 1$, $m = 6$ for different values of M.
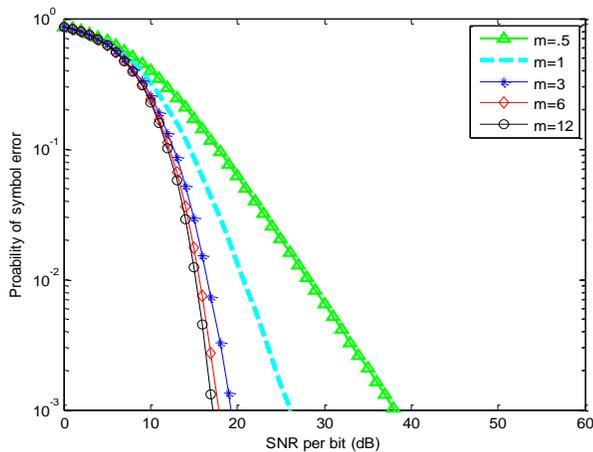
Fig. 11. Approximate symbol error rate for M-PSK with STTD over Nakagami-m fading channels when $M = 16$ for different values of m.
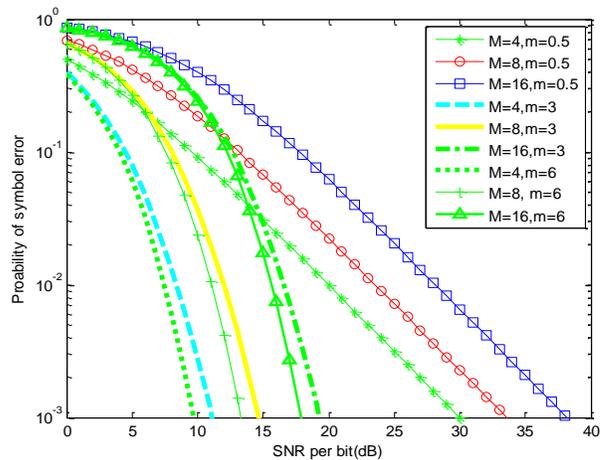
Fig. 14. Approximate symbol error rate for M-PSK with STTD over Nakagami-m fading channels when $m = 0.5$, 3 and 6 for different values of M.

VI. Conclusion

In this study, approximate expressions of the SER for M-PSK modulation scheme over i.i.d slow-flat Rician and Rayleigh fading channels using MRC space diversity are derived. In particular, an approximate expression of the SER for M-PSK modulation scheme over a Gaussian channel is used. From the simulation results, one can notice the superior effect of using MRC diversity technique on the overall performance. Moreover, the probability of symbol error with M-PSK modulation scheme over i.i.d slow-flat Rayleigh fading channels for various values of the diversity order N are illustrated. The simulation results show that by increasing the diversity order, the communication reliability will increase. (i.e., as N increases, less power is needed to achieve the same probability of error). Furthermore, comparisons between the exact and approximate SER expressions show that the results are comparable (2% dB). Here, by increasing the Rician parameter, k, the overall performance will be improved. However, by decreasing k, the worst case scenario can be achieved, i.e., the Rayleigh case when $k = 0$. Finally, an approximate expression for the SER for M-PSK over Nakagami-m i.i.d slow-flat fading channel is derived. Moreover, STTD diversity technique is used to enhance the coverage and capacity, where the STTD is applied using two transmit antennas and one receive antenna. From the simulation results, one can observe the effect of increasing the Nakagami-m parameter, m, on the overall performance of the model where the fading is less severe in this case. In particular, the overall performance is shown to be the worst when $m = 0.5$, and one gets the Rayleigh fading channel case. Furthermore, the SER is lower for higher values of SNR and the performance of the SER is worse for high order PSK modulation scheme.

References

[1] Aniruddha. Chandra, Ananya. Patra, and Chayanika. Bose, "Effect of Imperfect Phase Synchronization on the Error Rate Performance of MPSK in Rayleigh, Rician and Nakagami Fading Channel ", Annual IEEE, India conference (INDICON) , 2010.

[2] K. Sivanesan and N. C. Beaulieu, "Exact BER Analysis of Bandlimited BPSK With EGC and SC Diversity in Cochannel Interference and Nakagami Fading", IEEE Commun. Letters, Vol. 8, NO. 10, pp. 623 – 625, Oct., 2004.

[3] Ahmet. F, Coşkun, and Oğuz. Kucur, "Performance of Joint Transmit and Receive Antenna Selection in Nakagami-m Fading Channels", European wireless conference (EW), pp. 314 – 317, 12-15 April, 2010.

[4] K. Sivanesan and Norman C. Beaulieu, "Precise Performance Analysis of MRC Diversity in Micro-Cellular System with Cochannel Interference", Signal System and Computers, Conference Record of the Thirty-Eighth Asilomar Conference (VTC).

[5] R. Schober, Y. Ma, L. Lampe, and P. T. Mathiopoulos," Diversity Combining for Coherent and Differential M-PSK in Fading and Class-A Impulsive Noise ", IEEE Transaction on Wireless Communication, Vol. 4, No. 4, pp. 1425 – 1432, JULY, 2005.

[6] Adel. Ahmad. Ali and Khalid. ALkhudairi, "BER for M-QAM with space diversity and MRC in Rician fading channel", Mobile Computing and Wireless communication International conference (MCWC)., 17-20 Sept. 2006.

[7] K. Sivanesan, Norman. Beaulieu, "Precise Bit Error Rate Analysis of Bandlimited BPSK with EGC and SC Diversity in CCI and Nakagami Fading ",Vehicular Technology conference (VTC), 26-29 September, 2004.

[8] Jian-yun. Zheng, Yi-Sheng. Zhu and Li-ping. Zhu, "Symbol Error Probability Analysis of An FFH/MFSK Receiver with Noise Normalization Combining In a Fading Channel With Partial –Band Jamming", Signal Processing System (SIPS), IEEE Work shop, pp. 377 – 381, 17-19 Oct, 2007.

[9] Ronng. He and Junsong. Xie, "BER Performance of M-QAM and M-PSK in Nakagami Fading Channel with STTD", IEEE International Symposium on Personal, ln-door and Mobile Radio Communication Proceedings, vol. 2, pp. 1327 – 1330, 2003.

[10] Andrea. Goldsmith, " Wireless Communication", Cambridge University, 2005.

[11] Upamanyu. Madhow, "Fundamentals of Digital Communication", Cambridge University, 2008.

[12] S. Gradshteyn and I. M. Ryzhik, "Table of Integrals, Series and Products", San Diego, Fifth edition, CA: Academic, 1994.

[13] M. Khodeir and M. Jawarneh "Approximate Symbol Error Rate for M-ary Phase Shift Keying (M-PSK) using Maximum Ratio Combining (MRC) Technique over Fading Channels", Science and Information (SAI) Conference 2013, London, UK, October 2013

A Socially Driven, Goal-Oriented Approach to Business Process Management

Mohammad Ehson Rangiha

Centre for HCID, City University London,
UK

Bill Karakostas

Centre for HCID, City University London,
UK

Abstract—**Over the recent years, there has been much discussion about the concept of Social Business Process Management (SBPM) and how it is able to overcome some of the limitations of the traditional BPM systems. This paper aims to address gaps in social BPM research by working towards a goal-driven SBPM meta-model that seamlessly integrate the process design and enactment stages. This approach also makes use of a process recommendation system to guide the activities of the users based on their social behavior and social goals. We argue that this approach will lead to truly social driven process enactment environments.**

Keywords—*BPM; Social BPM; Goal-Based Modeling; Social Goals; Process Goals*

I. INTRODUCTION

Business Process Management (BPM) provides a platform for the management, measurement and improvement of business processes [2]. BPM has evolved over the years and an attempt has been made towards integrating social elements into the BPM lifecycle, although these discussions are still at their infancy [3].

Social BPM is when the essential elements of social software has been applied to the different stages of BPM. This is to effectively optimize and improve the efficiency of the traditional BPM systems. [20] defines social BPM as a "methodology for bringing more and diverse voices into process improvement activities".

Recent research has started to introduce elements of social software into the BPM lifecycle [1][3][5][7][13].

This paper aims to present the research which has been carried out so far in the area of social BPM, identify the gap in the area which is related to social BPM during enactment, and finally to propose a goal-driven model which is guided by the social user behavior in order to overcome the limitation of the traditional BPM model and to contribute to the research in social BPM.

The paper is structured in the following way: firstly some basic concepts of BPM are explained and a number of limitations of the current BPM model are presented (Section 2).

In Section 3, social software is discussed. Section 4 overviews social BPM, while in section 5, goal-based modeling has been proposed as a way forward in social BPM. In Section 6, a proposed recommendation mechanism in the context of user's social behavior is discussed. Finally in

Section 7, a conclusion is drawn from what has been discussed and the way forward for future research is proposed.

II. BUSINESS PROCESS MANAGEMENT

A. Overview of BPM

BPM is a field of knowledge where IT and management meet [9] and it "includes concepts, methods, and techniques to support the design, administration, configuration, enactment, and analysis of business processes" [10].

There are a various models regarding the overall BPM lifecycle, however for the purpose of our research we have adapted the model proposed by Aalst et al [12].

According to van der Aalst et al. [12], the BPM lifecycle consists of four main stages: process design, implementation, process enactment and finally process improvement/evaluation.

B. Limitations of BPM

The current state of BPM practice has a number of limitation which some of them are listed below [14]:

Lack of Information Fusion: The traditional BPM model follows a top-down approach where processes are designed and imposed on the users to follow [13] [23].

Model- Reality divide : Users not having a say in the process design stage results in them not using the processes; this in turn creates a gap between what the process actually is and what happens in real life [4][13][23].

Information Pass-On Threshold and Lost Innovation: Valuable feedback from users regarding the processes remains unused due to hierarchical controls which can prevent sharing of knowledge [4][13].

Strict Access-Control: In traditional BPM systems strict access control for the users is quite common and only actors which have been chosen and given specific access are allowed to execute them [11].

Lack of Context: The BPM engine focusses on controlling the flow and interaction of the data independent of the rest of the system. Therefore, the contexts of each case in which the data are appearing or the transactions are taking place are not easily accessible [16]. It is seen that majority of the business processes in one way or another interact with their surrounding environments, hence it is important to have an understanding of the context in which an activity or process is taking place in order to be able to improve them.

III. SOCIAL SOFTWARE

"Social software is a general term encompassing a set of tools and applications that enable group interaction and computer-mediated communication"[17]. Today, the trend of growth in the usage of social software in different platforms is increasing rapidly. We argue that social software concepts can be adapted in order to improve and overcome some of the challenges found in the current business process management systems.

The fundamental elements of social software is enabling interaction and maximizing participation between users and communities. Some of these have been mentioned in [15], listing the following characteristics: **Weak Ties** (allow spontaneous creation of contacts between non-predetermined individuals [19]), **Social Production** (Enable innovative and unexpected contributions and feedback regardless of geographical location [17] [1]), **Egalitarianism** (The maximization of participation through the role of trust and the equal rights that the users have to contribute to eventually reach the best practice [19]), **Mutual Service Provisioning** (The transformation of the organization's model from a one-way service system to a service-exchange system [19]. This results in the customer having a say in the design of the products and services they receive).

IV. SOCIAL BUSINESS PROCESS MANAGEMENT (SBPM)

A. Overview of SBPM

Social BPM is the intersection of social software and BPM to overcome some of the limitations of the traditional BPM systems. Social BPM is a "methodology for bringing more and diverse voices into process improvement activities" and it describes as collaboratively designed and iterated processes [20].Majority of the definitions given about SBPM until now mostly emphasize the role of collaboration during the process design stage of the BPM lifecycle and neglect how the SBPM engine could work during runtime.

SBPM offers a platform that enables contribution from various users of the community during all the stages of the BPM lifecycle. This increased participation ensures the users actually follow the processes during enactment; this overcomes the model-reality divide limitation which has been identified in the traditional BPM systems. However what has been discussed in SBPM is minor contributions of social software to BPM and not a framework explaining what the different elements of the SBPM are.

Therefore there is much to explore in order to exploit and fully benefit from the integration of BPM lifecycle and social software [3] [4]. BPM analysts believe that BPM application developers are socializing, not the end users themselves [8]. So in other words, based on the current understanding of SBPM, the developers are using social software in order to improve their experience of BPM and this ignores runtime social collaboration [15].

Social features not only play an integral role in the design of the processes, but the BPM lifecycle also benefits from these features during the execution of the processes. [22].

Adapting a SBPM framework not only needs a fully functional and validated model for it to be adaptable in an organization, but rather the organization also needs to be prepared to adapt to a whole new way of approaching support [27].

B. The Potential and Limitations of SBPM

There a number of potential benefits which SBPM can produce, some of these include as they have been elicited in [14]: **exploitation of weak ties and implicit knowledge** [1][3] (allowing the discovery and capturing of tacit and informal knowledge), **transparency** [3][5] (increase of procedure visibility in the organization), **decision distribution** [3][16]: (more well-informed and collective decision making), **knowledge sharing** [3][28][5] (allowing easier dissemination of knowledge).

Social software and traditional BPMS have different characteristics due to which their integration can also bring about a number of disadvantages and raise some challenges for the business. Potential limitations of SBPM have also been identified which include: **learning effort** [4][1] (major cultural shift required in the organization), **security** [4][5] (access rights given to many people can create an insecure environment), **quality** [4] (the quality of the content could be reduced as input is received from different people), **difficulty to evaluate** [5] (due to scattered benefits of SBPM it is difficult to evaluate them), **process management** (due to the flexible nature which exists it becomes a challenge to manage the interaction and overall processes)[14].

C. SBPM at Process Design and Enactment

There has been much discussion about how BPM can benefit from social software during the process discovery stage. In SBPM, users are all involved in the modeling of the processes [1], unlike the current approach in traditional workflow systems.

Brambilla et al. [7] have proposed an extension of the BPMN which includes using features of social software; it consists of various technologies and tools in order to include the contributions of the users from the outset of the process design stage. Although the idea of collaborative input at the design stage of the processes has been used previously by the Scandinavian School [28] in the 1970s.

Overall, although this approach eases the process of design and allows a wider collaboration from the users, however it does not remove the sequential and rigid nature of the traditional BPM systems which is the root cause for many of its limitation. Thus, this extension would not lead to a truly socially driven SBPM framework.

Kemsley [1] defines enactment or runtime collaboration as the modification of the processes during execution. This includes catering for unplanned participation of the users during the runtime of the processes. However no research in the area of social BPM, to date, has proposed a framework to

For SBPM [15]. Therefore this is the main area in which further research needs to take place and a clear way forward to overcome the sequential nature of the processes in BPM to be proposed.

V.   FURTHER RESEARCH INTO SBPM

A. Goal-Oriented Modelling

Goal-oriented business process modeling [32], strives to document the intention and goals behind the activities. Goals are defined as objectives that should be achieved by the system and its environment [34]. Goals are also commonly used in requirements engineering to ascertain the state which the user would like to reach after performing a specific activity [33].

Most of the research about social BPM mainly revolves around SBPM in the design stage. Adapting the approach of goal-based modeling and following this structure during the enactment stage, will also affect the design stage of the BPM lifecycle and it would need to adapt the proposed goal-oriented modeling approach [15].

As mentioned by Kemsley [1] social BPM may change the very nature of process modeling, so depending on the approach taken during the enactment, the process design approach might also need to be changed.

The defined goals may have sub-goals which need to be accomplished in order to fulfill the main goals [1] [24].This is in contrast to the traditional procedural based BPM systems [24] where performers are bound to follow a set of predefined rigid series of steps.

This approach would produce the desired flexibility and would be a platform for collaborative process discovery through defining end goals. The more the goals and sub-goals are defined in detail, the more robust and rigid the workflow, while the fewer and more general the goals , the more flexibility is given to the user to come up with their own set of activities. This heavily depends on the nature of the process, but in general cases, when there are many goals for a relatively small scenario, it is an indication that there is limited flexibility during the execution of the processes[15].

When the goals are fewer and are defined in a more general manner, it allows for a higher degree of flexibility and dynamic behavior for the user(s).

The priorities of the goal/sub-goals set, need to be in line with the general policies and aims of the company. These goals could be re-evaluated, subject to collective approval based on the suggestions for improvements received from the wider community and users. The continuous engagement of the end users and clients with the system and processes would also allow more insightful feedbacks from the customers which could be used to enhance the goals and processes [15].

Figure 1 presents a visual overview of our proposed goal driven SBPM process, out of which we plan to derive a meta-model.

Goal-based modeling is a potential approach in order to overcome the limitations of the traditional BPM models and propose a comprehensive model for social BPM where a dynamic and social behavior is visible in all the stages of the BPM lifecycle.

B. SBPM at Enactment

In most workflow management systems, users are expected to execute the predefined tasks in a specific order and sequence. When it comes to SBPM, this traditional approach is not in harmony with the flexible characteristics found in social software. Social BPM should enable the collaborative modeling and execution of processes with a certain degree of flexibility. In social BPM, the flow and sequence of actions should not be enforced upon the user, but rather it should allow and support the enactment of the processes, exploring the sharing of knowledge and the business best practices [5].

According to Brambilla et al. [6] participatory enactment transfers this collaboration in the process design stage to the execution and runtime of the processes. This is because the processes are designed on the flyer at the time of their execution therefore there is a seamless integration of the design and execution stage of BPM lifecycle.

By adapting a goal-based modeling approach in the process discovery stage of the BPM lifecycle, during the enactment of the processes, the users will have a certain degree of flexibility and freedom when executing the processes. On the one hand, the goals defined in the design stage act as controls and provide direction to the sequence of activities which need to be performed, and on the other hand this will leave an unstructured space for the user to decide their own course of action in order to accomplish the particular goals [15]. Figure 1 shows how the goal is explicitly mentioned, but how to achieve them is simply recommended by the system through different alternatives, and the user is free to decide how the tasks are executed.
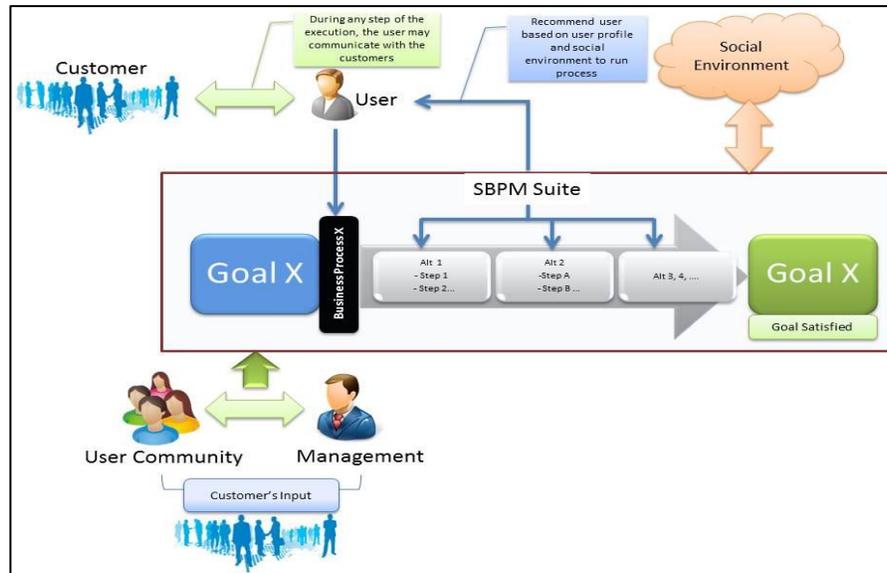
Fig. 1. Goal-Driven Social BPM Interaction Model

VI. SOCIALLY GUIDED BUSINESS PROCESS MODELLING

The model in figure 2 illustrates the different elements of the model which are explained in this section. This is a high level view of how these features are linked to one another and further research needs to be carried out in order to explore in more detail how they will function in practice.
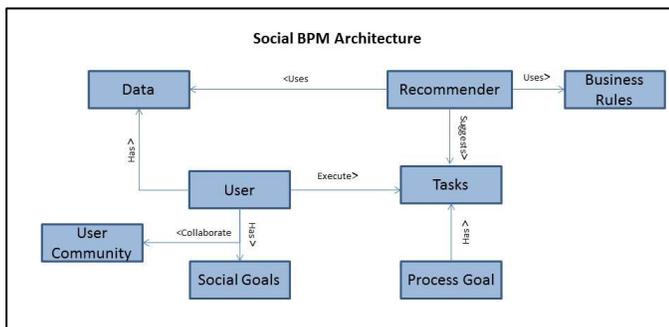
Fig. 2. Social BPM Architecture

A. User-System Interaction through Recommendation

In order to achieve a specific goal, the user is flexible to choose the course of action he chooses. This however is guided by the system in order to firstly ensure the steps taken lead to a state or goal which is expected, and secondly, so that the steps taken can be monitored, captured and re-used by the Social BPM suite in the future.The system can make process /step recommendations based on the data gathered from the user. This data would determine the suitability and interest of the user to be engaged in a process to achieve a given goal. The process recommendation system [26] based on this data interacts with the user in order to achieve the sub-goals which are needed in fulfilling the overall goal.

The concept of recommender systems have extensively been used previously in e-commerce, however they have not been utilized in the context of process/task recommendations.

For the purpose of the SBPM framework, recommendations based on Hybrid approach would be most suitable. This is because it combines recommendations that have been made in the past (content–based recommendations), with recommendations made by other users with similar tastes and preferences (collaborative recommendations)[18].

During this interaction, the user is given a certain level of flexibility to insert or remove steps into the processes. This can involve, for example, eliminating a certain step by proposing and achieve a sub-goals through a different route which might be better than the one suggested by the system. In this context, flexibility here is "the ability to adapt the process flow on demand through adding, skipping, or sequence reordering of process steps" [21].

Figure 1 captures this concept, by illustrating how the goals are set initially, and then based on them, how the system can make recommendations. Users then have the option to follow the recommendation or choose an alternative step.

B. Business Rules

Constraints and dependencies are from the essential elements of recommendation mechanism. Because before the recommender assigns or suggests a tasks, it should be guided by a number of business rules which capture any possible dependencies of the tasks. In other words, the sub-goals are recommended in such a fashion that their pre-requisites have already been fulfilled before the recommendation. Otherwise what is suggested may not be executable by the user or the system as there are one or more dependent tasks which need to be satisfied first.

C. Social Goals

As the proposed model is based on a goal-oriented approach, it is important to specify what type of goal is referred to when discussing goals in the context of social BPM.

The topmost goals are process goals, which are dependent on the cases study and specific scenarios. These might have various sub-goals which need to be fulfilled in order for them to be satisfied.

The other types of goals in this model are *social goals* and relate to the user/customer and their motivation for getting involved in a particular process. These goals can vary from person to person, and can belong to categories such as social

affiliation goals, social responsibility goals and social concern goals [25].

VII. CONCLUSION AND FUTURE RESEARCH

In conclusion, SBPM research is at an early stage and there is still lack of an effective method for applying the social elements to the enactment stage of BPM. The solution proposed in this paper is in terms of working towards a meta-model of the social BPM lifecycle that applies goal-oriented modeling principles to the design and enactment stages. With the use of goal-based modeling we are able to make enactment more flexible and allow collaborative participation of the processes.

This flexibility, however, needs to be controlled, so the enactment does not divert from its main path and from achieving its end goal. According to Ramdas [31] there are different degrees of flexibility found when it comes to the interaction of the processes:

1) Ad hoc/Unstructured: there is no ordered workflow, decisions and actions are performed on the fly and these processes are normally manually executed and consist of unstructured activities.

2) Structured: These types of processes are well-defined with the performers and series of activities predefined.

3) Collaborative: these processes include on the fly interactions, however within a certain defined boundary and framework.

Our research, in particular addresses the collaborative aspect by proposing a goal-based approach. A similar sort of categorization has also been proposed by Brambilla et al [30] ranging from a closed BPM system which the tasks are highly rigid and actors all predefined, to the process mining category where the activities are executed freely and there is no set of activities defined.

In our approach, the steps to be performed are simply suggested by the system based on the previous iterations and through a process recommendation mechanism, the user is then free to either accept or reject the proposed step. It is important to note that the sub-goals required to satisfy the overall goal of the process must be fulfilled by the user. This gives the required flexibility to the user and at the same time preserves the integrity of the system to ensure the process and steps are moving in the right direction.

The recommendations are based on the data gathered from the user to ensure they have an interest in the process they are recommended to be involved. Other users who also have interest in the same processes can participate in achieving the recommended tasks.

Goal-based modeling has of course, its own limitations that need to be addressed before its adoption for social BPM. One of the most important challenges in defining the goals in this model is dealing with conflicting priorities between goals and identifying the goals themselves and not confusing them with the steps involved in accomplishing the goals [29].

In summary, this paper has focused on the current stage of our on-going research in the area of social BPM with a special emphasis on the process discovery and enactment stage and how socially guided BPM can change both the enactment stage of the BPM lifecycle as well as the process discovery stage as there is seamless integration of these two stages.

In future research we plan to formalize a metal-model for the goal-oriented SBPM, and to apply this model to a case study in order to validate our proposed approach. Real goals will be used from a case study and the meta-model presented in this paper will be apply to the scenario. This, through an evolutionary processes will lead to a standardized meta-model for a socially enabled BPM system. The different elements of the recommender system also needs to be explored further to provide a comprehensive framework for a socially driven, goal-based BPM model.

REFERENCE

[1] Kemsley, S.: Leveraging social BPM For Enterprise Transformation: In Social BPM Work, Planning and Social Collaboration Under the Impact of Social Technology, Swenson, KD., Palmer, N., et al.: BPM and Workflow Handbook Series, pp. 77 – 83, Future Strategies Inc, USA (2011).

[2] R.G. Lee, B.G. Dale, "Business process management: a review and evaluation", Business Process Management Journal, Vol. 4 Iss: 3, pp.214 – 225, MCB UP Ltd (1998).

[3] Brambilla, m., Fraternali, p., Combining social web and BPM for improving enterprise performances: the BPM4People approach to social BPM, In Proceedings of the 21st international conference companion on World Wide Web, pp. 223-226 ACM Publishers, New York (2012).

[4] Filipowska, A., Kaczmarek, M., Koschmider, A., Stein, S., Wecel, K., Abramowicz, w., Social Software and Semantics for Business Process Management - Alternative or Synergy?, in Journal of Systems Integration, vol 2, No 3, pp.54 - 69 (2011).

[5] S. Erol, M. Granitzer, S. Happ, S. Jantunen, B. Jennings, A. Koschmider, S. Nurcan, D. Rossi, R. Schmidt, P. Johannesson. Combining BPM and Social Software: Contradiction or Chance? In Journal of Software Maintenance and Evolution: Research and Practice, vol. 22 pp. 449–476, John Wiley & Sons, Ltd (2010).

[6] Brambilla, M., Fraternali, P., Vaca, C.: A Notation for Supporting Social Business Process Modeling Third International Workshop, BPMN 2011 Proceedings, vol. 95, pp. 88-102 , Springer Berlin Heidelberg (2011).

[7] Brambilla, M., Fraternali, P., Vaca, C.: BPMN and Design Patterns for Engineering Social BPM Solutions: In BPM 2011 International Workshops, vol. 99, pp. 219-230, Springer Berlin Heidelberg (2012).

[8] Swenson, K.: Who is Socializing in Social BPM? 12 May 2010 [Blog Entry], Collaborative Planning & Social Business. Available: http://social-biz.org/2010/05/12/who-is-socializing-in-social-bpm-2/ [Accessed Dec 2013].

[9] Filipowska, A., Kaczmarek, M., Kowalkiewicz, M., Zhou, X., Born, M., Procedure and guidelines for evaluation of BPM methodologies: In Business Process Management Journal, vol. 15 Iss: 3, pp.336 – 357. (2009).

[10] Weske, M.: Business Process Management, Concepts, Languages, Architectures, Springer-Verlag Berlin Heidelberg (2007).

[11] Statistics about Facebook users, http://www.checkfacebook.com/ [Accessed on 1st Feb 2013]

[12] Van der Aalst, W.M.P.: Business process management: a personal view: In Business Process Management Journal, vol. 10, no.2, pp. 744-791 Emerald Group Publishing (2009).

[13] Schmidt, R., Nurcan, S.: BPM and Social Software: In BPM 2008 International Workshops, vol 17, pp 649-658, Springer Berlin Heidelberg (2009).

[14] Rangiha, ME., Karakostas, B.: Towards a Meta-Model for Goal-Based Social BPM. (In press) In: Post-Proceedings of Lecture Notes in Business Information Processing (LNBIP) series, Springer, 6th BPMS2 workshop in conjunction with BPM 2013 conference, Beijing, China (2013).

[15] Rangiha, ME., Karakostas, B.:Goal-Driven Social Business Process Management. In: Proceeding indexed by IEEE Xplore and various international databases, to be published by The Science and Information Organization, Science and Information (SAI) Conference Proceedings, pp. 894 – 901, London, UK (2013).

[16] Van der Aalst, W.M.P., Weske, M., Gr¨unbauer, D.: Case Handling: A New Paradigm for Business Process Support: In Journal Data & Knowledge Engineering, vol.53, issue.2, pp. 129 – 162, Elsevier Science Publishers, The Netherlands (2005).

[17] Abbattista, F., Calefato, F., Gendarmi, D., Lanubile, F.: Incorporating Social Software into Agile Distributed Development Environments: In proceedings of the 23rd Automated Software Engineering (ASE) workshop, pp.46-51 (2008).

[18] Burkhart, T., Weis, B., Werth, D. and Loos., P.: Towards Process-Oriented Recommender Capabilities in Flexible Process Environments -

State of the Art: 45th Hawaii International Conference on System Sciences (HICSS), USA, Hawaii (2012).

[19] Bruno, G., Dengler, F., Jennings, B., Khalaf, R., Nurcan, S., Prilla, M., Sarini, M., Schmidt, R., Silva, R.: Key challenges for enabling agile BPM with social software: In Journal of Software Maintenance and Evolution: Research and Practice 23, 297–326 (2011).

[20] Richardson, C., Is social BPM a Methodology, A Technology, Or just a lot of Hype? 20 May 2010 [Blog Entry]. Forrester. Available: http://blogs.forrester.com/clay_richardson/10-05-20-social_bpm_methodology_technology_or_just_lot_hype[Accessed Dec 2013] .

[21] Dorn, C., Burkhart, T., Werth, D and. Dustdar, S. : Self adjusting Recommendations for People-Driven Ad-Hoc Processes: in Business Process Management. pp. 327-342: Springer, Berlin / Heidelberg (2010).

[22] R. Schmidt, F. Dengler, and A. Kieninger.: Co-creation of value in IT service processes using semantic mediawiki: In BPM 2009 International Workshops, v.43, pp. 255-265. Springer Berlin Heidelberg (2010).

[23] Palmer, N.: The Role of Trust and Reputation in Social BPM: In Social BPM Work, Planning and Social Collaboration Under the Impact of Social Technology, Swenson, KD., Palmer, N., et al.: BPM and Workflow Handbook Series, pp. 35-43, Future Strategies Inc, USA (2011).

[24] Pucher, MJ.: How to link BPM Governance and Social Collaboration through an Adaptive Paradigm: In Social BPM Work, Planning and Social Collaboration Under the Impact of Social Technology, Swenson, KD., Palmer, N., et al.: BPM and Workflow Handbook Series, pp. 57 – 75, Future Strategies Inc, USA (2011)

Measuring Homophily in Social Network: Identification of Flow of Inspiring Influence under New Vistas of Evolutionary Dynamics

Hameed Al-Qaheri

Department of Quantitative Methods and
Information Systems
Kuwait Universit, Safat, Kuwait

Soumya Banerjee

Department of Computer   Science
Birla Institute of Technology, Mesra,
India

Abstract—**Interaction with different person leads to different kinds of ideas and sharing or some nourishing effects which might influence others to believe or trust or even join some association and subsequently become the member of that community. This will facilitate to enjoy all kinds of social privileges. These concepts of grouping similar objects can be experienced as well as could be implemented on any Social Networks. The concept of homophily could assist to design the affiliation graph (of similar and close similar entities) of every member of any social network thus identifying the most popular community. In this paper we propose and discuss three tier data-mining algorithms) of a social network and evolutionary dynamics from graph properties perspective (embeddedness, betweenness and graph occupancy). A novel contribution is made in the proposal incorporating the principle of evolutionary dynamics to investigate the graph properties. The work also has been extended towards certain specific introspection about the distribution of the impact, and incentives of evolutionary algorithm for social network based events. The experiments demonstrate the interplay between on-line strategies and social network occupancy to maximize their individual profit levels.**

Keywords—*Homophily; Affiliation; Embeddedness; Betweenness; Graph occupancy; Evolutionary dynamics*

I. INTRODUCTION

Different properties of social network have demonstrated potential interplay between events, participants and   social network itself.   There are numbers of instances, where, attribute of social network could drive the application area of the network itself **[21].** Hence, certain properties of social network have an emerging impact **[27]** and homophily like behavior is definitely one of them.   Prior research demonstrates impressive role of such behavior on the application and analysis of social network **[28] .**

Homophily **[1] [5] [6] [20]**, the tendency of individuals to form association with individuals of similar socio-cultural background, becomes the basic governing structural component of any social network and it has been the focus of many social network studies **[2] [7].** Social network studies reveal that *"social networks are homogeneous with regard to many socio-demographic, behavioral and interpersonal characteristics"* **[3].** In any existing social network such as Facebook or Twitter there lies some common functional attributes such as *'posting photos', 'sending messages', 'likes',*

'dislikes', etc.. These kinds of activities lead to the concept of affiliation towards a community **[20].** From influence of social propagation**,** Facebook and Twitter are dedicated to disseminating the information and thus the concept of Twitter follower graph and cascading of influence also reinforces the hypothesis of different influence measurement model **[4].** Considering the broader definition of the problem, this paper finds a close similarity between graph theory and a social network homophilic structure and explores the empirical significance of influence propagation or a popular community ranking and detection. The paper validates the existing graph postulates with a proposed mining algorithm and simulation applied a Facebook data set. Investigation yields certain significant results with regards to popular community structure and ranking based on different classical graph theory properties like path traversal, Betweenness and Embeddedness of social network nodes. After intial validation through graph simulation, an initiative has been solicited with a self-organizing   and   evolutionary   principle,   which   could dynamically trace the variants of social network. The role of evolutionary dynamics **[20]** is also considerably significant as it is defined as a study of the mathematical principles according to which life has evolved and continues to evolve. The evolution is also visible in the formation of social graph. Reinforcement and validation of graph properties has been demonstrated through the algorithmic strategies coined from evolutionary dynamics **[13] [14].**The remaining part of the paper is organized as follows: Section 2 elaborates the statement of the problem with the parameters of graph theory followed by existing methodologies, examples of social and affiliation graph in section 2A and motivation of the analysis has been discussed in section 2B. Section 3 describes mathematical treatments responsible for proposed algorithm, presented in section 4. Section 5 discusses the data set for experiments and their implication on the graph properties of social network. Section 5.1 introduced the role of evolutionary dynamics to validate the simple graph properties for social network instances, which may implicate in the mining of graph related inferences. Finally section 6 gives conclusion and mentions further scope of relevance research on the paradigm.

II. STATEMENT OF THE PROBLEM

A community is formed in order to propagate or transfer or share different knowledge across the network. And to disperse

the utility of one's community, it is required to apply some techniques, such as voting or polling, on which knowledge about the particular community has to be shared. When forming community, certain basic parameters along with the links among the members of the community are to be considered. Parameters and links are as follows:

- Friend list
- Community links
- Graph occupancy period: Initial time and final time
- Paths and connectivity
- Path traversed
- Embeddedness
- Betweenness of nodes
- Affiliation
- Co-evaluation of social and affiliation network

A. Friend List

Friendship is developed in a social site based on some common factor, like as, members belonging to same school, working area, community and so on. Similar type of characteristic people can be blocked into a common structure and even a few from the block can also belong to some other structure based upon their choice. [8] The homophily test of friendship in a social site can be interpreted with the help of the following example. Let there be a network where **"m"** fraction of all individuals are male and **"f"** fraction of all individuals are female. Considering a given edge in this network, if we independently assign each node the gender male with probability **"m"** and the gender female with probability **"f"** then the both ends of the edge will be male with probability **"m^2"** and similarly both ends will be female with probability **"f^2"**. But if the first end of the edge is male and the second end is female or vice versa then there exist *'cross-gender edge'*. This condition will take place with a probability **"2mf"**. Thus the test for homophily according to gender can be summarized as – if the fraction of cross-gender edges is less than **"2mf"** then there is a presence of homophily [8].

B. Community Links

Community can be created by a group of members by selection and social influence method. The tendency of people to form friendships with others who are like them are termed as *selection* **[8].** The selection criteria are mainly race or ethnicity or similar characteristics. People may also modify their behaviors to bring them more closely into alignment with the behavior of their friends. This process is vividly described as *socialization* and *social influence* **[8].**

The individual similar characteristics drive the formation of links but social influence is a mechanism by which the existing links in the network serve to share people's characteristics.

C. Graph occupancy period: Initial time and final time

It is the amounts of time spend on visiting a node. The duration of remaining in a focus is calculated by checking the difference in final time and initial time. This indicates the graph occupancy period.

D. Paths and connectivity

According to the social scientists John Barnes defines graph theory as "Terminological jungle, in which any newcomer may plant a tree."[8] A *path* is defined to be as a sequence of nodes with the property that each consecutive pair in the sequence is connected by an edge. The paths can also be analyzed as not just the nodes but also the sequence of edges linking these nodes. Connectivity can be described by saying a graph is connected if for every pair of nodes, there is a path between them. If a graph is not connected, then it is separated into a set of connected pieces. Connected components of a graph are a subset of the nodes such that the following two properties hold [8].

1) Every node in the subset has a path to every other [8]
2) The subset is not part of some larger set with the property that every node can reach every other [8].

E. Path Traversed

The path traversal is important in spread of important information. It is required to examine whether something flowing through a network has to travel just a few hops or more. The *"Length of a path"* is the number of steps it contains from beginning to end that is the number of edges in the sequence that comprises it. The path traversal technique used over here is the *"Breadth - First - Search"*. The method of the traversal is one just need to keep discovering nodes layer-by-layer, building each new layer from the nodes that are connected to at least one node in the previous layer. Since it searches the graph outward from a starting node, reaching the closest node first, it is named as breadth-first-search [8].

F. Embeddedness

The number of common neighbors the two end points in a network has referred to as embeddedness of an edge. This is illustrated with the help of a schematic diagram in **schema1**.
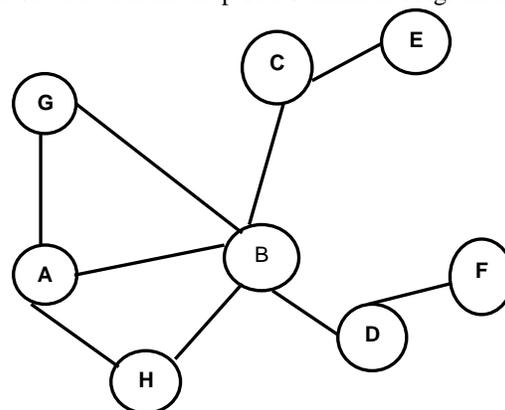
Schema 1 An affiliation network Case

Here, embeddedness for two node A and node B has two common neighbors' node G and node H. Thus the concept of embeddedness provides information that if two individuals are

connected by an embedded edge then this makes it easier for them to develop a trust level and generate confidence for transferring vital information or interacting with each other [8].

G. Betweenness

Betweenness of a node is explained as the total amount of flow that it carries when there exists a unit of flow between each pair of nodes is divided up evenly over shortest path. Nodes with high betweenness occupy critical roles in the network structure. To compute betweenness efficiently we use the notation of breadth-first-search. For a give graphical structure the calculation of betweenness is done on the perspective of time. For each given node the total flow from that node to all other is distributed over the edges. This technique is applied on every node in order to simply add up the flow from all of them to get the betwenness on every edge [8](shown in schema 2).
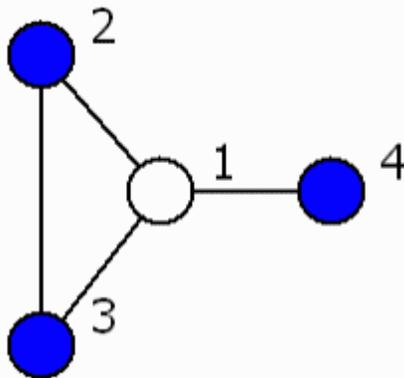
Schema 2: Local betweenness(the local betweenness of actor 1 is 2

H. Affiliation

Affiliation, a concept that is associated with homophily graph, **[8] [9]** can be used to represent the participants, i.e. a set of people, in a set of foci (representing some kind of community). For example, node A, representing a person could participate in focus X through an edge. These kinds of graph are said to be affiliation network, since it represents the affiliation of people (on left) with foci (on right). Affiliation network is one of the examples of the bipartite graph. **Bipartite graph:** A graph is said to be bipartite if its nodes can be divided into two sets in such a way that every edge connects a node in one set to a node in the other set [8].

Scheme 3 is an example showing nodes A and B representing people participating in and Literature Club and Soft Computing) foci.

I. Co-evolution of social and affiliation networks

New friend links are formed and people become associated with new foci over the period of time.
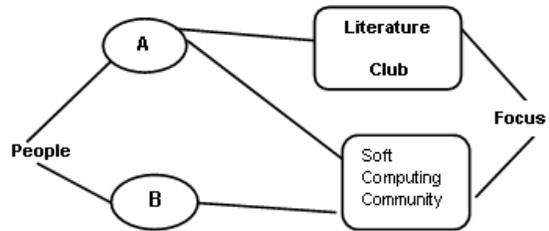
Schema 3 Affiliation through bipartite graph.

This kind of formation can lead to a kind of co-evaluation which might indicate the selection choice of each individual and there social influence. For example if two people belong to a same focus then there is a probability that they become friends and can influence each other with their community they belong. According to the graph theoretic representation nodes are used as both people and foci but the difference is created by distinct type of edges. Firstly, an edge in a social network, it connects two people and indicates friendship. Secondly, an edge in an affiliation network, usually known as 'social-affiliation network'. This edge connects a person to a focus and designates the operation of the person in the focus. These two parameters can be resembled in the following **schema 3**.
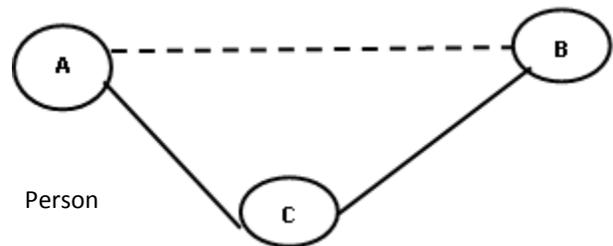
Schema 3 (a) When A, B, C are three different persons.
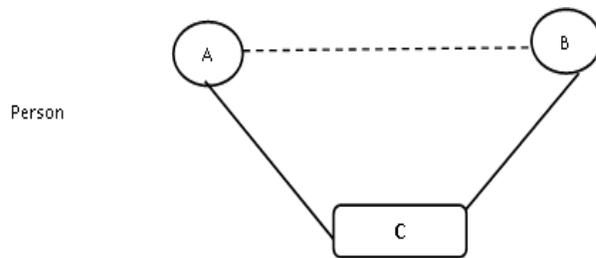
Schema 3 (b) A and B represent people but C denotes focus.

III.   RECENT TRENDS AND MOTIVATION

While exploring the incentive generation, the basic economics model of incentive distribution has become crucial trend to be studied. Research already revealed the impact of incentives on worker self-selection in a controlled and restricted laboratory experiment. Subjects face the choice between a fixed and a variable payment scheme. Considering the status of the treatment, the variable payment is a piece rate, a tournament, or a revenue-sharing scheme [22] [23].

The extension of applying the graph properties of social network is broadly inspired by Julia Poncela Casasnovas's research **[24].** She addressed the study of the evolution of cooperation on complex networks, using among the different social dilemmas. Emphasizing mainly on the Prisoner's Dilemma game as a metaphor of the problem, her research analysed possible outcomes of the dynamics, depending on the underlying topology. Very recently, in 2013, it has been pointed out that the topology not only highlights homophily or other associated properties but also it leads toward role discovery problem. role discovery problem **[25]** finds groups of nodes that share similar topological structure in the graph. But is it only topology or economies of incentive to quantify the distribution of pay off under network? We investigated from significant work on evolutionary games of Nowak et. al. **[26]** and found that even payoff determines reproductive rate and successful individuals have a higher payoff and produce more offspring. Still it cannot be assured that the payoff also could be signified by the carrying capacity of individual participants in networked games. Finally, this work adopts the emerging strategies of evolutionary game based incentive distribution for any instance of social network under test.

IV. EXPLORING MATHEMATICAL TREATMENTS

The most influential community (focus) can be determined by the frequency of the clicks made by the individual nodes. However, visiting a focus and being a follower and then a member of that focus are two different aspects. Visiting a node could mean only collect information while being a member means making the community well know. As such, to find the most influential focus we need to find all possible paths from the source to the destination and finding the shortest path by computing the Betweenness values. Considering the graph one node at a time, the distribution of the total flow over the edge from that node to the other nodes could be computed. And hence, the betweenness of the every node could be calculated as follows:

$$\text{Betweenness of every node} = \sum \text{flows from different nodes} \quad (1)$$

The shortest path traversed from the source to the destination can be found by the *Breadth-First* Search algorithm. Hence the number of shortest path to each node should be the sum of the number of shortest path to all nodes directly above it in the *breadth-first* search [8]. Valuation of the members of the community or focus can be denoted by the time spent on it by a particular node. Here lies the concept of Graph Occupancy, which can be calculated by calculating the time difference between Final time of leaving the focus and Initial time of entering the focus. Mathematically, this can be expressed as follows:

$$f_i = b^t \qquad /* \text{ frequency is non-linear relation with}$$

$$\text{betweenness.} \qquad (2)$$

Where $t = \{1, 2, \ldots \ldots m\}$

Since we concentrate on each node for calculating the betweenness and thus finding their frequency of participants in

making a focus famous thus we have the following set of Betweenness (B_i) and Frequency (F_i).This is expressed as:

$$B_i = \{b_1, b_2, \ldots \ldots b_n\} \text{ and } F_i = \{f_1, f_2, \ldots \ldots \ldots f_n\}$$

Equation 2, describes the frequency of acceptance of a focus by different nodes. Based on a number of choices made by each individual, we can select out the most beneficial or famous (leader) communities, in terms of the population focus. This entire logic can be expressed mathematically as:

$$g(p) = \Pi\left(\sum_{i=1}^{n} B_i + F_i \sin(\Pi p) + C\right) \quad (3)$$

where, p is the popularity of a node and C is a constant for each delay unit process. In equation [3], the number of paths to be followed to reach the destination is obtained by the betweenness calculation denoted as B_i and frequency of the particular node can also be obtained and denoted as F_i. The summation of these two values yields the minimum possible path to be traversed to point the popularity of the community. The maximum popularity of a community could be found by the product of this result, in other words, by applying the concept of *Max-Min* function, the maximum popularity of a focus can be obtained for the minimum path travelled from the source to the destination.

V. PROPOSED ALGORITHM -I

Finding the most influential community could be achieved through the following algorithm which has distinct blocks to evaluate the popularity and influence of node(s). The algorithm contains certain unconventional graph properties such as betweenness and embeddedness.

1. Begin
2. Initialization of links present in between the nodes and focus
3. Initialization of variables

Betweeness=b_i,   b= $\{b_1, b_2, \ldots b_n\}$

Graph frequency= (f_i), f_i= $\{f_1, f_2, \ldots f_n\}$

Popularity P

4. Finding value $f_i = b^t$ $\{t=1, 2, 3 \ldots n\}$

5. Calculating the betweenness:

Betweenness of every node= $\sum$ flows from different nodes

6. Values form the betweenness leads to find the shortest path

7. Calculating the shortest path based on Breadth first search

8. Select a random node n_i which has visited a particular community node at least once $n_i \epsilon N$

9. w*hile* $b_i \leq b_n$
do

        Calculate the Graph frequency f_i

      i*f* $f_i \geq f_{i+1}$
then

          F= f_i

end if

        Calculate g(p)= $\boldsymbol{\pi}(\sum_{i=1}^{n} B_i + F_i \sin(\pi P) + C)$

10. end while
11.   end begin

VI.   IMPLEMENTATION AND ANALYSIS

The data set for validating the proposed model was collected from Facebook community network. The reason to choose Facebook was its close resemblance, in terms of nodes, betweenness and edges, to the classical graph theory models. A set of nine homophilic nodes, at a particular time, were selected for initial validation. In these sets, each node is connected to its influential nodes, which is denoted by the links and their weights. There are communities that belong to the nodes and each of these nodes tries to promote their own communities.

To denote the most influential focus or community, we need to find the shortest path (which also motivates the other nodes to join the community and increase the popularity). *Embeddedness* of the nodes are also considered to identify the common neighbors among them. **Table 1** shows the parameters to be considered for the proposed algorithm. The output value in the table was computed after implementing the algorithm (using MATLAB version 7.10.0.499 (R2010 a)). The detailed explanations of the parameters are as follows: **Friend List:** 9 different Nodes represent 9 different friends with their communities

Link Present: edges among the nodes with weights.

Embeddness: A→B edge having common neighbors. There may be present or might not be present. If not present, then denoted by NIL. If present then node number is given.

Betweenness: involves reasoning about the set of all shortest paths, between pairs of nodes.

Graph occupancy: amount of time spent on a node. Leading to the further popularity of a community and increasing the node linkage. The corresponding outputs are given in the following **Table 1.**

TABLE I.   LIST OF NODES WITH THEIR CONNECTIVITY, FLOW OF INFORMATION AND FURTHER INCREASE IN POPULARITY

Parameters						
Friend list	**Links present**	**Embeddedness**	**Betweenness**		**Graph occupancy**	
Node 1	(4,5,6)	Node 4	(1,5) (1,6)	0.2100 1.5300	(1,1) (1,2) (1,3) (1,4) (1,5) (1,6)	0.0541 0.0462 0.0474 0.0148 0.0151 0.1478
Node 2	(3,5,6)	Node 3	(2,3) (2,5)	0.5100 0.3200	(2,2) (2,3) (2,4) (2,5) (2,6)	0.0591 0.0793 0.0131 0.0213 0.0106
Node 3	(4,5,6,7)	Node 4	(3,4)	0.1500	(3,1) (3,3) (3,4) (3,5) (3,6)	0.0046 0.0266 0.0145 0.0135 0.0081
Node 4	(1,3,5,6,8)	NIL	(4,1) (4,6)	0.4500 0.7900	(4,1) (4,2) (4,3) (4,4) (4,5) (4,6)	0.0435 0.0221 0.0446 0.0591 0.0065 0.1234
Node 5	(1,2,3,4)	NIL	(5,3) (5,4)	0.3200 0.3600	(5,1) (5,2) (5,3) (5,4) (5,5) (5,6)	0.0111 0.0135 0.0827 0.0381 0.0375 0.0194
Node 6	(1,2,3,6)	Node 4	(6,2) (6,3)	0.4100 0.2900	(6,1) (6,2) (6,3) (6,4) (6,5) (6,6)	0.0028 0.0377 0.0352 0.0030 0.0119 0.0591
Node 7	(3)	NIL	(7,3)	0.4500	(7,2) (7,3) (7,4) (7,5)	0.0191 0.0267 0.0046 0.0070
Node 8	(9)	NIL	(8,9)	0.5500	(8,2) (8,3) (8,4) (8,5)	0.0466 0.0652 0.0113 0.0171
Node 9	(4,8)	NIL	(9,4)	0.3000	(9,1) (9,3) (9,4) (9,5) (9,6)	0.0092 0.0531 0.0290 0.0271 0.0162

Fig. 1.   Nodes with their links and weights

Fig. 2.   Enhancements in new connections with the existing invoking nodes

Figure 1 represents nine different influential nodes with their links represented in the directed graph format. Each of the edges has their weights marked on it. Here, the shortest path has been considered from a source node say Node1 and the destination node, Node 6.Time duration, that is the amount of time spent by the nodes in their communities, can be an invoking factor to join their community. Using equation (3), **Figure 2**demonstrates that the most influencing or inspiring node enhances the number of functionally active node. In this figure, it is shown that the inspired nodes have added on more of the connectivity with other different nodes.

To crawl Facebook, we implemented a distributed, multithreaded crawler using Python with support for remote method invocation (RMI) **[11].** Facebook provides a feature to show 10 randomly selected users from a given regional network; we performed repeated queries to this service to gather 50 user IDs to "seed" our breadth-first searches of social links on each network[1]**[11].**

A. Role of Evolutionary Dynamics

Based on the intial simulation, it has been demonstrated that there exist a strong cohesive directions with graph theory and social network in the context of homophilic community detection.

We need further investigation for the specific attributes in the context of more homophily identification from the perspective of either directed or undirected graph. The extension of the algorithm will be significant to quantify the application specific investigation of homophily.   The extension also could revalidate the correlation of graph properties under social network.

Inspired by the phenomenal contribution in evolutionary game theory by Nowak and his colleagues, several non linear characteristics have been started adopting the concept of evolutionary dynamics [12] [13]. It is evident that evolutionary dynamics are defined by nonlinear differential equations and therefore can be imported for revalidating complex graph and network with the growth of a social graph as mentioned in the first part of the algorithm. An evolutionary dynamic assigns each population game F an ordinary differential equation [14] $\dot{x} = V^F(x)$ on the simplex x. One simple and general way to define an evolutionary dynamic is via a growth rate function:

$$g : R^n \times x \rightarrow R^n \ (4)$$

Here, g represents the (absolute) growth rate of strategy i and it will as a function of the current payoff toreward the strategy. The previous algorithm only considered the shortest path, breadth first search and graph frequency. As Evolutionary Dynamics can also retrospect the growth aspects, propagation and mutation of message under any state of graph, hence therefore the conventional graph properties have been revalidated using the extended algorithm. It should be mentioned that proposed algorithm tries to incorporate potential strength of evolutionary dynamics for simulating

---

[1]http://code.google.com/p/crawl-e/

the behavior and growth of social graph structure. Series of more precise plots have been accomplished in post simulation of extended algorithm **(Figure 4-8)**. Persistence across the network also indicates the out degree and also any average number of distinct tags of groups and of tag assignments of users having k_{out} neighbours could be evaluated from social network as shown. The users, who have more contacts in the social network, tend also to be more active in terms of tags and groups. Average number of distributive tags denoted as n_t. group tags on the specific message as n_g and subsequently n_w represents the list of predefined choices. Correlations between the activity of participants and their number of declared friends and neighbours can be identified: here also the k_{out} neighbours have been taken into consideration. The data has been log-binned **(Figure 3)**: by definition a bin of constant logarithmic width signifies that the logarithm of the upper edge of a bin (x_i+1) is equal to the logarithm of the lower edge of that bin (x_i) plus the bin width (b) **[19]**. Here, the symbols indicate the average, and the error bars with near optimal 25 and 75 percentiles for each bin. The algorithm deploys the Python 2.7.3, which was released on April 2012. The present implementation also allows the feature of Automatic numbering of fields in the str.format() method, which have been reflected in the post implementation stage of algorithm.

Algorithm II: Graph Pattern (G, P (t), μ)

1. Define the initial state of the graph , i.e. define $G_i = \delta_i$

   /* δ_i is the [2]Kronecker symbol*/

2. Solve for Pi(t), which provides functions P(t) and →i(t)
   /*P(t) : Probability of sample occupancy time for t_m

   Evaluate: $\dfrac{dP}{dt} = -\dfrac{dY}{dt} = -\sum_i \mu_i P_i$

   (5)

   /* The probability Yi(t) that at least one mutation has occurred while the system was at state i before time t*/

3. Initialize the system with N classes of social network instances at time t = 0, with its intialstart of connection, k_{out}: out-degree of graph, occupancy on graph, and termination instance:

   evaluate: $\rho_i(t_m) = \dfrac{\mu_i P_i(t_m)}{\sum_i \mu_i P_i(t_m)}$ (6)

   /* where: probability $\rho_i(t_m)$ that the social network is at state i, P(t): sample occupancy, μ represents probable mutation rate of messages across participants according to the proportion of occupation*/

4. Sample the next mutation time according to the cumulative probability P(t). This can be done via the

---

[2]It is simply a function on two variables, i and j which are integers, when for each social network instances cardinality of the variables is large, then it could assist to approximate inference based with a constrained, lower complexity, adaptively sized sum for the target cardinal value **[10]**.

inversion method, such that the next time $t_m = P -1(r)$, where r is a uniform random variable between 0 and 1.

5. Add t_m to the current time of occupancy of participants and betweenness.

6. Choose the specific score and plot according to their respective nore transition of the network and update the state of the system as per Step 4.

7. Remove extinct and redundant classes from the list and reduce the number N of classes accordingly.

8. Return to Step 1 until finished.
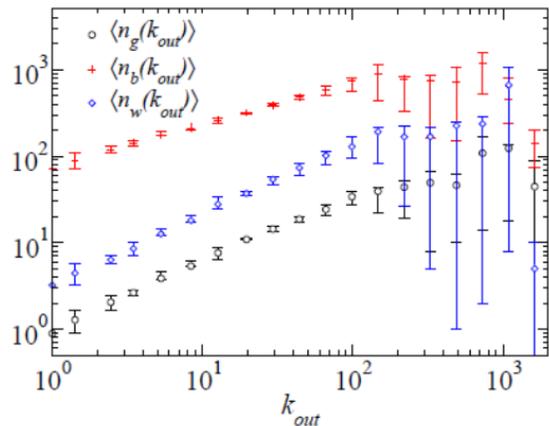
Fig. 3. Distribution of tagging with out degree (k_{out})
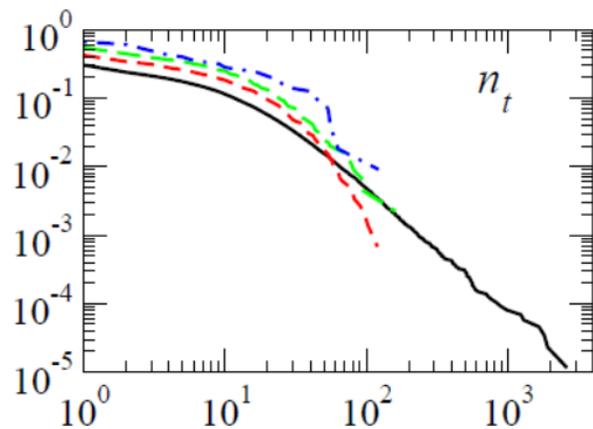
Fig. 4. Degree Correlation and distribution of Message tagging

In case of large scale network like Facebook, the most conventionally investigated mixing pattern involves the degree (number of neighbours) of nodes. This type of mixing improvises the likelihood; leading users with a given number of neighbours connect with users of similar degree. This property is emphasized by computing multi-point degree correlation functions. Complementary cumulative conditional distributions as mentioned with group tagging, specific message tagging and pre-defined choice tagging, compared with the global cumulative distributions denoted by black lines (Figure 4). Even among the subset of users with a given k_{out}, a strong disparity is still observed in the amount of activity and also around a specific community. Subsequently, the

occupancy on a specific graph instance results in the following plot: Log-log plot of the distribution of the contact durations and of the cumulated duration of all the contacts two individuals' m and n have over a day (w_{mn}). An interesting inference could be drawn that out of 88% of the total contacts sustained less than 1 minute on a specific tag or comments, but more than 0.2% persisted more than 5 minutes against a specific topic of interest. For the cumulated durations, 64% of the total duration of contacts between two individuals during one day last less than 2 minutes, but 9% last more than 10 minutes and 0.38% more than 1 hour. The small symbols reciprocate to the actual distributions, and the large symbols to the log-binned distributions [18] (Figure 5).
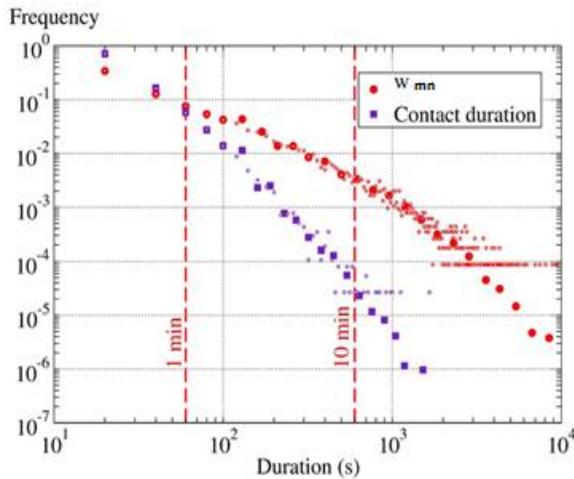
Fig. 5. Occupancy Graph for Social Network with Mutation Probability of Messages

Degree distributions of evolutionary dynamics and the empirical data: 'o' represents the data points from the simulated data and 'x' represents the data points from the empirical data new project The results of the degree distributions of the extended algorithm with dynamic values of mutation are shown in Figure 6. The upper figures are the comparison of developer and project degree distributions in linear coordinates.

The lower figures are the comparison of developer and project degree distributions in log-log coordinates. The R_2 of developer degree distribution from the simulated data in lower figure is 0.959 and the R_2 of project degree distribution from the simulated data in lower figure is 0.7657. Also the largest project size of the simulated data is just 1500. We can further lower this value by tuning the mutation parameter μ [17] and P (t),and graph function of the extended algorithm. H log represents homophily log evaluated from message exchanged towards any specific and common interest [17].
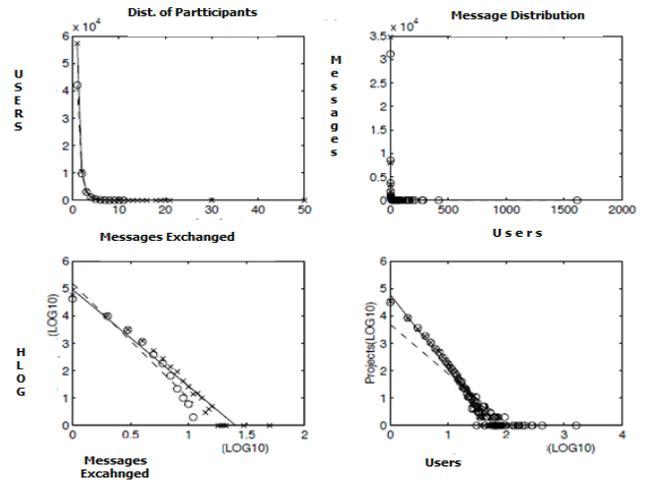
Fig. 6. Evolutionary Dynamics plot with varying mutation and Probability of occupancy
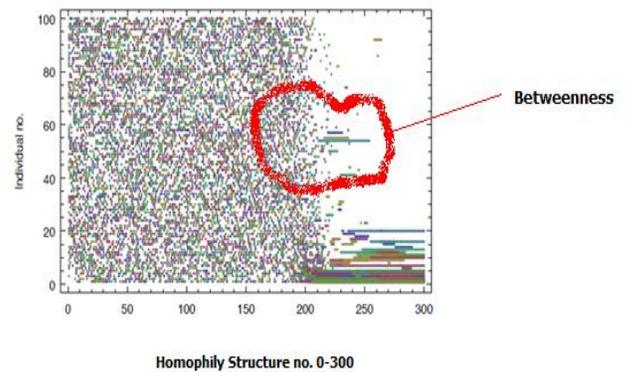
Fig. 7. Recursive simulation on test function towards Homophily from Betweenness

In order to quantify the complex large cardinal network, we incorporate classical test function like DeJong[3](The conventional De Jong's functions is the so-called sphere function, this function is unimodal and convex by nature)[16].Dimension is represented in the formulas by variable D, so as can be observed that it becomes simple to calculate selected functions for an arbitrary dimension for arbitrary number of participants. Similarly, the result of homphily structure from intial betweeness and other asscoited properties is safely turned out from local mimima due to the size of its popluation and avaiable best solution.The extended part of the algorithm demonstrates the frquencey of interplay and incentive distribution betwwen different communities played in social network.

$$1 + \sum_{i=1}^{D} \frac{x_i^2}{4000} - \prod_{i=1}^{D} \cos\left(\frac{x_i}{\sqrt{i}}\right)$$

represents Dejong Function

Algorithm III   Distribution in *Population graph*

The working strategy

Variables: N./*Homogeneous population of size*/

t /* Time Step t*/

r /* relative fitness */

/*temporal event could be Binomial time step[4] **(equation 8)** of propagation of base event under social network , if randomly chosen individual   from SN then fixation of probability of new mutant μ_1 represents probable mutation rate of messages across participants according to the proportion of occupation*/

$$\mu_1 = \frac{1-1/r}{1-1/r^N} \tag{7}$$

$$\mu_1 = \left[\frac{1-1/r}{1-1/r^N}\right][t_i + t^2{}_{i+1} + t^4{}_{2(i\ +1)} + \ldots + t^n{}_{2i!/N} \tag{8}$$

The distribution of trend on population graph of the new attributes concerning the people refers category of homophily structures (shown in red line) where the majority of people in the population have lower levels of the attribute. Gradually, as time elapses, the most influential message/ broadcast configuration propagates across the population (indicated by green line) but   subsequently; evolution dynamics pushes the higher level of the attribute with certain incentive. This is natural recipient of incentives under social network for those persons who are participating and interacting. In addition to, there are drifts of incentive distribution under different settings. Hence, eventually after a considerable period of evolutionary dynamic, drift increases at the high end of the attribute for homophilic structure and the distribution approaches reverts back again to normal.
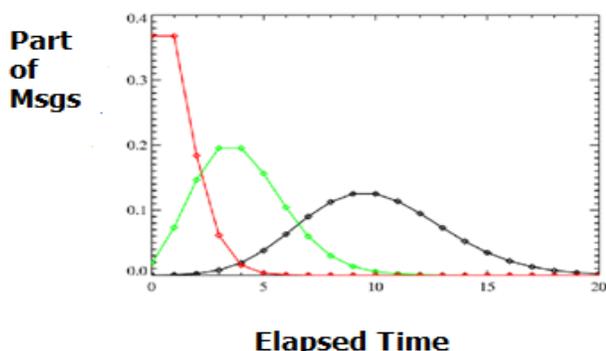
Fig. 8.   Time versus messages
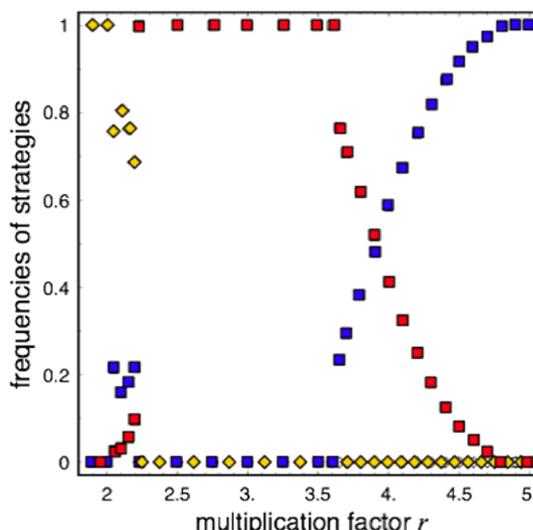
Fig. 9.   Simulation of Pay off distribution

At this stage of simulation, it was not evident   how the appropriate interaction could entail higher range of incentive among the participants. Therefore, a specific data set was chosen to exhibit different interplay among the social participants. The observations show that the frequency of cooperators (indicated by blue), deviators (indicated byred) and loners (represented as yellow) under smooth participation. This is also measured as the multiplication factor according to trust of discussion and interaction. Individuals are arranged on random regular graphs where each node has eight neighbors and they interact in randomly formed groups of size $N = 5$. For small multiplication factors,individuals dominate. The reason is simple: in this case even in a group of cooperators, the payoffs do not exceed the incentive of individual.   All of the proposed components of three algorithms improvise dynamic equilibrium considering the population of social network. The range of multiplication with trust increases the pay off a very small discrimination of*r*.And above the threshold deviators exist.. Only for much larger value of *r*~4.056 cooperators reappear and co-exist with deviators. Since individuals are absent, the dynamics again retain voluntary participation into compulsory interaction. Finally, for r= 4.6 cooperators take over and manage to displace deviators (**see Figure 9 and table 1).**

Interaction (Session 1)			Interaction (Session 2)			…….Interaction (Session n)		
	Incentive			Incentive			Incentive	
	Intial	Final		Intial	Final		Intial	Final
Group 1	10 Units	Ω	10 units	Ω		10 units	Ω+ donated value	
Group 2	10 Units	Ω	5/7/12 units	Ω+ intial value		6/7/15 units	Ω+ intial value	
Group n	10 Units	Ω	10 units	Ω		Ω	Not defined	

Table 1. Configuring design of Incentives on SN

The proposed design involves 4 stages and 4 groups described below and summarized in the following Table:

- Group 1: This group comprises of more than 65 subjects. The default initial value of the participating reward is 10 units and the final value could be measured as Ω.

- Group 2: This group represents almost 50 subjects with initial incentive but in subsequent sessions the experience differs with different fractions of incentives depending on the frequency of interaction as shown in Figure 8.

- Group n: The group enhances its subject line > 100, but in addition to the normal incentive distribution, there will be different treatments for incentive either there should be donation or additional amount the participants will push into it.

B. Comment on final Homophily Sturucure

The homphily evlalution from the social graph is completely based on the increase of posts, tags and other social network action artifacts. Keeping in mind about the non-linear aspects and growth strategy of socail network, evolutionary dynamics, comparatively better visualization of density of attributes has become possible. Especially, betweenness and graph occupancey and mutation could be the one of those key factors. Recent research of Tom A.B. Snijders and his colleagues [15] demonstrated the relation between evolutionary dynamics and homophily of social network in some common aspects of friendship, recommendation, group study and selection etc. **Figure 7** exhibits trend of precise visualization of homophily structure as guided by the extended algorithm with mutation as prime parameter. The following table 2 should be considered to understand the relationship shown in Figure 8: the red circle signifies higher cluster density and therefore providing the remaining count of homophily over the population.

Population variables	Mean	Std. Dev.	AutoCorr. St.No.	Homophily	Record Set tracking From snap.stanford.edu/data/
Male and Female	0.53	0.51	0.075	55	240000
Age	28.76	8.59	0.365	0-160	197000
Density of Propagation	24.60	13.21	0.35	160-240	230615
Network Variables					
Out Degree	*301.4*	*177.21*	0.15	58- 90	210043
Local Betweeness	*34.74*	*91.76*	0.017	170>	228456

VII. CONCLUSION AND FURTHER SCOPE OF RESEARCH

Affiliation being an important factor of homophily graph has a relevant role in promoting the focus through the nodes. The idea is how a friend in the Facebook or any social network can influence his friends to join the community he or she belongs to. This paper investigates such possibilities by exploring the homophily community in social network by using graph property based algorithm and simulation. Further, incorporation of evolutionary dynamics also contributed for investigating homophily property with better approximation. Subsequently, the growth of social network, temporal behavior and trend of it, can also be investigated by augmenting the existing evolutionary dynamics algorithm. On the other hand, the ranking algorithm or conventional classifier model can also be extended using initial attributes of embeddedness and graph traversal with graph occupancy time. Soft computing based (fuzzy and rough set) homophily identification from graph properties could be an emerging research on computational social network. Finally, in the extended analysis part, certain significant observations are made in terms of maximizing benefit or profit, based on their role at specific instances under social network. Experiments have incorporated certain public data sources to demonstrate mutual interplay of social network participants, their specific contribution towards the network and share of incentive if any.

REFERENCES

[1] Vojtek, P., &Bieliková, M. Homophily of neighborhood in Graph Relational Classifier. Pages 721–730 of: Geffert, V.,Karhumäki, J., Bertoni, A., Preneel, B., Návrat, P., &Bieliková, M. (eds), SOFSEM 2010: 36th Conf. on Current Trends in Theory and Practice of Computer Science, SpindleruvMlyn, Czech Republic, 2010, Proc. LNCS, vol. 5901. Springer.

[2] Halil Bisgin, Nitin Agarwal and Xia oweiXu, Investigating Homophily in Online Social Network, 2010,IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Web Technology, pp. 533-536.

[3] M. McPherson, L Smith Lovin and J.M. Cook, Birds of a feather: Homophily in Social Network, Annual Review of Sociology, 27(1): pp. 415-444, 2001.

[4] Eytan Bakshy, Winter A mason, Jake M.H .Ofman, Duncan J. watts, Everyone's an Influencer : Quantifying Influence on Twitter , published in Proceedings WSDM 2011, Hongkong, February 2011.

[5] Reza Zafarani, William D Cole, and Huan Liu. Sentiment Propagation in Social Networks A Case Study in LiveJournal. In Patricia Chai, Sun-Ki And Salerno, John And Mabry, editor, SBP 2010 - Advances in SocialComputing, Lecture Notes in Computer Science 6007, pp. 413–420. Springer Berlin, 2010.

[6] James H. Fowler, Jaime E. Settle, and Nicholas A. Christakis.Correlated genotypes in friendship networks.PNAS, pp. 108:1993, 2011.

[7] Cosma Rohilla Shalizi and Andrew C. Thomas. Homophily and Contagion Are Generically Confounded in Observational Social Network Studies. Sociological Methods and Research (arXiv:1004.4704v3), 2011.

[8] *Networks, Crowds, and Markets: Reasoning about a Highly Connected World.* By David Easley and Jon Kleinberg.Cambridge University Press, 2010.

[9] Benjamin Goluby Matthew O. Jackson, How Homophily Affects the Speed of Learning and Best-Response Dynamics, Q*uarterly Journal of Economics* Vol. 127, Iss. 3, pp 1287--1338, 2012.

[10] Chris Pal, Charles Sutton, Andrew McCallum. Constrained Kronecker Deltas for Fast Approximate Inference and Estimation.Submitted to UAI, 2005.

[11] Bryce Boe and Christo Wilson.User Interactions in Social Networks and their   Implications, accessableat http://cs.ucsb.edu/~bboe/public/pubs/interaction-eurosys09.pdfcrawl-e: Highly distributed web crawling framework written in python. Google Code, 2008.

[12] Chatterjee K, D Zufferey, MA Nowak (2012). Evolutionary game dynamics in populations with different learners.*JtheorBiol* 301: 161-173. http://dx.doi.org/10.1016/j.jtbi.2012.02.021

[13] Allen B, MA Nowak (2012). Evolutionary shift dynamics on a cycle.*JtheorBiol*   311:   28-39. http://dx.doi.org/10.1016/j.jtbi.2012.07.006.

[14] Survival of dominated strategies under evolutionary dynamics, Josef Hofbauer and William H. Sandholm,

Theoretical Economics 6 (2011), pp. 341–377.

[15] Alessandro Lomi, Tom A.B. Snijders, Christian E.G. Steglich, and Vanina Jasmine Torló (2011). Why Are Some More Peer Than Others? Evidence from a Longitudinal Study of Social Networks and Individual Academic Performance. *Social Science Research*, 40 (2011), 1506-1520.

[16] X.-S. Yang, Test problems in optimization, in: Engineering Optimization: An Introduction with Metaheuristic Applications (EdsXin-She Yang), John Wiley & Sons, 2010.

[17] Y. Gao,G. Madey, V. Freeh, Modeling and Simulation of a Complex Social   System: A   Case   Study,   accessible   at http://www3.nd.edu/~oss/Papers/SimulationSym37_gao.pdf

[18] J. Stehlé, and Et al, High-Resolution Measurements of Face-to-Face Contact Patterns in a Primary School.Published: Aug 16, 2011,DOI: 10.1371/journal.pone.0023176

[19] E. P. White, B. J. Enquist, and J. L. Green. Ecological On estimating the exponents of power-law frequency distributions.   Archives E089-052-A1.   2008.   Ecology   89:905-912.accessable   at http://www.esapubs.org/archive/ecol/E089/052/appendix-A.pdf

[20] Al-Qaheri, H, Banerjee , S., Ghosh, G. Evaluating the power of homophily and graph properties in Social Network: Measuring the flow of inspiring influence using evolutionary dynamics, Science and Information Conference (SAI), London, 2013 , IEEE XPlore, pp. 294 – 303.

[21] Himabindu Lakkaraju, Julian McAuley, Jure Leskovec What's in a name? Understanding the Interplay between Titles, Content, and Communities in Social Media, Association for the Advancement of Artificial Intelligence (2013

[22] Dohmen, T., and Falk, A., Performance Pay and   Multi-Dimensional Sorting: Productivity, Preferences and   Gender, American Economic Review, 101(2), pp. 556-590, 2011**.**

[23] DellaVigna, S., J. A. List, and Malmendier U.,   Testing for Altruism and Social Pressure in Charitable,   Giving, Quarterly Journal of Economics, 127(1), pp.1-56. 2012.

[24] J. Poncela Casasnovas, Evolutionary Games in Complex Topologies, Springer Theses DOI: 10.1007/978-3-642-30117-9, Springer-Verlag Berlin Heidelberg 2012.

[25] Sean Gilpin, Tina Eliassi-Rad and Ian Davidson , Guided Learning for Role Discovery (GLRD): Framework, Algorithms, and Applications KDD'13, August 11–14, 2013, Chicago, Illinois, USA.

[26] Sebastian Novak, K. Chatterjee and M.Nowak, Density games Journal of Theoretical Biology 334, pp.26–34, 2013.

[27] Himabindu Lakkaraju, Indrajit Bhattacharya, Chiranjib Bhattacharyya, Dynamic Multi-Relational Chinese Restaurant Process   for Analyzing Influences on Users in Social Media .IEEE International Conference on Data Mining (ICDM), 2012.

[28] Debanjan Mahata and Nitin Agarwal. Grouping the Similar among the Disconnected Bloggers, Social Network Analysis and Social Media Mining: Emerging Research. Guandong Xu and Lin Li (Eds.). IGI Global, 2012.

APPENDIX: PROOF OF STATEMENT IN EQUATION 8:

The proposed model is to trace   the time distribution mode in between different intermediate events. Considering the normal binomial distribution function, which specifies the number of times (x) that an event occurs in n independent trials where p is the probability of the event occurring in a single trial. Why we considered this trial? As, the exact probability distribution for any number of discrete trials may represent the number of mutant messages, therefore it becomes obvious that for any discrete time t, say for any temporal event the value of number of events could be large. Hence function becomes continuous. Thus relative fitness r also changes respective to bi-nominal distribution of mutant messages.

$$\text{Hence: } \mu_1 = [\frac{1-1/r}{1-1/r^N}][t_i + t^2_{i+1} + t^4_{2(i+1)} + \ldots + t^n_{2i!/N}$$

is a normal representation of r with bi-nomial time t

Generating a Domain Specific Checklist through an Adaptive Framework for Evaluating Social Networking Websites

Roobaea AlRoobaea

Faculty of Computing and
Information Technology,
Taif University, Saudi Arabia, &
School of Computing
Sciences,University of East Anglia
Norwich, UK

Ali H. Al-Badi

Department of Information Systems,
Sultan Qaboos University,
Oman

Pam J. Mayhew

School of Computing Sciences,
University of East Anglia, Norwich,
UK

Abstract—The growth of the Internet and related technologies has enabled the development of a new breed of dynamic websites and applications that are growing rapidly in use and that have had a great impact on many businesses. These websites need to be continuously evaluated and monitored to measure their efficiency and effectiveness, to assess user satisfaction, and ultimately to improve their quality.

The lack of an adaptive usability evaluation checklist for improvement of the usability assessment process for social network sites (SNSs) represents a missing piece in usability testing. This paper presents an adaptive Domain Specific Inspection (DSI) checklist as a tool for evaluating the usability of SNSs. The results show that the adaptive social network usability checklist helped evaluators to facilitate the evaluation process, and it helped website owners to choose the specific-context usability areas that they feel are important to their usability evaluations. Moreover, it was more efficient and effective than user testing and heuristics evaluation methods.

Keywords—Heuristic evaluation (HE); User Testing (UT); Domain Specific Inspection (DSI); social networks domain; social networks checklist

I. INTRODUCTION

It is clear that Heuristic Evaluation (HE) and User Testing (UT) are the most important traditional usability evaluation methods for ensuring system quality and usability [1; 2]. Currently, complex computer systems, mobile devices and their applications have made usability evaluation methods more critical; however, usability differs from one product to another depending on product characteristics. It is clear that users have become the most important factor impacting on the success of a product; if a product is produced and is then deemed not useful by the end-users, it is a failed product; nobody can use it and the company cannot make money [3]. [4] asserted, "companies are endeavoring to understand both user and product, by investigating the interactions between them".

Traditional usability measures of effectiveness, efficiency and satisfaction are not adequate for the new contexts of use [5]. HE has been claimed to be too general and too vague for evaluating new products and domains with different goals; HE can produce a large number of false positives, and it is unlikely to encompass all the usability attributes of user experience and design in modern interactive systems [6; 7]. UT has been claimed to be costly, time consuming, prone to missing consistency problems and subject to environmental factors [8]. To address these challenges, many frameworks and models have been published to update usability evaluation methods (UEMs) [9; 10]; however, these frameworks and models are not applicable to all domains because they were developed to deal with certain aspects of usability in certain areas [11].

The adaptive framework was originally constructed and then the DSI method and its checklist for educational domain was generated and evaluated against HE and UT methods [12; 13]. For further validation of the adaptive framework, social networks domain was chosen and then the DSI method was generated for SNSs and it evaluated against HE and UT methods [14]; in those experiences, the DSI method delivered interesting results by discovering more real usability problems in specific usability areas than HE method or UT method. An adaptive checklist based upon the DSI method for facilitating the social network sites evaluation process was developed. The main objective of this paper is to address the challenges that were raised and to present this checklist which can be applied to any website in the social network domain as a tool that can be used by designers, developers, instructors, and website owners to design an interactive interface or assess the quality of existing website. It also allows anyone to adopt any area of usability or any principle to determine the usability problems related to the seven specific areas in social network sites.

This paper is organized in the following way. Section 2 starts with a brief literature review, including a summary of the adaptive framework. Section 3 highlights the research methodology followed in this research. Section 4 presents a discussion of the findings. Section 5 presents the conclusion and future work.

II. LITERATURE REVIEW

A. Background and Motivation

The primary concern of interaction design is to develop interactive products or technologies that are usable. A website

is a product, and the quality of a product takes a significant amount of time and effort to develop. Web design is a key factor in determining the success of any website, and users should be the priority in the designers' eyes because usability problems in a website can have serious ramifications, over and above the users failing to meet their needs [15]. A high-quality product is one that provides all the main functions in a clear format, and that offers good accessibility and a simple layout to avoid users spending more time learning how to use it than satisfying their needing; these are the fundamentals of the 'usability' of a product. Poor product usability may have a negative impact on various aspects of the organization, and may not allow users to achieve their goals efficiently, effectively and with a sufficient degree of satisfaction. The website consultants and marketing sectors have understood that the number of hits, customer return rate, and customer satisfaction are extremely affected by the usability of a website [16].

The success of SNSs has gained a great deal of attention of researchers in latest years. Because, the impact of these sites on business is still largely unexplored. For example, impact SNSs on knowledge management (e.g. customer relationship management), collaboration, communication, innovation, and training [17; 18]. Currently, companies, educational systems and governments adopt SNSs tools to their environment work to save time, make money and to improve their corporate productivity.

In this regards, designing interactive websites and evaluating them are common stages of product development. On the other hand, the current traditional usability methods to measure quality attributes, such as, effectiveness, efficiency and satisfaction are not adequate for the new contexts of use, and are not stable in the modern dynamic environment such as SNSs and e-catalogs systems [19; 20]. Consequently, several studies have emphasized the importance of developing new kinds of usability evaluation methods and of constantly improving and making modifications to existing methods as a matter of priority, in order to increase their effectiveness [21]. Having extensively reviewed the existing literature on web usability evaluation methods; this research is unique in systematically constructing an adaptive framework that is applicable across numerous domains. This DSI framework generates DSI checklist as available tool for assessing and improving the usability of a product.

B. Description of the Adaptive Framework

The adaptive framework was developed according to an established methodology in HCI research [12; 22]. It consists of four development steps as follows:

Development Step One (D1: Familiarization): This stage starts by justifying the need to develop a method that is specific, productive, useful, usable, reliable and valid, which can be used to evaluate an interface design in the chosen domain. It entails reviewing all the published material in the area of UEMs but with a specific focus on knowledge of the chosen domain. Also, it seeks to identify an approach that would support developers and designers in thinking about their design from the intended end-users' perspective.

Development Step Two (D2: User Input): This stage consists of mini-user testing (task scenarios, think aloud protocol and questionnaire). Users are asked to perform a set of tasks on a typical domain website and then asked to fill out a questionnaire. The broad aim of this stage is to elicit feedback on a typical system from real users in order to appreciate the user perspective, to identify requirements and expectations and to learn from their errors. Understanding user needs has long been a key part of user design, and so this step directly benefits from including the advantages of user testing.

Development Step Three (D3: Expert Input): This stage aims to consider what resources are available for addressing the need. These resources, such as issues arising from the mini-user testing results and the literature review, require a discussion amongst experts (in the domain and/or usability) in order to obtain a broader understanding of the specifics of the prospective domain. Also, it entails garnering more information through conversations with expert evaluators to identify the areas/classification schemes of the usability problems related to the selected domain from the overall results. These areas provide designers and developers with insight into how interfaces can be designed to be effective, efficient and satisfying; they also support more uniform problem description and they can guide expert evaluators in finding real usability problems, thereby facilitating the evaluation process by judging each area and page in the target system.

Development Step Four (D4: Draw Up DSI: data analysis): The aim of this step is to analyse all the data gathered from the previous three. Then, the DSI method will be established (as guidelines or principles) in order to address each area of the selected domain.

C. Description of validation process for the Adaptive Framework

After constructing the DSI framework, the researchers test it intensively through rigorous validation methods to verify the extent to which it achieves the identified goals, needs and requirements that the method was originally developed to address. The validation process of the DSI checklist included analytical test, empirical test and statistical test. These tests were conducted using the newly developed DSI checklist alongside heuristics evaluation (HE), user testing (UT) and SPSS package.

III. RESEARCH METHODOLOGY

A. Evaluation of the Practicality of the Framework

In the first stage, the researchers conducted a literature review on the materials relating to usability and UEMs as well as on the requirements of social network sites (SNSs). In stage two, a mini-user testing session was conducted through a brief questionnaire that entailed four tasks, which were sent to ten users who are regular SNS users, to gain an appreciation of which elements or features they expect to be in any SNS, their more general expectations of these sites and to learn from their errors.

In stage three, a focus group discussion session was conducted with experts in usability and/or the SNS domain (i.e. single and double experts). Cohen's kappa coefficient was used on the same group twice to enable a calculation of the reliability quotient for identifying usability problem areas. In stage four, the researchers analysed the results of the three stages and incorporated the findings. The intra-observer test-retest using Cohen's kappa yielded a reliability value of 0.9, representing satisfactory agreement between the two rounds. After that, the usability problems areas were identified to facilitate the process of evaluation and analysis, and to help designers and programmers to identify the areas in their website that need improvement. Then, the DSI method was established. It is common for social networks as well as business networking websites to take into account what is called 'user experience'. The DSI method was classified according to the usability problem areas, and checklist was developed, as shown in Appendix A.

B. Selection of the targeted websites

The first step in an initial preparation phase is selecting the websites. The researchers sought to ensure that the selected websites would support the research goals and objectives. The selection process was criteria-based; five aspects were determined and verified for each website, and these are: 1) Good interface design, 2) Rich functionality, 3) Good representatives of the social network domain, 4) Not familiar to the users, 5) No change will occur before and during the actual evaluation. In order to achieve a high level of quality in this research, the researchers chose three well-known websites in this domain, which are LinkedIn, Google+ and Ecademy. All of these have all the aspects mentioned above.

C. Recruitment of Experts and Users

The selection of usability experts and users is the second important step in the initial preparation phase in this experiment. The researchers decided to recruit six expert evaluators, divided into two groups of three, who were carefully balanced in terms of experience. In each group, there are two double expert evaluators (usability specialists in SNSs) and one single expert evaluator (usability specialists in general). Each group employed two methods, namely DSI checklist and HE, to evaluate the three different websites. The evaluation was carried out in a prescribed sequence, i.e. Group 1 used DSI checklist on Google+ and then HE on LinkedIn, and finally DSI checklist on Ecademy, while Group 2 used HE on Google+ and Ecademy and then DSI checklist on LinkedIn. The researchers adopted this technique to avoid any bias in the results and also to avoid the risk of any expert reproducing his/her results in the second session through over-familiarity with one set of heuristics, i.e. each evaluation was conducted with a fresh frame of mind.

Selecting and recruiting users must be done carefully; the participants must reflect the real users of the targeted website because inappropriate users will lead to incorrect results, thereby invalidating the test. Appropriate users will deliver results that are more reliable; they will also be encouraged to conduct the experiment [30]. There is no agreement on how many users should be involved in usability testing. [30] suggested that 6 to 12 users are sufficient for testing, whereas other studies have recommended that 7, 15 and 20 users are the optimal numbers for evaluating small or large websites; particularly 20 users if benchmarking is needed [31]. At this point, 30 users were engaged; they were chosen carefully to reflect the real users of the targeted websites and were divided into three groups for each website, i.e. a total of 10 users for each website. The majority of the users are students and employees, and they were mixed across the three users groups in terms of gender, age, and education level and computer skills.

D. Piloting the Adaptive Checklist

A pilot study was conducted by two independent evaluators. They checked the adaptive DSI checklist by applying it in a real experiment to make sure that there were no spelling or grammatical errors and no ambiguous words or phrases, and that all of the sentences in the adaptive checklist were sufficiently clear to be used by the evaluators. A fewer minor improvements were made,

E. Actual Evaluation

The Heuristics Validation phase started with a training (familiarization) session for the six expert evaluators. They were given a UEM training pack that contained exactly the same information for both groups, except for the information pertaining to their respective UEM. The researchers emphasized to each evaluator groups that they should apply a lower threshold before reporting a problem in order to avoid misses in identifying real problems in the system. Then, the actual expert evaluation was conducted and the evaluators evaluated all websites consecutively, rating all the problems they found in a limited time (which was 90 minutes). After that, they were asked to submit their evaluation report and to complete a five-point scale on an SUS questionnaire (1 for strongly disagree and 5 for strongly agree) to rate their satisfaction on the evaluation method they had used (DSI checklist or HE), and to give feedback on their own evaluation results.

The Testing Validation phase started with a training (familiarization) session for the 60 users; it involved a quick introduction on the task designs, the think-aloud approach and the purpose of the study. The next step entailed explaining the environment and equipment, followed by a quick demonstration on how to 'think aloud' while performing the given tasks. Prior to the tests, the users were asked to read and sign the consent letter, and to fill out a demographic data form that concluded details such as level of computer skill. All the above steps took approximately ten minutes for each test session. The actual test started from this point, i.e. when the user was given the task scenario sheet and asked to read and then perform one task at a time. Once they had finished the session, they were asked to rate their satisfaction score relating to the tested website, to write down their comments and thoughts, and to explain any reaction that had been observed during the test, all in a feedback questionnaire. This was followed by a brief discussion session.

IV. DISCUSSION AND FINDINGS

The researchers extracted the problems discovered by the three methods from the problems sheet and removed all false

positive problems, subjective problems, and duplicated problems during the debriefing session. The problems agreed upon were merged into a unique master problem list (see Table 1, Table 2, and Table 3), and any problems upon which the evaluators disagreed were removed.

TABLE I.　TOTAL PROBLEMS FOUND (WITHOUT DUPLICATES) IN GOOGLE+

Method / Problem type	UT	HE	DSI checklist	Total problems
Catastrophic	4 (100%)	0 (0%)	0 (0%)	4
Major	9 (82%)	3 (27%)	11 (100%)	11
Minor	11 (37%)	13 (43%)	28 (93%)	30
Cosmetic	10 (37%)	6 (22%)	16 (59%)	27
No. of problems	34 (47%)	22 (31%)	55 (75%)	72

TABLE II.　TOTAL PROBLEMS FOUND (WITHOUT DUPLICATES) IN LINKEDIN

Method / Problem type	UT	HE	DSI checklist	Total problems
Catastrophic	2 (33%)	0 (0%)	6 (100%)	6
Major	5 (39%)	5 (39%)	11 (85%)	13
Minor	8 (32%)	8 (32%)	19 (76%)	25
Cosmetic	11 (92%)	0 (0%)	11(92%)	12
No. of problems	26 (46%)	13 (23%)	47 (84%)	56

TABLE III.　TOTAL PROBLEMS FOUND (WITHOUT DUPLICATES) IN ECADEMY

Method / Problem type	UT	HE	DSI checklist	Total problems
Catastrophic	0 (0%)	0 (0%)	0 (0%)	0
Major	3 (50%)	0 (0%)	6 (100%)	6
Minor	6 (50%)	8 (67%)	11 (92%)	12
Cosmetic	11 (37%)	4 (13%)	16 (53%)	30
No. of problems	19 (40%)	12 (25%)	33 (69%)	48

Overall, UT, HE and adaptive DSI checklist revealed different types and numbers of usability problems. One-way ANOVA reveals that there is significant difference between the three methods in terms of discovering usability problems on the whole (F = 13.32, p < 0.001). UT, HE and DSI revealed 47%, 31% and 75% of the usability problems found in Google+, respectively. One-way ANOVA-Tukey HSD was used and the results show that there is a strongly significant mean difference amongst the methods in finding usability problems in Google+ between HE and UT, where p < 0.03 and the mean difference = -14.667, as well as between DSI checklist and HE, where p < 0.003 and mean difference = -16.767. In LinkedIn, UT, HE and DSI checklist revealed 46%, 23% and 84% of the found usability problems, respectively. One-way ANOVA-Tukey HSD was used and the results show that there is a strongly significant difference amongst the methods in finding usability problems in LinkedIn, particular between HE and DSI checklist (p < 0.046 and mean difference = -14.333) and between HE and UT (p < 0.009 and mean difference = -15.367). Finally, UT, HE and DSI checklist revealed 50%, 32% and 87% of the found usability problems in Ecademy, respectively. One- way ANOVA-Tukey HSD was used and the results show that there is significant

difference amongst the methods in finding usability problems in Ecademy between HE and DSI checklist, where p = 0.012 and mean difference = -15.000. The performance of HE in discovering usability problems during the experiment ranged from 23% to 31%. UT discovered usability problems ranging from 40% to 47%, while DSI checklist discovered usability problems ranging from 69% to 84%. Also, UT and HE performed better in discovering major, minor and cosmetic real usability problems, but DSI checklist was the best in discovering more catastrophic, major, minor and cosmetic real usability problems. Thus, it can be seen that DSI checklist was the best in discovering real problems; this was followed by UT, and then finally HE.

Furthermore, each method revealed different types of problem (both unique and overlapping). For example, DSI checklist found 41% uniquely of the total number of real usability problems (n = 73 out of 176). HE found 14% uniquely of the total number of real usability problems (n = 24 out of 176), and UT identified 32% uniquely of the total number of real usability problems (n = 56 out of 176). 23 (13%) real problems out of 176 were found to be 'overlapping' by the three methods. In terms of the definition of missed problems given by [25], we can consider that the problems that were found by one method but not found by the others to be missed problems. From this point, DSI checklist missed 80 real usability problems; however, HE and UT missed 129 and 97 real usability problems, respectively. These findings should facilitate any decision-making with regard to which of these methods to employ, either on its own or in combination with another, in order to identify usability problems on social websites.

V.　CONCLUSION AND FUTURE WORK

The main aim of this experiment was to evaluate the adaptive DSI checklist for the social network websites through its ability to discover usability problems by comparing its results with usability testing (UT) and Heuristic Evaluation (HE). The adaptive DSI checklist was built based on the views of users and usability experts. It seemed to guide the evaluators' thoughts in judging the usability of the website through clear principles that include all aspects of the social networks' quality, which was represented in the seven usability areas.

Also, the DSI checklist outperformed both HE and UT, even when taken together. This finding facilitates decision-making with both regard to which of these methods to employ. Also, it addresses the shortcomings of these methods; hence, to avoid wasting money and time, an alternative method that is well-developed, context-specific and adaptive checklist to the situation in hand, such as what has been generated here and for the educational domain in [13], should be employed. This research contributes to the advancement of knowledge in the HCI field by introducing the adaptive DSI checklist that is specific for evaluating the social network websites. In order to consolidate and confirm the findings, future research could include testing the adaptive DSI checklist by applying it on different SNSs for example. Also, we need to further test the adaptive framework by developing an adaptive DSI checklist for different fields, such as e-commerce or news sites.

References

[1] G. Lindgaard and J. Chattratichart, "Usability testing: what have we overlooked?" in Proceedings of the SIGCHI conference on Human Factors in Computing Systems, pp. 1415-1424, ACM, 2007.

[2] J. Chattratichart and J. Brodie, "Applying user testing data to UEM performance metrics", in CHI'04 extended abstracts on Human Factors in Computing Systems, pp. 1119-1122, ACM, 2004.

[3] J. Nielsen, "Did poor usability kill e-commerce", in www.useit.com, 2001.

[4] F. Nayebi, J.-M. Desharnais and A. Abran, "The state of the art of mobile application usability evaluation", in Electrical & Computer Engineering (CCECE), 2012 25th IEEE Canadian Conference on, pp. 1-4, IEEE, 2012.

[5] P. Zaharias and A. Poylymenakou, "Developing a usability evaluation method for e-learning applications: Beyond functional usability", Intl. Journal of Human-Computer Interaction, vol. 25, no. 1, pp. 75-98, 2009.

[6] M. Hertzum and N. Jacobsen, "The evaluator effect: A chilling fact about usability evaluation methods", International Journal of Human-Computer Interaction, vol. 13, no. 4, pp. 421-443, 2001.

[7] J. Chattratichart and G. Lindgaard, "A comparative evaluation of heuristic-based usability inspection methods," in CHI, vol. 8, pp. 05-10, 2008.

[8] A. Oztekin, Z. J. Kong and O. Uysal, "Uselearn: A novel checklist and usability evaluation method for e-learning systems by criticality metric analysis", International Journal of Industrial Ergonomics, vol. 40, no. 4, pp. 455-469, 2010.

[9] N. Alias, S. Siraj, D. DeWitt, M. Attaran and A. B. Nordin, "Evaluation on the usability of physics module in a secondary school in Malaysia: Students retrospective", The Malaysian Online Journal of Educational Technology, p. 44, 2013.

[10] C. Gutwin and S. Greenberg, "The mechanics of collaboration: Developing low cost usability evaluation methods for shared workspaces", in Enabling Technologies: Infrastructure for Collaborative Enterprises, 2000 (WET ICE 2000). Proceedings IEEE 9th International Workshops, pp. 98-103, IEEE, 2000.

[11] C. K. Coursaris and D. J. Kim, "A meta-analytical review of empirical mobile usability studies", Journal of Usability Studies, vol. 6, no. 3, pp. 117-171, 2011.

[12] R. AlRoobaea, A. Al-Badi, P. Mayhew. " Generating a Domain Specific Inspection Evaluation Method through an Adaptive Framework: A Comparative Study on Educational Websites. " International Journal of Human Computer Interaction (IJHCI), 4(2), 88, 2013.

[13] R. AlRoobaea, A. Al-Badi, P. Mayhew, "Generating an Educational Domain Checklist through an Adaptive Framework for Evaluating Educational Systems." International Journal of Advanced Computer Science and Applications(IJACSA), 4(8), 2013.

[14] R. AlRoobaea, A. Al-Badi, P. Mayhew. "A framework for generating a domain specific inspection evaluation method: A comparative study on social networking websites." In Science and Information Conference (SAI), 2013, pp. 757-767. IEEE, 2013.

[15] S. Y. Chen and R. D. Macredie, "The assessment of usability of electronic shopping: A heuristic evaluation", International Journal of Information Management, vol. 25, no. 6, pp. 516-532, 2005.

[16] Y. Rogers, H.Sharp, and J.Preece," Interaction design: beyond human-computer interaction," Wiley, 2007.

[17] F. Bonchi, C. Castillo, A. Gionis, and A. Jaimes," Social network analysis and mining for business applications," ACM Transactions on Intelligent Systems and Technology (TIST), 2(3), 22, 2011.

[18] S. Andriole, "Business impact of Web 2.0 technologies," Communications of the ACM, 53(12), 67-79, 2010.

[19] A. Thompson and E. Kemp, "Web 2.0: extending the framework for heuristic evaluation", in Proceedings of the 10th International Conference NZ Chapter of the ACM's Special Interest Group on Human-Computer Interaction, pp. 29-36, ACM, 2009.

[20] A. Alrobai, R. AlRoobaea, A. Al-Badi, P. Mayhew. "Investigating the usability of e-catalogue systems: modified heuristics vs. user testing." Journal of Technology Research, 2012.

[21] Y. Guo, R. Proctor and G. Salvendy,"A conceptual model of the axiomatic usability evaluation method," Human Interface and the Management of Information, Interacting with Information, pp. 93-102, 2011.

[22] R. AlRoobaea, A. Al-Badi, P. Mayhew , "Generating a Domain Specific Inspection Evaluation Method through an Adaptive Framework." International Journal of Advanced Computer Science and Applications(IJACSA), 4(6), 2013.

[23] J. Dumas, and J. Redish, A practical guide to usability testing. Intellect Ltd, 1999.

[24] J. Nielsen, and H. Loranger, Prioritizing web usability, New Riders Press, Thousand Oaks, CA, USA,2006.

[25] L. Masip, T. Granollers and M. Oliva, "A heuristic evaluation experiment to validate the new set of usability heuristics," In Proceedings of the 2011 Eighth International Conference on Information Technology: New Generations, pages 429-434. IEEE Computer Society,2011.

APPENDIX A: THE ADAPTIVE DOMAIN SPECIFIC INSPECTION CHECKLIST FOR EVALUATING SOCIAL NETWORK WEBSITES USABILITY

Usability problem area	The adaptive Domain Specific Inspection (DSI) checklist
Layout and formatting (LF)	Design consistency: o Are all links and button styles throughout the site consistent? o Are all the pages organized /structured in a similar style? o Are the font choices, colours and sizes consistent with good user screen design? o Is the navigation of the site consistent? o Does site has access to Home, Contact Us and other relevant information link on all the pages?
	Simple user interface: o Does the site provide brief, constructive, unambiguous descriptions of the task? o Are the most important items in a list placed at the top? o Does site have search & help option? o Does the site use minimal page scrolling (i.e. the pages are not too long)? o Does the site highlight important changes (i.e. most viewed, most discussed, favourite feeds and recent updates)? o Does the site use glyphs and icons (metaphors) for representation and recognition in a context that is relevant, and not just for decoration? o Does the site use alternative text for the graphics/images? o Does the site categorize content into primary (absolutely necessary to show) and secondary (can be hidden), and show secondary information only on user demand? o Is the site layout, and architecture logical and hierarchical o Does the colour scheme override the content (undesirable)? o Is the site easily readable? o Does the site make important keys larger than other keys? o Are pages easy to bookmark? Is it possible to bookmark a person? o Is a casual user able to return to using the site after some period without having to learn everything all over again? Are all functions and information well-presented and easy to remember? o Is the screen layout efficient and visually pleasing? o Does the site provide the minimum number of clickable actions, selections and scrolling to complete one main task? o Is the site constantly used pop-up windows? o Can users switch between windows during overlapping windows? o Are users allowed to move backward, forward and skip data entry screens among all the pages? o Do all pages have a title? o Does the site helps user to pre populate data during registration, search etc.?
Content quality (CQ)	Correct, relevant, up to date and reliable information: o Is the content updated frequently, the last update statement being displayed in a prominent place? o Does the site display only information that is relevant for its purposes? o Does the site display only the available content, and is the content suitable to the page length? o Does the site provide concise and non-repetitive information? o Is there a link provided to the homepage? Was the site built by a reliable institution? o Are the reliability, stability and continuity of the site content guaranteed?
	Error-free: o Are errors, confirmation, and prompt messages displayed consistently throughout the site? o Is the site free of typographical errors and spelling mistakes? o Do error messages prevent potential errors from happening? o Does the site provide solutions that help the user avoid errors, such as providing 'undo' and 'redo' features? o Can errors be averted or minimized when possible? o Can corrective action be taken to rectify errors? o Are the details of the error messages available with indication to what actions are that users need to take to correct the error?
	Representation with familiar terminology & understandable content: o Is the content readable, scannable and easy to understand? o Do the content blocks need to be visually separated? o Are the vocabulary and terminology used familiar to users? o Does the site provide correct spelling and grammar, and understandable graphic symbols?
	Appropriate & approachable content: o Is the organization of the content suitable for achieving the primary goals of the site? o Are users provisioned with FAQ? o Does the site offer an appropriate amount of information for the page length, and is all the text of a viewable/readable size? o Does the site provide an icon for help next to a field? o Does the site show error in different colour and layout to read easily?
	Site upload time & memory utilization: o Is the site upload-time reasonable? o Is the site free from heavy coding /unwanted scripting which could consume more time/memory? o Are response period suitable to the member's cognitive processing? o Are response period suitable to each task?

Security and privacy (SP)	Awareness of security mechanism/settings & protection: o Are sensitive areas of the site protected against hackers by credentials and SSL security (e.g.,VeriSign™)? o Is it easy to change privacy and security settings? o Does the site protect customers' personal data adequately? o Can the uploaded content still be displayed outside the site if the user decides not to permit it (undesirable)?? o Is the adult content accessible to anyone without asking them to declare whether or not they are over 18 (undesirable)?? o Are users who are over 18 allowed to solicit personal information from under 18s (undesirable)? o Are all protected areas wholly inaccessible? o Does site has taken adequate measure of penetration testing to improve the security? o *Does site displays what are the security measure has been taken care to the user?* o Does site support industry defined standards like OWSAP, W3C.
	Transparency of transactions: o Is the adopted security mechanism and policy clearly displayed? o Does the site provide transparency of transactions and data use to build user confidence and trust, unless the user gives a clear indication not to expose it? o Are links to 'privacy policies' and 'terms & conditions' clearly displayed? o Is it clearly stated that any data submitted will not be used for other purposes, in order to build user confidence and trust? o Are there processes in place to check the number of memberships or access statistic data? o Should users upload, post, email, transmit or otherwise make available any content that is unlawful, harmful, pornographic and racial, do other users have the option to report any suspicious activity or inappropriate content that breaches the terms of service directly to the customer service or site manager? o Does site provide details like what is the user's information are going to be stored? o Does site declares about sharing the user's information to 3rd party for any purpose? o Does site informs user to contact for promotion, marketing and others such communication?
Business support (BS)	Advertising or sales pitches mechanism: o Is the advertising experience on the site too intrusive, disturbing the user's primary actions? o Does the site have pop-up advertisements (undesirable)? o Does 'multimedia help' make advertising enjoyable/attractive? o Can users leave comments and "likes" (these are social media terms)? o Can users classify advertisements easily? o Do the features of the paid membership are clearly described with giving hot offers?
	Trust & credibility of information sources and company advertising: o Is the user interested in the advertisement characters because they are drawn from the user's own culture? o Does the user have confidence that the site is operating in the way it was designed to?
	Easy to follow & share: o Can users share the content easily (text and links)? o Are the videos and photos easy to upload, download, share, retrieve and organise? o Can users share (i.e. post to friends' profiles) and tag other members in photographs and videos. o Are users able to access each other's profile information? o Are users allowed to share their content with other SNS services?
	Forum/blog facilities and connectivity with different groups/businesses: o Do users become engaged with the site through a set of facilities that are designed to promote engagement (e.g. by creating a group, blog, business)? o Will information posted on users' walls appear on their fans' walls? o Is it easy to create polls, pages and forums? o Are blogs and forums used to get ideas about markets, customers, and strategies? o Is it easy to use site mail to communicate with friends? o Is it allowed to make free calls between computers and/or phones? o Is it easy to create events or select widgets using a calendar? o Can users join regional, educational or workplace networks? o Do websites use 'crowdsourcing' approach to stimulate innovation, solving problem and sharing knowledge?
	Syndication of Web content (such as RSS tools): o Is there a news feed on users' home pages that provides them with friends/ company activity updates? o Can users publish RSS feeds to their profiles? o Are RSS filters used to create content streams to improve customer relationship management?
	Frequent posting & updating: o Are interactive tools such as post text, single chat and multiple chat provided? o Is it easy to modify, update and remove posts? o Can the users participate as much as they want?
User usability, sociability and management activities	Manageable personal profile & user-driven content: o Is it easy to register on the site? o In case of theft and/or a forgotten password, is recovery option available? o Can customers personalise (customise) their online workplace? o Can users edit/delete the content that they have posted? o Can users easily collect and access the content that they have found and liked/marked as favourite? o Can users create and modify their personal profile, and delete it if necessary? o Reporting mechanism: Can users report content that they may have a problem with (such as sexual, religious, illegal, etc.) easily? o Can the network delete a content that has received a lot of complaints? o Can the user manage all the activities pertaining to the site with ease, and have overall control? o Are items logically labelled and grouped in a control panel?

	Easy functionality, participation & user privileges, such as revoking & accepting friends/connections: ○ Private messaging: Can users who are directly connected chat/ message each other in a private conversation? ○ Public messaging: Can users broadcast and share messages with other users with whom they are directly connected? ○ Is it easy to accept new friends and blocking unwanted friends/connections? ○ Can users choose who they want to be directly connected to? This should be a two way agreement - where both users approve of the connection. ○ Can a conversation take place between more than just 2 users? ○ Can users register a group or book or band? Can they create a fan club for a band?
	Supporter of users' skills & freedom, such as the customization of users' content/messaging and notifications: ○ Does the site allow the user to initiate actions? ○ Can users create their own templates or page graphics? ○ Are there enough options for organising page layout or templates? ○ Can users choose a number of applications to be displayed on their profile page? ○ Does a website use e-mail notifications to encourage members? ○ Does site provides customisation based on users choice?
	Offers of informative feedback - action & reaction: ○ Is there confirmation for each action? ○ Is feedback given in proportion to the action performed (not too much and not too little)? ○ Are errors conveyed in context and written in a way that users will understand? ○ Does the site provide an overview of the work process that has been completed by the user (e.g. completing a user's profile)? ○ Is the feedback given at any specific time tailored to the content or problem being studied by the user? ○ Does the site feedback provide the user with meaningful information concerning their current level of achievement within the program? ○ Is the message of current status related to the user's task? ○ Does the site program provide the user with opportunities to access extended feedback from instructors through email and internet communication, and are adequate FAQs also offered? ○ Does the performance support tools provided mimic their real–world counterparts?
	Appropriate multimedia with complete user control: ○ Are the videos and images on the site of high quality, with the inclusion of alternative text for visually impaired people? ○ Can users change video, audio and image settings easily? ○ Is a mechanism provided to skip/stop animation and video without disruption? ○ Does the site include sound and visual effects, these effects providing meaningful feedback or hints, designed perhaps to stir particular emotions? ○ Does the site include surprises, humour and interesting representations for the user, while avoiding unnecessary multimedia representations that could confuse a user who has just started to work with the site? ○ Is there unnecessary animation and 'flash' on the site (undesirable)? ○ Is it easy for users to set up their own channels (e.g. YouTube channels)? ○ Are video ratings and comments available on the site? ○ Can users modify photo, audio and video submissions? ○ Are users allowed to play videos outside the site (e.g. YouTube) which would mean that they could be 'embedded' into other websites?
Accessibility and compatibility	Accessibility and compatibility of hardware devices: ○ Is the site compatible with various platforms and hardware, and can its features be adapted to individual user preferences? ○ Do potential users have to have special computer skills to be able to use site? ○ Are all the input devices/buttons that have no function disabled to prevent user-input errors? ○ Are the lessons accessible to users with physical impairments, and their contents available in various languages? ○ Does the site is properly load tested and support agreed number of users at a time. ○ Does the site have proper Disaster Recovery in place? ○ Does the site is supported by text reader or other such devices?
	Accessible path-contact details, help and support: ○ Is a site map and /or table of contents available, as well as a calendar? ○ Is there accessible and appropriate help available on demand? ○ Does the site provide clear contact details, using multiple contact formats (email, forms, etc.)? ○ Is the FAQ page easy to find? ○ Is everything on the site clearly understandable by the user, including how to access options for additional guidance (chatting, editing, adding, seeking instruction or other forms of assistance) when needed? ○ Does user allowed to resume work where they left off after getting help? ○ Does the performance of the site is satisfactory and it loads most of the content in less than a second?
	Easy access through universal design: ○ Has a universal design been implemented to cater for diversified user groups? ○ Is the structure too tight (strangling) or too loose (lacking cohesion), both of which are undesirable?
Navigation site and search quality	Correct & reliable navigation/directions: ○ Do all links and buttons lead to the correct location? ○ Does the site provide a breadcrumb (cookie crumb trail) to identify the path to the current location? ○ Does the site match the menu structure to the task structure, and can the user distinguish between options and content on the pages?
	Easy identification of links and menus:

	o Are the navigation objects and tools placed in consistent, clearly defined positions, and are they of an adequate size? o Are icons and links labelled? o Is an item still visible when it should be hidden from view, and vice versa? o Are the menus straightforward and easy to understand, the items being logically grouped and labelled? Do buttons, links and features have a 'mouseover' or pop-up window that provides meaningful feedback?
	Search support & functionality: o Are the functionality of buttons and controls obvious from their labels or from their design? o Are there clearly visible search buttons and search input fields consistently placed across all pages? o Are there live search results and filtering? o Does site help to auto fill the search query? o Does the search response are fast enough? o Are the results of searches clear, visible, informative, advisable and relevant? o Does the site support different search criteria (e.g. groups, people, interests, content, suggestions, and companies)? o Does the results page show the user what was searched for, and is it easy to edit and resubmit the search? o Are all the necessary functions of the site available without having to leave the site, and do they work correctly? o Are all the functions clearly labelled, thus facilitating successful completion of the task? Is the status of each task made clear on every page? o Is the search engine accurate? o Does the site support onsite searches within country/region, language, interests, industry, keyword videos, channels, play lists, and groups? o Can the moderated or restricted content be viewed by members with "SafeSearch" switched on?

Sociomaterial analysis of Music Notation Lessons: Virtual work and digital materialities

Demosthenes Akoumianakis

Department of Informatics Engineering, Technological Education Institution of Crete (TEI Crete)
Heraklion, Crete, Greece

Abstract—The present research rests and elaborates on sociomaterial aspects of virtual practices, as manifested through distributed and collaborative work. This is approached through an interpretive case study of music notation lessons (MNLs) using the DIAMOUSES system. Our empirical data suggest that sociomateriality shifts the focus of designing interactive technologies from mere considerations of digital manifestation (i.e., forms of representation) towards explicit accounts of the representational practices (i.e., the particular material properties of these forms) and the quality attributes to be embedded in technology.

Keywords—Virtual work, affordances, design qualities, case study research

I. INTRODUCTION

Sociomateriality is a recent construct aiming to explain the relationship between the social and the material across work settings and organizational contexts. Such an intriguing challenge turns out to be difficult due to the lack of sufficient ground to anchor the varieties of 'agencies' implicated in the social and material realities of organizational life. Recent management and organization science scholarship seeks for theoretical insights by recapitulating constructs such as social practice [1] [3], affordances [2] and virtual team work [4] [26]. Information Systems (IS) researchers [5] [6] [9] explore and classify facets of sociomateriality. The Computer-Supported Cooperative Work (CSCW) community is also engaged in debates about the configurations of the social and material [7], the way in which new digital technologies establish new materialities [8] and the techniques which may offer useful insights [10]. In most of these studies, sociomateriality rests on the recognition that technologies, people and organizations are constitutively entangled, rather than a priori self-contained entities, distinctively and separately studied [3]. Nevertheless, this notion of 'entanglement' may not offer an appropriate lens for design. Instead, as Leonardi [4] [22] convincingly argues 'imbrications' may serve better design-oriented thinking, especially in fields such Human-Computer Interaction (HCI) and CSCW, which constitute the venues motivating the present research.

The present work aims to contribute to the on-going debate about digital materiality [8] and the way in which it is crafted and implicated into social / organizational practices. The normative perspective adopted is that digital materiality anchors new practices which do not follow inexorably from the material features of established technologies; instead, they are improvised on the basis of old practices that work

differently in new technological circumstances and frequently lead to changes in social configurations [11], [12]. In this vein, our empirical ground stems from on-going work in collaborative music practicing. Such a focus should complement other research works making claims about how social media, networking platforms, blogging and micro-blogging services drive cultural shifts in the way people socialize online [13], use language [14] and organize offline arrangements, such as vacations and Network Music Performance (NMP) [15], [16]. The present work is seen as complementary to these efforts in so far as it offers insights into intrinsic properties of technology that bring about, enable or constrain these changes to the volume and extent that justifies cultural shifts or changes in social practice. Recent scholarly works acknowledge the need for better understanding the IT artefact and call for challenging the prolonged concerns with immaterial properties of information systems [17], [8]. They also point to a direction for research to advance a theory of digital materiality, thereby revisiting our understanding of computer-mediated artefacts and the ways in which they invoke material concerns of whatever practice is engaged.

In light of the above the specific aim of this paper is two-fold. Firstly, it sets out to explore and untangle features embedded in digital technologies such as abstraction, modularity, persistence, sharing, connectivity, etc., whose presence or absence may determine what people do (individually or collectively) and the respective enacted phenomena. Secondly, it seeks to establish a preliminary empirical ground by examining these features in a particular virtual work setting – that of online music notation lessons (MNLs) – and the digital artefacts involved. Phrased differently, our goal is to investigate sociomaterial aspects of computer-mediated music practicing and the intrinsic properties, such as representations, tools and software quality attributes, that shape and augment the practice in virtual settings. The approach builds on two threads of research; on the one hand, it reflects upon recent empirical findings about online music ensembles [10] and on the other hand, it transforms these findings into design proposals anchoring distributed music making as virtual work [18]. The main contributions are to be found not so much in the theoretical treatment of the concepts involved, as on the (design-oriented) mechanics that may determine what is possible under certain circumstances and the methodological challenges confronting HCI researchers interested in understanding how new technologies enable or constrain what people do online.

The rest of the paper is structured as follows. The next section motivates the present work by reviewing relevant scholarships and establishing the research focus. Then, we present the methodology for studying a specific NMP scenario, namely music notation lessons. The approach entails revisiting an existing research setting through an interpretive case study and instruments that bring to the surface sociomaterial considerations. The paper is concluded with an outline of implications and ongoing and future research.

II. THEORETICAL MOTIVATION AND RELATED WORK

Despite wide acknowledgements of the new primary beneficiaries established by new technologies [19] [20], it is not yet clear exactly what may be the features embodied in these technologies (material) and how these implicate novel (social) practices. This is the rationale for bringing sociomateriality at the forefront to assess not only the concept's theoretical underpinnings but also some of its implications for designers. To this effect and by elaborating on different theoretical literatures, such as digital materiality [17] [8], imbrications [22] [23], practice theory [21], remediation [24] and digital assemblages [25], the present research seeks to establish its theoretical footing.

Attempting a broad classification, it may be claimed that current thinking on sociomateriality emphasizes two metaphors coined by the 'entanglement' of agencies and the 'imbrication' of structures. Both perspectives share common ground, but they lead to slightly different insights. For instance, they both advocate the notion of 'affordances', initially introduced by ecological physiologist James Gibson [31], to anchor the material grounds of either computer-mediated work or daily activities [2], [22]. However, entanglement favours certain (perceivable) product-oriented affordances, while the imbrications perspective emphasizes design-embedded affordances (which may not be easily perceived at first sight). It is also common to relate sociomateriality with remediation [24] and the notion of digital assemblages [25], but there are various forms of remediation that may not implicate new digital assemblages.

In such a broad setting, our current effort adopts a design –oriented perspective focusing on digital artefacts, their transformative capacity as well as the way in which they are inscribed into different technological configurations and settings. As artefacts we consider not only the digital manifestation of objects but also the material qualities through which these objects become embedded into organizational life. This perspective allows us to theorize about the 'cultural' history of artefacts in terms of media-specific representations and affordances as well as to foresee the tactics through which they are remediated to ascribe (new) material agency to certain technologies.

A. Affordances and virtual work

The concept of affordances was introduced by Gibson to coin opportunities for perception and action offered by the environment to an organism, whether human or not [31]. Norman [32] suggested a slightly different interpretation focusing on '…perceived and actual properties of the thing, primarily those fundamental properties that determine just how the thing could possibly be used (p.9)'. In subsequent works, Norman [33] acknowledged misuse of the term and made a further clarification, stating that '… it is very important to distinguish real from perceived affordances' and that '… design is about both, but the perceived affordances are what determine usability…' (p. 123). Despite differences in orientation, Gibson and Norman recognize that material properties of physical objects determine the actions each object invites for or constrains. Accordingly, agency is critical since possibilities of action are not given. In a slightly different vein, Gaver [34] by-passes the issue of agency claiming that it is the design that suggests an affordance and therefore, affordances 'can be designed and analyzed in their own terms.' (p. 81). According to this view, affordances of computer-mediated environments can be related to design qualities such as abstraction, translucence, interoperability, connectivity and plasticity. Then, it may be argued that it is the presence or absence but also the intertwining of these design qualities that enables or constrains use of computer-mediated artifacts as well as the part of such use that is retained and made available for exploration. Arguably, this is useful but in need of further refinement if it is to provide a basis for designing systems.

Following a different strand and building on semiotics, Bailey, Leonardi & Barley [26] have advanced a conceptual lens that is useful for understanding virtual work in terms of the intrinsic properties of computer-mediated practices and affordances. The authors distinguish between digitization and virtuality by anchoring their implications. Thus, they claim that digitization implicates computer-based representations of physical phenomena, while virtuality occurs when digital representations stand for, and in some cases completely substitute for, the physical objects, processes, or people they represent. Building on this distinction, the authors qualify computer-mediated work as virtual teams, remote control or simulations, while in each case they classify the human activities involved as operations *with* or *on*, and *within* or *through* representations. By this account Bailey, Leonardi & Barley resurface the notion of affordances to assess how certain representations (and the practices they implicate) obtain material features through the way in which they become embedded into whatever technology is at hand.

B. Remediation

The concept of remediation was only recently introduced in the information systems literature in an effort to assess how certain practices are aligned and re-aligned to certain media. Lanzara [24] defines remediation as a migration of an assemblage of embedded agencies established in a certain medium to a new assemblage in a new or multiple media. Lanzara's [24] also acknowledges that 'the more deeply embedded is the practice in a specific medium, the greater the amount of restructuring involved in the migration to a different medium'. At core what is called for is an assessment of the degree of embeddedness of practices into

certain media through accounts of intrinsic constructs such as representation of meaning, means of transmission and synergistic use of media. There are practices, such as painting and music composition that tend to avoid detail and value abstraction. They also tend to embrace ambiguity as enabling richness of meaning and bring into play our intuition and imagination that determine how the practice is enacted. These practices – frequently referred to as creative practices – explore visual, spatial, textural and audio representations that afford abstraction, rich meanings and interpretation. In contrast, there are technical practices that aim to eliminate ambiguity, seek for certainty and pursue correctness, completeness and detail. They rely on formalism, symbolic representations and logical reasoning.

It then stands to ask 'How are practices remediated (using technology) and what effects such remediation may bring about?' Clearly, new technologies and media can have variable effects on practices. For instance, [14] examines how a typographic convention, the hashtag, operates as a linguistic marker, thus rendering the language searchable and driving a cultural shift for electronic discourse from online conversation to such 'searchable talk'. Similarly, Dourish and Mazmanian [8] describe how digital photography and the material constitution of digital images allow for forms of manipulation quite different from those that film photography affords. These examples are indicative not only of the scope of remediation but also of its implications that may vary from mere improvements in the conduct of practices to establishing totally new practices whose domain of discourse is anchored on new grounds [16].

C. Practice theories

Attempting to establish some sort of benchmarks or thresholds to understand intended and unintended effects of remediation, practice-oriented thinking and theories [21], [27] – despite their different orientations – offer a potentially useful roadmap. Of particular relevance to the present work is the notion of the practice lens [27], [28] as a metaphor for analytical inquiries. According to the practice lens, technologies can be seen as prerequisites for particular outcomes but the existence of prerequisites does not determine the outcome. Thus, by appropriating features inscribed in technology it is possible to enact new structures which were not initially foreseen during the development of the technology. Such a line of thinking is consistent with Pickering's concept of 'temporarily emergent' activities [30] and Suchman's notion of 'socio-materiality' of practice [7].

For our purposes, practice-based theorizing translates to a rather distinct line of argumentation. Specifically, it is argued that although functionality embodied in technological artefacts is clearly important, it is not likely to be, by itself, the sole determinant of how the artefact is used in practice. Instead, the presence or absence of non-functional design qualities embedded in technology can catalyze and determine technology use [29]. This is evidenced when examining what constitute embedded qualities in established and emerging technological paradigms. Specifically, technological transitions such as the evolution from the WWW, to Web 2.0

and Social Semantic Web as well as the associated practices (i.e., from device-dependent to device independent mark-ups, from authoring to collaborative editing and from consuming information to social networking and end-user content creation) can be explained by the increasing embeddedness of qualities, such as abstraction, portability and interoperability, in technological artefacts and tools [10].

D. Consolidation and reasearch questions

Attempting to synthesize the discussion thus far, it is argued that three issues stand out very prominently. Firstly, practices comprise activities on digital artifacts, which in turn, are representations of objects (human and non-human). Secondly, the affordances of these representations determine by and large the possibilities and effects of what is possible in a certain medium as well as the viability of remediation (i.e., shifting the practice form a certain medium to another). Thirdly, the material concerns invoked by digital technologies stem (at least in part) from the presence or absence of quality attributes that ascribe certain affordances to artifacts embedded in technology. Based on the above, several interesting questions stand out very promptly with regards to: (a) How are social and material agencies configured and enacted in a virtual setting? (b) What is the unit of analysis through which designers can gain insight to material agency, and (c) Does this focus on sociomateriality improve upon current design-oriented thinking and activities in the fields of HCI and CSCW?

III. METHODOLOGY

To shed light to these concerns, this section elaborates on an interpretive case study of collaborative MNLs. MNLs in the broader context of NMP present an interesting and challenging case that fosters the migration of an assemblage of embedded agencies established through sensory-based media to a new digital assemblage amenable to virtual work. Arguably, sociomaterial concerns constitute a core theme in pursuing pathways for such migration. Our case rests on an existing system, called DIAMOUSES [35]. DIAMOUSES offers an appropriate research setting to explore sociomateriality for several reasons. Firstly, it was designed to support MNLs as well as other NMP scenarios. Secondly, it was the first system to promote a separation of concerns; one component undertakes community management functions while another one implements music co-practicing [16]. This suits our purposes as it makes it easier to locate and 'trace' community- and practice-oriented features. Finally, DIAMOUSES could be easily modified to facilitate virtual ethnographic analysis of online music ensembles, which is one of the methods used to solicit data.

A. The research setting: DIAMOUSES notation lessons

DIAMOUSES is a system for NMP which demands a specialized set up in terms of equipment and software. Figure 1 depicts a typical DIAMOUSES configuration which can serve a variety of NMP scenarios ranging from rehearsals, improvising and learning. A MNL with DIAMOUSES represents a case of reconstructing online widely accepted practices based on established music constructs.

Figure 1: A typical DIAMOUSES configuration

Figure 2: DIAMOUSES piano lesson

To conduct MNLs with DIAMOUSES, moderators (or music theory tutors) prepare shared music materials (i.e., score, recordings, videos), schedule and organize the music lesson and invite participants. There are two prerequisite for taking part in MNLs. The first is the users' acceptance of the moderator's invitation, which is followed up by registration to a virtual 'room' containing the shared material of the lesson. Registration is a two-stage process where participants first become members of the community (by building their music profile) and then register to 'rooms'. The second prerequisite entails downloading the dedicated practice-specific software suite, which allows members to engage synchronously in the micro-negotiations of a specific music lesson. This toolkit provides a virtual space for participants to access shared music materials and negotiate them against their personal technical virtuosity. The scenario addressed in our experiment represents a multi-site engagement in a piano lesson with one moderator and several participants (see Figure 2).

B. Instruments and data collection

In order to set the focus on social and material aspects of MNL, it is compelling to define each constituent and identify possible measurable features of each. In our current analysis, the 'social' constituent is conceived as the online ensemble that is brought together to take part in a MNL. The agency of online ensembles is traceable by account of 'cultural' artifacts of practice. These are broadly defined as the digital remains that reveal individual or collective activity taking place either prior, during or following a MNL.

The 'material' constituent coins the digital representations embedded in technology which drive the technology's performative capacity. These can be assessed by examining designated quality attributes that enable or constrain use in any particular setting. This gives rise to the theoretical scaffold summarized in Figure 3. Thus, our normative perspective is that social structures such as online ensembles operate *with*, *on* or *within* and *through* representations embedded in technology by ways which determine (i.e., enable or constrain) certain uses. Then, such line of thinking can inform the design of an interviewing instrument which will comprise questions on cultural artefacts and design qualities.

To solicit information on what may be the range of plausible cultural artefacts, available scholarship was

consulted pointing out generic artefacts [20] (i.e., user profiles, expressing opinion & communication, finding & sharing information, establishing connections) as well as practice-specific [16] (i.e., organizing and performing own work, contributing to the shared practice agenda, coordinating / aligning with others and making sense of what is expected, aligning online and offline activities). In terms of quality attributes, our analysis relies on earlier works [29] indicating the role of certain attributes such as abstraction, portability, translucence, information and social connectivity. For each quality attribute a number of screening criteria were

established to facilitate an interpretation of the quality attribute in the users' language.

Thus, a data collection strategy was devised comprising qualitative interview data and digital traces highlighting online activities in the course of virtual ethnographic studies. Interviews were scheduled and carried out after the virtual ethnography of an online music ensemble with the researcher becoming actively involved in the practice of an online MNL. Further details on the instruments devised, the screening criteria, the specific questions, the data compiled and the data analysis methods have been elaborated elsewhere [10].
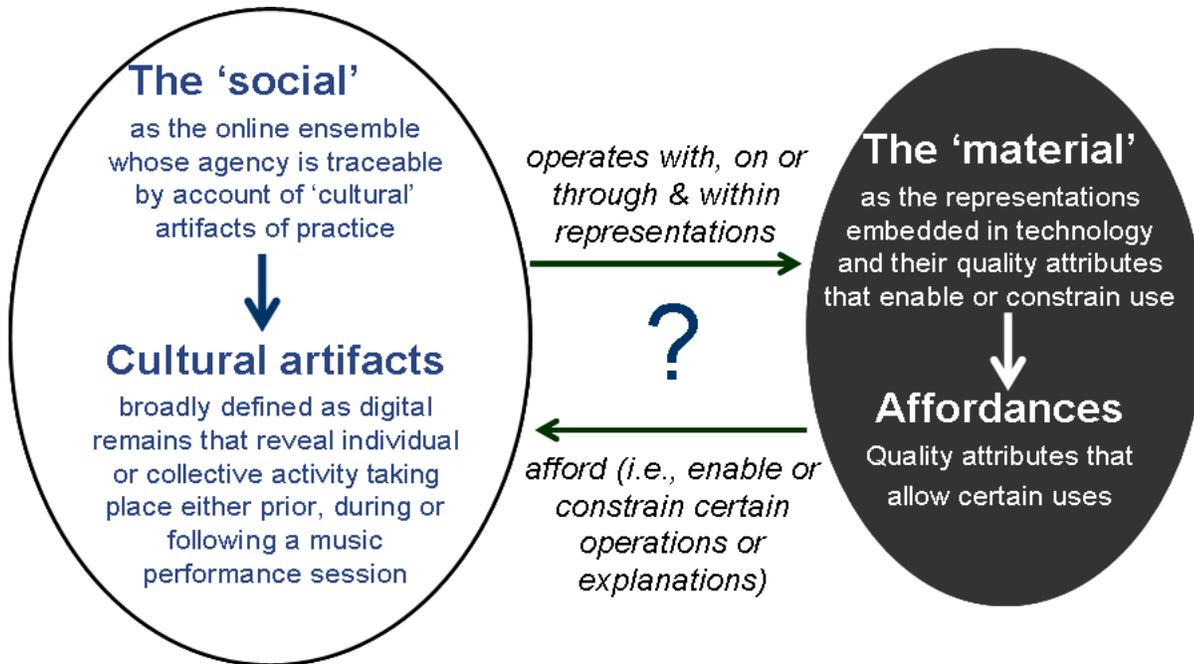
Figure 3: Theoretical scaffold

C. Summary of findings

Table I consolidates participants' responses to questions intended to unfold generic and practice-specific cultural artefacts and the virtual settlements in which their traces are retained. All participants confirmed that online ensembles emerge and sustain their function through the members' recurrent interactions in two separate but interrelated virtual spaces, namely the community management system (coined by the 'LR' code in Table I) and the music toolkit for co-engagement in synchronous MNLs (coded as 'MT' in Table I).

With respect to generic cultural artefacts participants acknowledged all those revealed by earlier studies [20], namely user profiling, tools for expressing opinion and communicating, finding and searching for information and establishing connections. As for practice-specific cultural artefacts (shaded rows), participants identified a wide variety, broadly classified in three sub-categories/sub-codes: (a) performative artefacts used to execute a shared agenda (b) coordinative artefacts that facilitate coordination and distributed organizing in the course of executing the agenda and (c) collective artefacts of historical

value that reveal the outcome of the online music ensemble. As shown in Table I, most of these artefacts are traceable in a single settlement – in most cases the music toolkit – with the exception of the music score and the supporting documentation that were deemed as having cross-settlement traces.

In the category of practice-oriented performative artefacts the score, as means for representing music (through XML scripts in the LR community management system or visual representations in the Music Toolkit), was the most frequently cited cultural artefact. From the participants' responses, it turns out that the score, in addition to being a learning material, it also serves as a boundary object intertwining between the online and offline setting, thus serving the purpose of structuring (and re-structuring) unknown contexts and/or actions and assigning them with meaning. In contrast to performative artefacts, practice-oriented coordinative artefacts were considered to be those that facilitate micro-negotiations during a MNL. In this category participants identified the metronome, the floor manager and objects for controlling remote user's performance. The final category of practice-oriented cultural artefacts includes those of historical value that

consolidate a collaborative engagement. All respondents identified the recorded performance (or recorded audio signal) as a socially constructed cultural artefact offering long-standing and traceable evidence of the ensemble's existence. It can therefore be concluded that the remains of a virtual ensemble engaged in MNLs comprise on the one hand the dynamics of collaboration during the lesson and on the other hand, the 'packaged' outcome codified in an audio format.

Assessment of the design qualities inscribed in technology and the extent to which they enable or constrain use turned out to be more challenging, as specific methods to anchor such features are lacking. As a result, heuristic assessment was recruited to provide the required evidence. Table II summarizes the results and confirms that DIAMOUSES exhibits several limitations and shortcomings. The first observation is the lack of any evidence for portability, abstraction, translucence and plasticity of the cultural artefacts coined as generic. This is not surprising as DIAMOUSES was not conceived or designed as social web site (in the sense suggested in [20]). In terms of digital manifestation (i.e., form), the system's focus was on GUI toolkit enhancements and CSCW oriented inscriptions for replicating domain-specific artefacts such as the score and synchronizing client applications. This constrained the range of digital representations implicated in practice and imposed strict schemes for binding across representations. As for the material features of these representations, the design of DIAMOUSES failed to acknowledge the properties that condition the forms of social action invoked by the embedded digital representations. This is derived from the weak support (i.e., Xw in Table II) evidenced by our users. More detailed analysis of participants' responses (see [10]) reveals loose accountability for explicit and implicit boundaries and lack of boundary spanning mechanisms. Similarly, the social protocols for co-engaging in practice are undermined and underserved.

Having outlined the limitations, it is important to notice that at the time of designing DIAMOUSES, such concerns were not widely shared or easily accommodated in technical specifications. Subsequent extensions in web services and standards, the increasing availability of public APIs for interoperability and the trend towards virtualization and cloud services, have created new opportunities and reset the research agenda.

TABLE I.   SUMMARY OF FINDINGS ON CULTURAL ARTIFACTS

Question	Artifact	Virtual settlement	
		LR	MT
Objects/artifacts intended to register and obtain roles	Custom registration system	X	
Tools for commenting & communicating	Forum	X	
	Chat		X
	Camera-based surveillance		X
Objects / artifacts for locating codified information or uploading user-created content	Liferay room, dedicated information templates and custom portlets	X	
Tools for building and maintaining social ties	Extend invitation (tutor)	X	
	Accept invitation (participants)	X	
Objects/artifacts whose manipulation results in traceable digital remains of my own performance	Music score as XML script	X	
	Music score as Interactive graphical object		X
Objects/artifacts that tell me what to do in relation to others	Synchronous chat		X
	Social proxies on music score		X
Objects/artifacts for making sense of the joint agenda and what is expected	Metronome		X
	Floor manager and states i.e., occupied vs. released		X
Objects/artifacts for understanding what others are doing	Remote audio signals (output)		X
	Mute (on/off)		X
Objects/artifacts for online & offline alignment	Local audio stream (input)		X
	Mute (on/off)		X
Objects/artifacts for collective performance	Mixed recorded performance	X	
	Supporting documentation	X	X

TABLE II.   DESIGN QUALITIES INSCRIBED IN TECHNOLOGY

|	PR	AB	TR	Pl	IC	SC	Virtual settlement	
							LR	MT
User profiles						X	X	
Expressing opinion & com/tion					X		X	X
Finding and searching information					X		X	X
Establishing connections					X	X	X	
Music score	X	X	Xw	Xw	X	X	X	X
Metronome			Xw		Xw			X
Floor manager			X		Xw			X
Remote performance controls					Xw			X
Recorded performance						X	X	
PR: Portability, **AB**: Abstraction, **TR**: Translucence, **Pl**: Plasticity, **IC**: Information Connectivity, **SC**: Social connectivity;								
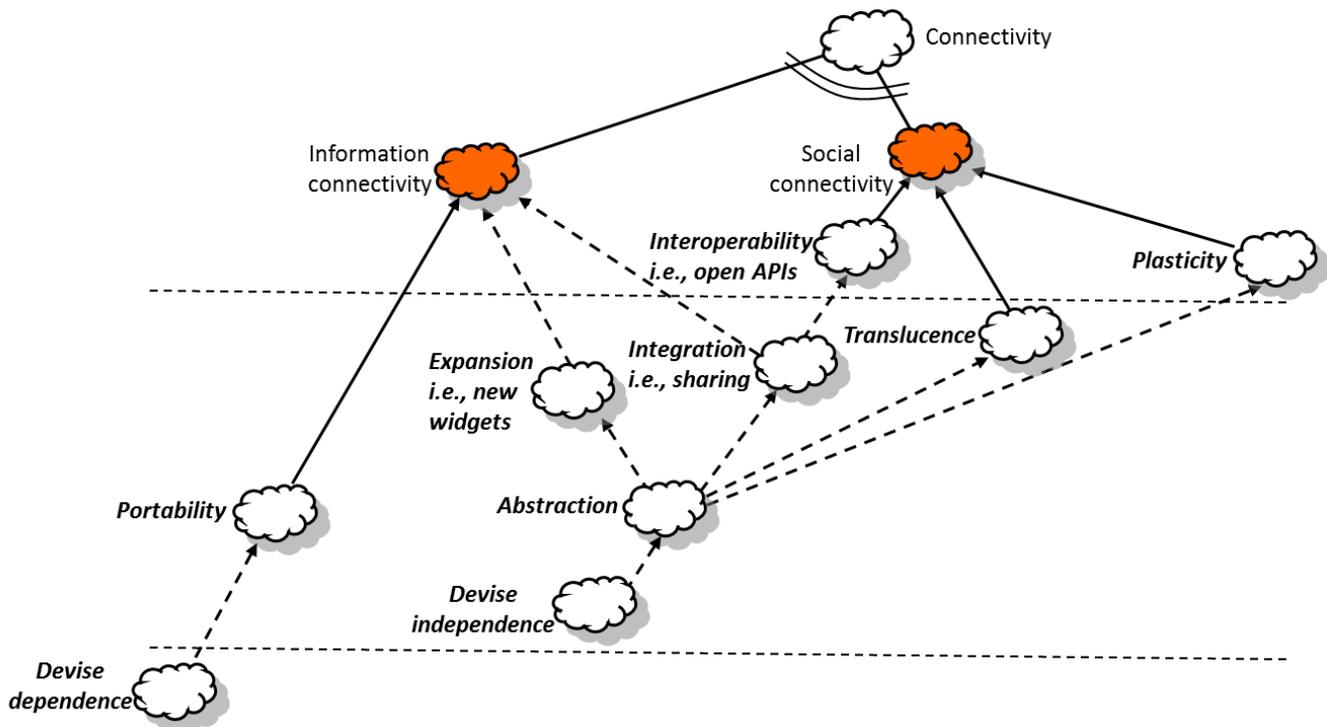
Figure 4: Inscribing design qualities into code

IV.   DESIGN IMPLICATIONS & DISCUSSION

A. Revisiting the research questions

The analysis presented thus far, allows us to reformulate our understanding of sociomateriality in the context of online MNLs. Thus, it is claimed that online music ensembles can be conceived as enacted cyber-structures whose agency in virtual space stems from the members' capacity to operate either *with*, *on* or *through* and *within* digital representations embedded in technologies. At the material level, such capacity can be further qualified not so much by the medium-specific affordances – after all, in the virtual setting everything is digital – as by the design affordances of the representations. It is these affordances and the presence or absence of quality attributes that enable or constrain not only what users can do, but also what is retained of their activities in the form of digital traces. In light of the above, it is worth revisiting the three questions relevant to present work (see section II.D).

With regards to the enactment of agencies, it is argued that social structures such as online music ensembles exercise control of representations embedded in technology to facilitate designated representational practices, such as MNLs. Material agency is enacted to anchor the technology's performative capacity which is determined by embedded design affordances and the inscriptions for portability, abstraction, translucence, portability, connectivity, etc.

As for the unit for analyzing material agency, our case study points to (a) digital traces and remains as the locus of cultural artifacts (see theoretical scaffold introduced in Figure 3) and (b) the transformative capacity stemming from the presence or absence of designated quality attributes. Through this lens, it is possible to dis-entangle the 'social' and the

'material' constituents of a digital assemblage, so as to re-configure material capacities thereby creating new capabilities for action. According to this view technologies can be taken apart and reassembled so as to facilitate new (and improved) social agency. Similarly, social agency is enacted in response to whatever material concerns are invoked by technical configurations.

Finally, in terms of sociomateriality's contribution to design-oriented thinking in the fields of HCI and CSCW (third research question), it is fair to conclude that our analysis favors the notion of 'imbrication' as a guide to designing digital artifacts and novel virtualities. Nevertheless, it should be acknowledged that using 'imbrication' as a lens for design requires further attention and detailed treatment which forms a part of our ongoing work [18].

B. Implications for design

Without underestimating the variety of issues pending attention, it is claimed that the material described in this paper raises implications for designers of computer-mediated representational practices. The most prominent is that it brings to the forefront the need to address (non-functional) quality attributes as first class design properties. To grasp the point one may recall briefly the pathway driving the evolution towards Web 2.0 and the Semantic Web. Specifically, the compelling need for portability in the initial WWW era was accommodated by virtual machine environments, Graphical User Interface toolkits and device-dependent mark-up. The next wave targeted abstraction and interoperability and was enabled by device independence and public Application Programming Interfaces. As a result web site authoring practices made way for blogging and collaborative editing practices, while end users became content generators (across different micro-

context such as blogging platforms, social networking services and other virtual settlements) from mere information consumers. In the current computing era, it stands to argue that technologies broadly classified under the Web 2.0 paradigm can be assessed in relation to a set of common provisions which inscribe in code certain degrees of connectivity. Schematically this is depicted in Figure 4 where design qualities (presented as clouds) intertwine to facilitate information and social connectivity. The point to be underlined is the variety of digital materialities that emerge as a result of different configurations of the designated qualities and the specific ways in which they become intertwined in code. Phrased differently, it may be argued that at any point in time there is a variety of digital materialities emerging from design commitments to different strategies that prevail within a certain computing paradigm. Thus, what is of interest is not so much the particular mix of quality attributes, but the extent to which such a mix and the ways in which it may be supported (i.e., open APIs, social aggregation services or any other kind or form of interoperability) leads to new capabilities and social practices.

V. SUMMARY & CONCLUSION

The present research is a step towards understanding how technologies are enacted in practice and the material concerns invoked by the presence or absence of designated quality attributes. It turns out that such qualities define a space of opportunities that anchor what can be done with the technology at hand. It is also evident that there is no absolute prescription for devising and undertaking analysis of sociomaterial considerations. At any one time, the social and the material are co-defined [3]. Nevertheless, digital technologies (as design artifacts) can be taken apart and re-assembled to convey different material properties and thus different sociomaterial realities [4]. Our case study confirms this conclusion and provides justification for what is possible, as well as what is desirable but not feasible, with the current version of DIAMOUSES. It also suggests pathways in which systems such as DIAMOUSES can be isolated and re-configured to create new opportunities for remediated practices [18].

Thus, in the light of our findings, it stands to argue that sociomateriality brings to design several concerns and challenges, including the compelling need for (a) devising the appropriate mix of flexible representations to be embedded in technology; (b) inscribing imbrications of representations to smooth out discontinuities in practice, and (c) tracing use to assess the socio-material aspects of representational practices in virtual space. Through this lens, it is possible for each representation to zoom-in on some selected elements of practice which become the focus of attention, while others are ignored or pushed into the background. At the same time, imbrications of (different) representations undertake to establish the new virtuality which dissolves ambiguities and/or breakdowns and re-orients users' practice.

REFERENCES

[1] W. J. Orlikowski (2007). "Sociomaterial Practices: Exploring Technology at Work.", Organization Science, Organization Studies 28(9), pp. 1435–1448.

[2] R. Zammuto, T. Griffith, A. Majchrzak, D. Dougherty and S. Faraj (2007). "Information technology and the changing fabric of organization," Organization Science 18(5), pp. 749-762.

[3] W. J. Orlikowski and S. V. Scott (2008). "Sociomateriality: Challenging the separation of technology, work and organization" The Academy of Management Annals, 2(1), pp. 433−474.

[4] P. M. Leonardi (2013)."Theoretical foundations for the study of sociomateriality", Information and Organization 23, pp. 59–76.

[5] M. Aanestad and T. Bratteteig Eds. (2013). Nordic contributions in IS Research, LNBIP 156, Springer-Verlag Berlin Heidelberg.

[6] A. Johri (2011). "Sociomaterial bricolage: The creation of location-spanning work practices by global software developers". Information and Software Technology 53 (9), pp. 955‑968.

[7] L. Suchman (2007). Human-Machine Reconfigurations: Plans and Situated Actions, (2nd ed.), New York: Cambridge University Press.

[8] P. Dourish and M. Mazmanian (2012). Media as material: Information Representations as Material Foundations for Organizational Practice. In Paul Carlile, Davide Nicolini, Ann Langley and Haridimos Tsoukas (eds.) Perspectives on Process Organization Studies: How Matter Matters: Objects, Artifacts and Materiality in Organization Studies. Volume 3. Eds. Oxford University Press.

[9] E. Parmiggiani and M. Mikalsen (2013). "The Facets of Sociomateriality: A Systematic Mapping of Emerging Concepts and Definitions". In M. Aanestad and T. Bratteteig (Eds.): SCIS 2013, LNBIP 156, pp. 87–103, Springer-Verlag Berlin Heidelberg.

[10] D. Akoumianakis (2013). "Socio-materiality of online music ensembles: An analysis based on cultural artifacts & affordances", The Science and Information Conference 2013 (SAI'2013), IEEE Computer Society.

[11] T. M. Harrison and B. Barthel (2009). "Wielding new media in Web 2.0: exploring the history of engagement with the collaborative construction of media products", New media & Society, 11 (1&2), pp. 155–178.

[12] K-B., Jensen and R. Helles (2011). "The internet as a cultural forum: Implications for research", New Media & Society, 13, pp. 517-533.

[13] A. Gruzd, B. Wellman and Y. Takhteyev (2011): "Imagining Twitter as an Imagined Community", American Behavioral Scientist, 55(10), pp. 1294–1318.

[14] M. Zappavigna (2011). "Ambient affiliation: A linguistic perspective on Twitter", New media & Society, 13 (5), pp. 788–806.

[15] D. Akoumianakis, N. Vidakis, A. Akrivos, G. Milolidakis, D. Kotsalis and G. Vellis (2011). "Building 'Flexible' vacation packages using collaborative assembly toolkits and dynamic packaging: The Case Study of the eKoNES", Journal of Vacation Marketing, 17(1), pp. 17-30.

[16] D. Akoumianakis and C. Alexandraki (2012). Collective Practices in Common Information Spaces: Insight from Two Case Studies, Human-Computer Interaction, 27 (4), pp. 311-351.

[17] J.-F. Blanchette (2011). A Material History of Bits. Journal of the American Society for Information Science and Technology, 62(6), pp. 1024-1057.

[18] D. Akoumianakis, G. Ktistakis, G. Vlachakis, P. Zervas and C. Alexandraki (2013). Collaborative music making as 'remediated' practice: Blending boundary objects and cloud services, Proc. of the 18th International Conference on Digital Signal Processing (DSP2013), Santorini, Greece, IEEE Press.

[19] P. H. Gochenour (2006). "Distributed communities and nodal subjects", New Media & Society, 8(1), pp. 33-51.

[20] W. Kim, O-P. Jeong and S-W. Lee (2010). "On social web sites", Information Systems, 35, pp. 215-236.

[21] R. T. Schatzki, K. Knorr-Cetina, K. and E. von Savigny (2001). The Practice Turn in Contemporary Theory, pp. 1–14. London: Routledge.

[22] P. M. Leonardi (2011). When flexible routines meet flexible technologies: Affordance, constraint, and the imbrication of human and material agencies. MIS Quarterly, 35(1), pp. 147−167.

[23] D. L. Introna and N. Hayes (2011). On sociomaterial imbrications: What plagiarism detection systems reveal and why it matters. Information and Organization 21, pp. 107-122.

[24] G. F. Lanzara (2010). "Remediation of practices: How new media change ways we see and do things in practical domains", First Monday, 15(6-7).

[25] J. Kallinikos, A. Aaltonen and A. Marton (2010). "A theory of digital objects", First Monday, 15 (6-7).

[26] D. E. Bailey, P. M. Leonardi and S. R. Barley (2012). "The Lure of the Virtual", Organization Science 23(5), pp. 1485-1504.

[27] W. J. Orlikowski (2000). "Using Technology and Constructing Structures: A Practice Lens for Studying Technology in Organizations", Organization Science, 11(4), pp. 404–28.

[28] U. Schultze and W. Orlikowski (2004). "A Practice Perspective on Technology-Mediated Network Relations: The Use of Internet-Based Self-Serve Technologies", Information Systems Research, 15 (1), pp. 87–106.

[29] D. Akoumianakis, D. (2010). "Tracing community life across virtual settlements", International Journal of Virtual Communities and Social Networking, 2(4), pp. 51-63.

[30] A. Pickering (1993). "The mangle of practice: agency and emergence in the sociology of science", The American Journal of Sociology, 99(3), pp. 559–589.

[31] J. Gibson (1979). The Ecological Approach to Visual Perception. Boston: Houghton Mifflin.

[32] D. A. Norman (1988). The Psychology of Everyday Things. New York: Basic Books

[33] D. A. Norman (1998). The Invisible Computer. Cambridge: MA, MIT.

[34] W. W. Gaver (1991). "Technology affordances", Proc. of the Annual Conference on Human Factors in Computing Systems (CHI'91), pp. 79-84, ACM Press.

[35] C. Alexandraki, and D. Akoumianakis (2010). "Exploring New Perspectives in NMP: The DIAMOUSES Framework", Computer Music Journal, 34, pp. 66-83.

Strategic Analysis towards the Formulation of Micro Sourcing Strategic Trusts

Noor Habibah Arshad

Department of Information Systems
Faculty of Computer and Mathematical Sciences,
University Teknologi MARA,
Shah Alam, Malaysia

Siti Salwa Salleh, Syaripah Ruzaini Syed Aris,
Norjansalika Janom, Norazam Mastuki

Faculty of Computer and Mathematical Sciences,
Faculty of Accountancy University Teknologi MARA,
Shah Alam, Malaysia

Abstract—Malaysian government, realising its responsibility to upgrade the quality of life, has identified micro sourcing industry as one of the potential industry to elevate the livelihoods of the poor especially the B40 group. The B40 in Malaysia is defined as household income level of less than RM 2,300 per month. The huge potential impacts of micro sourcing industry provide motivation for this research. In determining the best way for Malaysia to implement micro sourcing industry using the available resources, strategic analysis was conducted. Tools such as SWOT and Gap analysis were used to perform the strategic analysis. Thus, the objective of the paper is to develop a full awareness of the situation through Strengths, Weaknesses, Opportunities and Threats (SWOT) analysis. In order to determine the factors that define its current state, gap analysis was used to list down the factors needed to reach the target state and to fill the gap between these two states. Through these analyses it helps with both strategic planning and decision-making. Workshops were held to gather information from stakeholders and to discuss on the internal Strength and Weakness and the external Opportunity and Threat of micro sourcing. The discussions reveal the gap between where we are and where we want to be and also reveal areas where it must improve to meet the micro sourcing goals. The findings from the SWOT and Gap analyses will provide perspective, will reveal connections, areas for action and also identify deficiencies. The analyses will also build on the strengths, minimize the weaknesses, seize opportunities and counteracts threats and fine tuning one process. Finally, micro sourcing strategic trusts will be formulated.

Keywords— micro workers; job providers; crowd sourcing; B40 group; SWOT analysis; Gap analysis

I. INTRODUCTION

In December 1996 the Malaysian National IT Agenda (NITA) was launched by the National IT Council (NITC). The Agenda provides the foundation and framework – known as the National IT Framework (NITF) – for the utilisation of ICT to transform Malaysia into a developed nation in its own mould consistent with Vision 2020 [1]. NITA's vision is to utilise ICT to transform Malaysian society into an information society, followed by a knowledge society and finally to a values-based knowledge society. NITA focuses on the development of people, info-structure and applications to create value, to provide equity and access to all Malaysians, and to qualitatively transform the society into a values-based knowledge society by the year 2020 [1].

One key strategy to transform the society into a values-based knowledge society is to elevate the livelihoods of the poor. With this strategy in mind, Digital Malaysia was established as an enabler towards a knowledge-based economy to drive wealth creation and enhance quality of life by harnessing and building upon Malaysia's varied ICT initiative [2]. Prime Minister of Malaysia also mentioned about Malaysia intention to build an ecosystem that promote the pervasive use of ICT in all aspect of the economy [3].

Under Digital Malaysia [4], micro sourcing industry has been identified as a potential industry to uplift the income of the population in the bottom 40 percent household income (B40). Involvement of B40 group in micro sourcing activities will allow them digital access and be paid for completing micro tasks. The household income level of the B40 is less than RM 2,300 per month [5]. Majority of these households have single income earners. More than half of the household's heads (52 %) have no recognised education background. With no education background, low skills level and in certain cases, living in remote locations, the B40 households are limited in their economic mobility and ability to secure higher paying jobs as well as income opportunities. Studies have also shown that there is low ICT adoption among poor communities in Malaysia, which is part of the B40 group [6].

The micro sourcing industry in Malaysia is still at its infancy stage and not properly structured [7]. There exists micro sourcing platforms but the number is relatively small and their roles are very limited, as mediator between demand (job providers) and micro workers (supply), and advertising tasks that are sourced from demand. Number of demand is also quite limited and concentrated within the private sector. The tasks available in the market are not targeted to any specific micro workers and these workers are given proper training to perform the task. The existing scenario could make the industry unsustainable in the long run.

Thus, the objectives of this paper are: (1) to come up with the positives and negatives of micro sourcing within the demand, supply and platforms (S-W) and outside of it, in the external environment (O-T), and (2) to present current situation and also highlighting the gaps exist that need to be filled. By developing a full awareness of the situation can help with both strategic planning and decision-making.

II. Micro Sourcing Current Scenarios

Micro sourcing is defined as "the art of taking a job traditionally performed by a designated agent (usually an employee) and outsourcing it to an undefined, generally large group of people in the form of an open call" [8]. In another word it is "to outsource a job to a large, anonymous crowd of workers, the so-called human cloud, in the form of an open call"[9]. When a task needs to be completed, it can be done faster and more efficiently with the help of others via micro sourcing. This is the fundamental understanding of what micro sourcing is from an employer's perspective. Micro sourcing has become a cost effective way for companies to give opportunities for individuals outside of the companies to use their skills and time for good use and earn additional income. These companies pay people based on the amount of hours of works, and save millions of ringgits by doing so. Companies also will be able to tap into a large pool of talents, allowing these talents to choose what works suit them best. Micro sourcing also allows companies to employ a large group of skilled people to handle projects within a specific time frame for a fixed price. Typical micro tasks are translation, data validation, image tagging, research, writing, editing, categorisation and data entry.

Some of existing micro sourcing platforms available globally supporting crowd-sourced micro tasks are AmazonMechanicalTurk, CrowdFlower, SamaSource, Ushahidi, Micro sourcing and ODesk. AmazonMechanicalTurk was launched in November 2005 and which the requestors are restricted to US-based entities, however the workers can be sourced globally [10]. CrowdFlower was founded in 2007 and uses the MTurk platform to distribute work, but provides its own interface on which work is completed. It also has sophisticated APIs to create and manage works [11]. SamaSource was founded in 2008 and claims to have a dedicated team of remote workers but does not post jobs on a public portal like Mturk [12]. Ushahidi was founded in 2008 and provides a platform for information collection, visualization and interactive mapping, especially for crises [13]. Micro sourcing is Philippines based company providing traditional outsourcing solutions [14]. Meanwhile, ODesk [15] was founded in 2003 and focus more on long-term work through remote staffing than real micro sourcing.

Global micro sourcing industry recorded substantial growth in the past few years [16]. The industry's Completed Task grew exponentially from 400,000 in 2008 to 291.8 million in 2011. The industry's total revenue meanwhile grew approximately 52.6% in 2010 and accelerated further to grow by 75% in 2011. Internet Services contribute the largest share of micro sourcing industry's revenue, 29% of total industry's revenue, followed by Media and Entertainment (20%) and Technology (18%) sectors. While Manufacturing and Financial Services sectors have a relatively low share in of the industry's revenue at 13% and 8% respectively, these sectors present significant untapped opportunity for micro sourcing penetration in the future.

Demand in the global micro sourcing industry is driven by start-up and small companies. Collectively they account for over 60% of the market revenues. Start-up companies drive majority of the revenues in the industry, contributing 39% of the total revenues. Large enterprises with revenue of more than US$1 billion represented only 8 % of total job providers but contributed 21% of total revenues due to huge transaction volume [16].

It was also reported that [16], geographically, companies in North America and Europe are the largest job providers in the global micro sourcing industry, offering 90% of jobs collectively. Supply of workers are more diverse with North America provides the biggest number of workers, 45% of total workers, followed by Asia. Comparing the distribution of the job provider and the workers, Europe is the largest net job provider while Asia is the largest net supplier. Europe provides 36% of global micro sourcing jobs but supply only 18% of global micro sourcing worker. Asia meanwhile, provides 35% of the global workforce but only 7% of global micro sourcing jobs.

However, there is no data available on the number of job providers in local micro sourcing industry. To have an idea of the job providers' scenario in Malaysia, we use data available from a single platform case study by Human Capital Connection (YourPartTime.com)[17]. The traditional outsourcing industry in Malaysia is growing with number of jobs advertised and job providers increasing. As shown in Figure 1, number of job providers grew three-fold from 520 in 2009 to 2,400 in June 2012. Number of jobs advertised also grew three-fold from 619 to 3,000 in the same period. Figure 2 shows the value of job advertised. Based on value of jobs advertised in the period of January to June in 2012, 77% of them are sourced from Small Medium Enterprises (SMEs), followed by multinational firms (18%). The Government provided only 1% of total market.
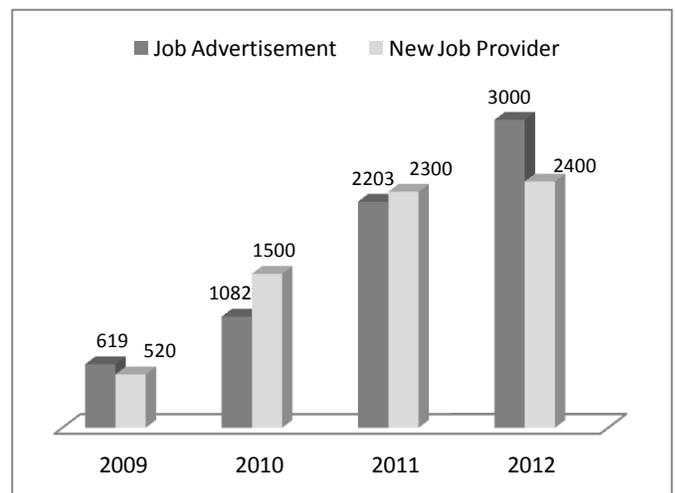

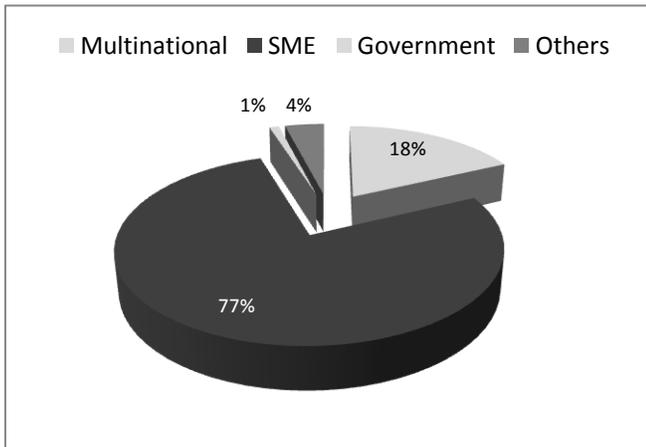Fig. 1. Jobs Advertisement and Job Providers in Malaysian outourcing industry (2009 to June 2012) [17]

Fig. 2.   . Malaysian outsourcing market share by Job Providers [17]
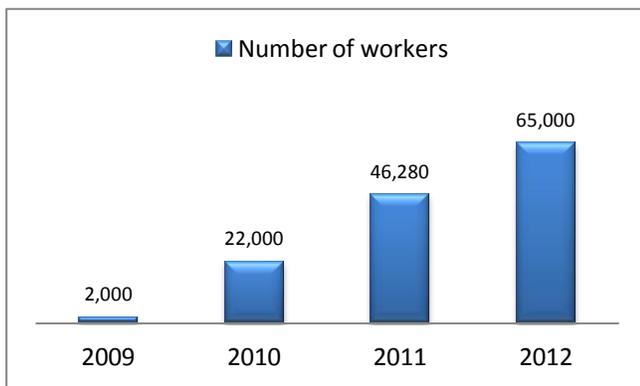

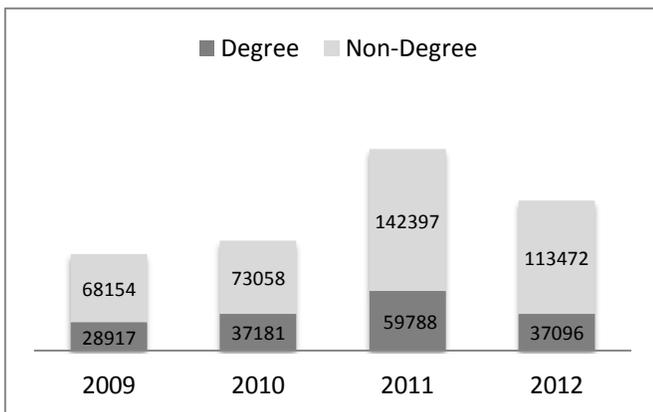Fig. 3.   Number of Malaysian micro workers [17]

Fig. 4.   Job placement by Job Malaysia for micro sourcing tasks [18]

The same as job providers, there is no comprehensive data on the number of micro workers in Malaysia. Thus, data from the same study by YourPartTime.com is used as proxy. Figure 3 showed that the number of micro workers in Malaysia is growing exponentially, in line with global trend. With initial number of 2,000 in 2009, the number increased to 22,000 in 2010, 46,280 in 2011 and 65,000 as at June 2012. This data is supported by increasing number of job placement for micro

sourcing tasks by Job Malaysia [18] that showed similar trend as shown in Figure 4.

III.   METHODOLOGY

Our first source of reference is documentation relating to micro sourcing and Digital Malaysia to understand the micro sourcing industry and the overall concept of Digital Malaysia. Two workshops were held to gather information from stakeholders and to discuss our initial findings for the study. The first workshop was held to gather information from stakeholders and to discuss on micro sourcing potentials, opportunities, potential growth and critical enablers from the perspectives of stakeholders. During the SWOT and Gap analysis, all stakeholders involved are asked to pool their individual and shared knowledge and experiences. Follow up interviews were done with some related government agencies and private sectors. The interviews help to uncover current practices, requirements and challenges in existing Malaysian micro sourcing environment. The second workshop was held to discuss the findings and recommendations on the micro sourcing strategic thrusts, framework, ecosystem and directions. Participants of both workshops were from Micro Sourcing for B40 Consultative Committee (MSCC), government ministries and agencies, Non-Governmental Organization NGOs) and private sectors. The information gathered from all these sources was then compiled and analysed as inputs for this paper.

IV.   RESULTS AND DISCUSSION

Strategic Analysis using SWOT and Gap analyses have been conducted in order to achieve the desired and targeted results. Strategic trusts are formulated based on these analyses. Based on the analysis of the current scenario of local micro sourcing industry, detailed description of SWOT, GAP analyses of the current scenario of local micro sourcing industry are described below.

The SWOT and GAP analyses were to analyse the three groups involved in the micro sourcing. The first group is Demand who provide the tasks. The second group is Supply or Micro Workers who are the people who complete the micro tasks. The third group is the Malaysian Micro sourcing Platforms. The platforms are the software and processes to run micro works and micro sourcing projects for use with internal or external crowd. The workshop captured the four-cornered SWOT analysis that can be used to support in building the strengths, minimize the weaknesses, seize opportunities and counteract threats. Meanwhile, the Gap analysis highlighting the gap exists that need to be filled.

1)   SWOT Analysis
Details of the findings are described below.

A.   *SOWT Analysis – Job Provider (Demand)*

The results for SWOT Analysis for Job Providers are:

a) Strengths

- Existing government guidelines on Homeworking: Incentives to job providers: Digital Malaysia has identified micro sourcing as one of projects under the Social Dimension Thrust. Some incentives shall be

available to induce local companies to outsource micro tasks especially those for B40 group. Public sector can also provide micro sourcing jobs such as digitisation of Government processes.

- Growing micro sourcing international market and demand. The global traditional outsourcing market size is estimated at be valued at US$1.7 trillion. Malaysian outsourcing industry is predicted to be worth US$1.9 billion by 2013, which include BPO, Systems Integration (SI) and IT Consulting.

b) Weaknesses

- Lack of awareness and low level of acceptance by local corporations:
Job Malaysia indicated that the awareness and acceptance by employers to employ micro workers are still low. There are only very few micro sourcing platforms that promote and stimulate the growth of the micro sourcing activities in Malaysia.

c) Opportunities

- Interest from socially responsible companies: Increasing number of "Triple" bottom line companies which focus on social, environment and financial results.

- Micro sourcing jobs from public sector: Amendment of procurement policies will promote micro sourcing jobs. Government procurement on supplies and services is very sizable, expected to be: RM14.2 billion and RM13.5 billion respectively in 2012 and 2013. Government jobs to be used as catalyst for demand although this should not pose additional cost to Government since it should reduce other labour costs.

- Regional market and OIC member countries market: Global micro sourcing market size was estimated at US$4.5 billion in 2010 with 144,000 workers. The market size is estimated to grow to US$20 billion in 2015 with 780,000 workers. Organization of Islamic Conferences (OIC) can also become a potential market to supply job for micro sourcing.

- Local SMEs (Over 600k): There are 645,136 SMEs that can be job providers for the micro sourcing industry. Most of these SMEs have already embrace ICT in their businesses.

d) Threats

- Quality and reliability of local micro workers: Many local firms have reservation on the quality and reliability of the tasks done by micro workers. These reservations can be summarised below:

- Will B40 be reliable?

- Will the quality of products and services provided meet expectation?

- Will the confidential data safe?

- Will the job delivered as the specified deadlines?

- Will there be enough supply of B40 with the right skill and attitude?

- Cheap foreign labour: Some local firms still prefer employing cheap foreign labour rather than exploring micro sourcing to keep costs low.

- Public sector procurement policy and process for crowd sourcing / micro sourcing activities: Government agencies are not willing or incapable of dividing jobs into micro tasking. They prefer to provide jobs to large contractors for simpler management.

B. *SWOT Analysis – Micro Workers (Supply)*

The results for SWOT Analysis for Micro Workers are:

a) Strengths

- Government support: Raising the living standards of low income household is one of the six National Key Result Areas (NKRAs). The Low income Households NKRA aims to completely eradicate hardcore poverty, reduce the incidence of poverty and enhance the productivity of low-income households.

- Existing guidelines for home working (JTKSM) can be extended to cover micro sourcing labour activities: Existing guidelines for home working under Department of Labour Peninsular Malaysia Ministry of Human Resources or *Jabatan Tenaga Kerja Semenanjung Malaysia* (JTKSM) can be applied to micro sourcing industry. Home working income such as income from micro works is exempted from income tax.

- Large potential workforce with access to basic ICT infrastructure: The B40 group was evenly distributed between urban and rural areas. Despite majority of rural household are self-employed, the main source of income is from paid employment. Total of 55% of urban B40 and 39.7% of rural B40 are salaried workers. These provide large potential workforce with some specific job related skills.

- Support from the Low Income Household NKRA to eradicate poverty: Data from the Ministry of Rural and Regional Development or *Kementerian Pembangunan Luar Bandar dan Wilayah* (KPLBW) shows that most of the B40 households have mobile phones. Some of them have PCs at home while those who do not own PC or have Internet connection are able to access Internet at community centres and cyber cafes. Therefore, most of the B40 households can participate in micro sourcing using mobile phones and PCs to communicate with micro sourcing platforms.

b) Weaknesses

- Majority of B40 with low level of education: Data from Department of Statistics (DoS) shows that most B40 households have low level education. In year 2009, 52.3% of people in the B40 group have no education certificate. Only 4.4% have post-SPM certificates (post high school certificates), while 25.4% and 17.9% have SPM (high school certificate) and PMR (lower school

certificate) respectively as their highest education certificates. It will be quite a challenge to prepare them as micro workers.

- Most of B40 are low-skilled workers: Data from Human Capital Sdn. Bhd., Yayasan Basmi Kemiskinan (YBK) and Yayasan Pembangunan Islam Malaysia (YAPEIM) shows that the B40 group possesses moderate to low level skills. They have moderate competency in basic computer and mobile phone skills, and low competency in English, Internet skills, communication, financial management and multi-tasking.

- Most of B40 are low level in attitude: Feedback from Ministry of Rural and Regional Development or Kementerian Pembangunan Luar Bandar dan Wilayah (KPLBW) and Pusat Zakat Selangor (PZS) shows that the common personality or features of the B40 are that they display low level in attitude, internal motivation, and goal orientation. They also do not have a long-term goal, lacking internal focus; have low self-esteem and low motivation to get extra income or reward. They are not autonomous and do not like responsibility. They have moderate level of creativity, productivity and initiative. Many of them are not interested on getting and doing work to earn their living and prefer to receive financial assistance and aids from Government or relief agencies.

- Financial-related issue: A Focus group discussion highlighted the fact that the B40 group may be willing to use their own money in advance only if the amount is little (for example 5-10 ringgit for phone prepaid reload and this is also subject to the level of their socioeconomic background). The B40 may also be willing to join the micro sourcing work force if the income is attractive (RM100 and above).

- Low level of English proficiency to cater for International demand: Total of 52.3% of people in the B40 group have no education certificate and most of them have low to moderate competency in computer and Internet skills, as well as English proficiency.

c) Opportunities

- Flexible and income generation opportunity for the B40: Large range of tasks is available in the micro sourcing industry, which provides opportunities for the B40 group to participate in micro tasks provided by international firms.

- Opportunities to upgrade skills in industry-relevant areas: Opportunities to upgrade skills in industry-relevant areas by skills development centres such as Pusat Giat MARA, Jabatan Kemajuan Masyarakat (KEMAS) and community colleges. The B40 group also needs soft skills training in the area of Internet usage, words processing, spreadsheet, presentation software, web or blog development, basic communication skills, marketing skills, time

management skills, platform training and motivational courses on self awareness and personal development.

- Availability of wide range of tasks in micro sourcing industry: Many parties can engage in activities to attract B40 into micro sourcing industry. Some of the activities are:

- Awareness campaigns such as road-shows.

- Reaching out to NGOs and community to support, motivate and gather feedback from existing and new micro workers.

- Masterminding and coaching to ensure continuous micro workers development.

d) Threats

- Inability if B40 micro workers to compete against skilled workers: Those not in B40 might compete with B40 in completing micro tasks, which could reduce the effectiveness of the Project.

C. SWOT Analysis – SWOT Analysis – Platforms

The results for SWOT Analysis for Platforms are:

a) Strengths

- Government support via the Digital Malaysia initiative: Under the Social Dimension Thrust, Digital Malaysia will look at infusing technology to uplift quality of life by spurring Netizens to move from digital consumption to digital production and also expanding digital access to all levels of society, especially the B40 group.

- Local platforms understand the B40 needs and issues: Some platforms have been working with B40 and understand their needs. With the support of Job Provider (Demand) and Community Champions, large range of tasks can be executed through micro sourcing.

b) Weaknesses

- Very basic functionality: There is no established national minimum standard to increase credibility of local platforms to both local and international markets. Certification program for micro sourcing platforms have yet to be created.

- Lack experience and exposure to international market: There is a lack of skilled/ experienced human resources to support micro sourcing platforms. There is no structured program to promote growth of local platforms. Specialised project management and business development teams are needed to promote and market micro sourcing solutions.

- Very basic micro business model: Currently, micro tasks advertised are selected directly by micro workers. There is no auto mechanism to ensure that micro tasks are selected by the most suitable micro workers.

- Weak payment mechanism: Payment for completed micro tasks is directly from organisation to micro workers. This mechanism creates issues such as guarantee of payment to be received by the workers. A

better mechanism is needed to ensure micro workers receive payment for tasks completed.

- Lack of escrow account functionality: Currently, job provider pays the imbursement directly to workers. In ability to ensure more secure environment in the payment mechanism, escrows account functionality should be introduces.

c) Opportunities

- Collaboration with established international platforms/ players: There is potential to develop various specialisation and niche area micro sourcing platforms catering for local and regional micro workers. Local platforms can collaborate with establish international platforms such as Samasource and Crowdflower.

- Leveraging on cloud computing technologies and solutions: MSC status companies offer cheap cloud computing resources which enable low operating cost for micro sourcing platforms players. The low operating cost will make Malaysian platforms competitive in international market.

- New model/ localized platforms for local, regional market needs: International micro sourcing platforms might not be suitable for the local micro workers especially the B40 group. A unique platform can also be attractive for new job providers, local and international.

- E-working opportunities: Malaysia has among the highest internet penetration with 89% of the population are internet users, based on data as at July 2012. This data shows the potential huge pool of micro workers for local micro sourcing platforms.

d) Threats

- Acceptance by local market as international platforms are more established and experience: International platforms are more efficient and have more experience, which will provide great competition for local platforms.

- Cyber security issues including online scams/ fake platforms:
Occurrences of cybercrime-have instilled fear and lacked of confidence with ICT platforms. Lack of motivation and attitude problem of B40 workers-might make them not interested to new type of income generating jobs.

The SWOT analysis which focuses on the four elements which are the strengths, weaknesses, opportunities and threats are able to address or at least recognized the full situation of micro sourcing in Malaysia. Furthermore, this analysis could be a source of micro sourcing strategic planning where it helps to identify core competencies and setting up the objectives for strategic planning.

2) GAP Analysis
The Gap analysis shows the present situation and also highlighting the gaps exist that need to be filled.

A. Gap Analysis - Job Providers (Demand)
The results for Gap Analysis for Job Providers are:

a) Local Market - Lack of awareness and low industry acceptance
In general, awareness and acceptance by employers to employ micro workers are still low either because of lack of knowledge about the industry or lack of trust on local platforms. These issues have to be mitigated by awareness programs and proper governance in the industry.

- *Public Sector*

The Government provides only 1% of the total jobs in local traditional outsourcing industry. The share of outsourced jobs from the Government should be increased via its procurement policies, tapping into Government's expenditure on supplies and services that range between RM10 billion to RM15 billion a year. These outsourced jobs can be break down as micro tasks such as archive digitisation and e-Government related.

- *Private Sector*

Even though SMEs contribute 77% of total jobs provided in local traditional outsourcing industry, there are still lots of potential that can be tapped especially in micro sourcing. In 2010 [19], 97.3% of 645,136 registered companies are categorised as SMEs as shown in Figure 5. 90% of these SMEs are in the services sector, 6% in the manufacturing sector, while the remaining in construction, agriculture and mining sectors. In the services sector, SMEs mainly operate in distributive trade subsector (wholesale and retail trade services).
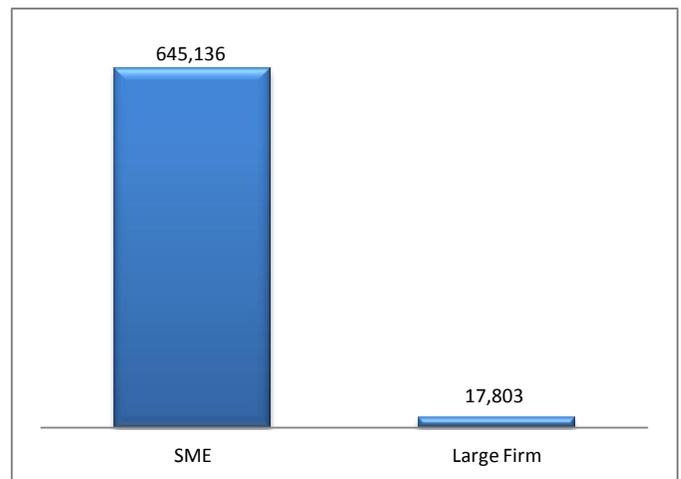

Fig. 5. Registered companies in Malaysia in 2010 [19]

SMEs have also contributed significantly to the Malaysian economy.

Figure 6 shows that SMEs have consistently contributed more than 65% of total private sector's value added to the Malaysian GDP every year
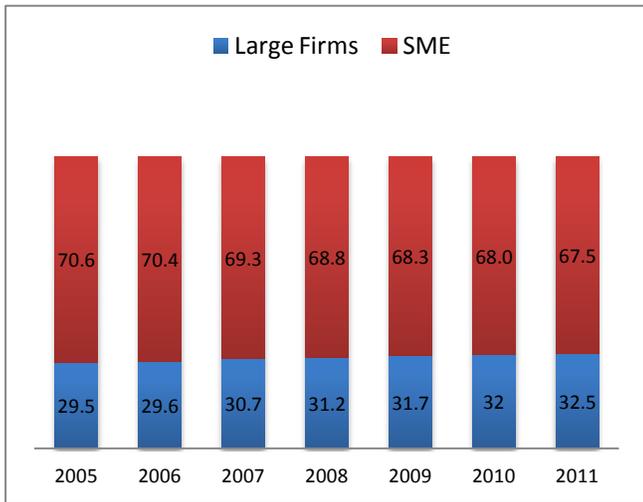
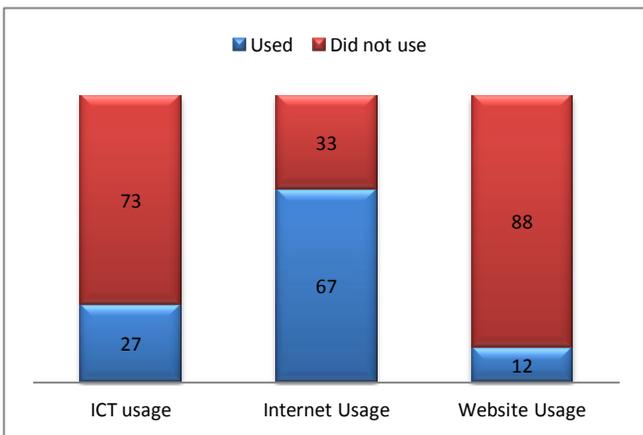Fig. 6.   Private sector's value added to Malaysian GDP [19]

Fig. 7.   ICT usage by SMEs [19]

Even though SMEs are main provider of outsourcing jobs, usage of ICT by SMEs is still low in Malaysia. Currently, only 25% of SMEs uses ICT in their businesses and only 12% of them have their own websites as shown in Figure 7. The usage however, is expected to grow as Google Malaysia, in partnership with MCMC, has launched the Get Malaysian Business Online (GMBO) campaign, a new initiative to bring more Malaysian businesses online. Google Malaysia will invest up to RM10 million to help 50,000 local SMEs set up their websites [19]. These new ICT users should be tapped as job providers for micro sourcing industry. There is also a need to increase awareness among the SMEs of the potential of lower operating costs with micro sourcing.

Other than SMEs, large firms should also be encouraged to provide jobs in micro sourcing industry especially in IT-related tasks. Socially responsible companies can also be roped in to provide micro works especially to targeted micro workers.

A case study was conducted to identify local companies that currently outsource some of their tasks and their potential involvement as job providers in micro sourcing industry. The study's findings are summarised in the Table 1.

TABLE I.   CASE STUDY'S FINDINGS: POTENTIAL LOCAL JOB PROVIDERS

	HR Company	IT Services	IT Consultant	FM Radio	Telco Berhad
EXISTING OUTSOURCED TASKS					
Out-sourced Task	Payroll, Allowance, Travelling claims, Leaves, Medical.	Multi-media. Creative, Data entries, Tedious repetitive works.	All areas which are not their core expertise	Sales promoter in each state	Application Support & Infra-structure Services
Costs	1.2 million (RM) per year	10 K (RM) to 500 K (RM) per project	1 million to 2 million (RM) a year	No amount revealed	About 10 million (RM) per month
Local/ Off-shore	Local	Local	Development is done offshore	Local	Mostly local
POTENTIAL MICRO SOURCING TASKS					
Area Out-sourced	All IT-related	Mainly IT	Only IT-related.	Technical area	IT-related & Infra-structure Services
Potential for Micro sourcing	Hosting of server and messaging	Clerical task	IT Maintenance and printing documents nationwide	No specific task identified	Application Support, Trans-formation Project, Infra-structure Services

b) International Market

The global outsourcing market as shown in Figure 8 is expected to grow at a CAGR of 4.31% from 2010 to 2015. By 2015, the market is estimated to reach US$574 billion, with Information Technology Outsourcing (ITO), which comprises both Application Outsourcing and Infrastructure outsourcing, accounting for US$363 billion, and BPO at US$211 billion in revenue [20].

Malaysia should leveraged on the expected growth in global micro sourcing industry especially Impact Sourcing. Malaysia with its diverse population with multicultural background should be promoted as a regional micro sourcing hub to attract international firms outsourcing their micro tasks in Malaysia.

Fig. 8.   Global outsourcing market (US$ Billion) [20]

According to the Monitor Group [21] the following trends will drive the growth of the Impact Sourcing market in the coming years:

- The increasing reach and decreasing cost of telecommunications is expanding the list of countries and regions that can compete for outsourcing work;
- There is continuing pressure on corporations and, to some extent, Governments to deliver services, particularly non-core services, in a cost effective manner;
- There is increasing demand for the digitisation of non-digitised records across the private and public sectors;
- There is an increasing level of Government activity that lends itself to outsourcing tasks, such as archive digitisation, the growth of e-Government platforms and universal identity number projects; and
- There exists latent demand for tasks such as data management, content editing and low-end transcription from large global companies.

A study by Accenture shows that companies in Asia Pacific are more likely to participate in Impact Sourcing. The study also shows that companies are willing to allocate 25% of their overall outsourcing portfolio for Impact Sourcing especially in Application Outsourcing and BPO. The study also highlighted barriers that prevent companies from considering Impact Sourcing which are Security, Language and Skills Availability, Infrastructure/ Technology Reliability, and Experience. Thus, these barriers need to be handled to attract micro sourcing jobs from international market.

B. *Gap Analysis - Micro Workers (Supply)*

The results for Gap Analysis for Micro Workers are:

   a) Attitude and motivation of B40

The most glaring gap on the supply side of the micro sourcing industry is the quality of the micro workers. While majority of micro workers in the international market are well-educated, the targeted micro workers under this Project are

from the B40 group where majority of them have no education certificate. The B40 group also have low to moderate skill levels in basic computer and mobile phone skills as well as English and communication skills. In addition, most of the B40 group display low level in attitude, internal motivation, and goal orientation.

Therefore, trainings to improve the education and skills level of the B40 group to handle micro tasks is utmost important. There is a need to rope in NGOs, foundations and individuals as Community Champions to engage with these communities and provide support to help them embrace the micro sourcing industry.

Types of training for these micro workers should also suit the types of tasks that are suitable to them. Some of micro sourcing tasks that are available and might be suitable for the B40 are:

- *Data Gathering*
    - Gathering price watch data
    - National population census taking
- *Promotional Services*
    - Positive word-of-mouth on social networking
- *Back Office*
    - Data entry
    - Data processing
- *ICT and Internet related*
    - Web development
    - Search engine optimisation
- *Creative Services*
    - Web design
    - Graphic design
    - Advertisement design
- *Language-related services*
    - Customer support
    - Telemarketing
    - Transcription

   b) Language proficiency

52.3% of people in the B40 group have no education certificate and most of them have low to moderate competency in computer and Internet skills, as well as English proficiency

C. *Gap Analysis – Platforms*

The results for Gap Analysis for Platforms are:

   a) Very basic functionality

Existing micro sourcing platforms in Malaysia have limited roles [22]. For example, they advertise micro tasks to micro workers but do not take the role of matching the tasks to the most suitable and able micro workers. Analysis of existing platforms has revealed the following:

- *Platforms act only as mediator*

A platform receives a task from a job provider and advertises the task to micro sourcing workers. Once the micro sourcing workers agree to accept the task, the platform will introduce the job provider and the micro workers.

- *Platforms do not have auto matching mechanism*

Once a task is advertised, it will be selected by micro workers. There is no mechanism to ensure that the micro workers are the most suitable and able workers to complete the task.

- *Platforms are not responsible for payment to micro workers*

Payments for completed tasks are made directly from job providers to micro workers. Some issues might arise from this arrangement such as guarantee of payment to be received by the workers. Therefore, there should be a mechanism to ensure that the payment is successfully received by the micro sourcing workers.

- *Platforms lack of credibility and reliability*

Platforms must ensure that tasks to be completed are properly executed by the micro workers. They must have mechanism to ensure the quality of the tasks completed.

b) Early stage models

To compete with international platforms and strengthen the local micro sourcing industry, local platforms' roles have to be expanded with well-established international platforms made as benchmark. As this Project is targeted to raise income of B40 group, platforms created must be well diverse and have niche/ focus areas to cater for targeted groups.

Strengthening and widening platforms' roles requires skilled/ experienced talents, and specialised project management and business development teams. Currently, there might not be enough local human resources to cater for this need. A strategy is required to attract and develop these talents.

3) Strategic Thrusts

Based on the strategic analysis, five strategic thrusts have been identified. The Strategic Thrusts are [23]:

a) Strategic Thrust 1: Harnessing Demand Side (Job Providers) of Domestic and International Market;

b) Strategic Thrust 2: Platform Capacity and Capability Building;

c) Strategic Thrust 3: Leverage and Utilise Existing Infrastructure;

d) Strategic Thrust 4: Uplift and Enhance Capability of the Supply Side (Micro Workers); and

e) Strategic Thrust 5: Instruments to Expedite Growth of Local Micro sourcing Industry.

These five Strategic Thrusts have been identified as the foundation for the Strategic Framework of the micro sourcing industry development in Malaysia. The rationale behind the Strategic Framework is to create micro sourcing Ecosystem [24] by narrowing or eliminating the gaps identified previously. The development of a complete Micro sourcing Ecosystem is to ensure industry sustainability and participation of B40 micro workers in an organised way. Success of the Ecosystem relies on effective roles played by the relevant stakeholders.

V. CONCLUSION

In general, awareness and acceptance by employers to employ micro workers are still low either because of lack of knowledge about the industry or lack of trust on local platforms. To compete with international platforms and strengthen the local micro sourcing industry, local platforms' roles have to be expanded with well-established international platforms made as benchmark. As this research is targeted to raise income of B40 group, platforms created must be well diverse and have niche/focus areas to cater for targeted groups. Strengthening and widening platforms' roles requires skilled/experienced talents, and specialised project management and business development teams.

Currently, there might not be enough local human resources to cater for this need. This is due to the majority of the targeted micro workers under this research are from the B40 group where majority of them having no education certificate. The B40 group also has low to moderate skill levels in basic computer and mobile phone skills as well as English and communication skills. In addition, most of the B40 group displays low level in attitude, internal motivation, and goal orientation. Therefore, trainings to improve the education and skills level of the B40 group to handle micro tasks is utmost important.

As a new industry, a strategic directions and guideline implementation need to be formulated. Therefore, the five strategic trusts formulated will be the foundation for the Strategic Framework of the micro sourcing industry development in Malaysia. The proposed framework is designed to ensure that the micro sourcing industry development will benefit all levels of society especially the B40 group. Although any individuals can become a micro worker, the opportunity to become micro workers should be targeted at B40 group with basic ICT knowledge and people outside the traditional workforce who want to increase their income.

The strategy is not only to attract and develop talents but also could leverage on the expected growth in global micro sourcing industry. With this right strategy and planning, Malaysia with its diverse population with multicultural background could be promoted as a regional micro sourcing hub to attract international firms outsourcing their micro tasks in Malaysia.

REFERENCES

[1] National IT Agenda – NITA. Retrieved Oct 2, 2012, from http://nitc.mosti.gov.my

[2] Prime Minister of Malaysia, 19 Oct 2011. Retrieved Oct 1, 2012, from http://www.mosti.gov.my/index.php?option=com_content&view=article&id=1778%3A19-october-2011-pm-chairs-23rd-icm-and-14th-msc-msia-iap-meetings&Itemid=64&lang=en

[3] New York, 17 May 2011, Digital Malaysia programme on track. Retrieved Oct 3, 2012, from http://www.btimes.com.my/Current_New...cle/index_html

[4] Multimedia Development Corporation (MDeC). 2012. Digital Malaysia Lab Report.

[5] Economic Planning Unit, 2010, Tenth Malaysia Plan 2011-2015. EPU. Prime Minister's Department, Putrajaya

[6] Nair, M and Vaithilingam, S. 2012. Innovative Use of ICT Among Urban-poor Communities: Challenges and Opportunities. *ICT Strategic Review 2012/13 Innovation fo Digital Opportunities,* pp 182-191.

[7] Arshad, N. H., Salleh, S.S., Aris, S.R.S, Janom, N., Mastuki, N., "Micro sourcing: The SWOT analysis on the demand, supply and platform". Science and Information Conference (SAI), October, London, 2013, pp. 768-773.

[8] Howe, J. (2006, June 2). Crowdsourcing: a definition. *Wired Blog Network: Crowdsourcing*. Retrieved Oct 1, 2012, from http://crowdsourcing.typepad.com/cs/2006/06/crowdsourcing_a.html

[9] Hoßfeld, T., Hirth,M.,and Tran-Gia, P. (2011) Modeling of Crowdsourcing Platforms and Granularity of Work Organization in Future Internet. *Proceedings of the 2011 23rd International Teletraffic Congress (ITC 2011)* p.142-149.

[10] AmazonMechanicalTurk, (accessed 29 Septembrer 2012), https://www.mturk.com/mturk/.

[11] CrowdFlower, (accessed 29 September 2012), http://crowdflower.com/.

[12] SamaSource, (accessed 29 September 2012), http://samasource.org/.

[13] Ushahidi , (accessed 29 September 2012), http://www.ushahidi.com/).

[14] Micro sourcing, (accessed 29 September 2012), http://www.micro sourcing.com/.

[15] ODesk , (accessed 29 September 2012), https://www.odesk.com/

[16] Crowd Sourcing Industry Report, Market Trends, Composition and Crowdfunding platform, 2012.

[17] Human Capital Connection (YourPartTime.com), 2012 (unpublished document)

[18] Jobs Malaysia, 2012, Kementrian Sumber Malaysia (unpublished document)

[19] SMECorp, 2012, Google to invest RM10mil to Help SMEs to Set Up Websites URL (accessed 19 February 2013), http://www.smecorp.gov.my

[20] Gartner Forecast Analysis, 2012, Forecast Analysis: IT Outsourcing, Worldwide, 2010-2015, 4Q11 Update

[21] Parikh, K. 2012, Impact Sourcing: Transforming Global Sourcing into a Social-Economic Endeavor. URL (accessed 21 December 2012), http://www.globalizationx.com/impact-sourcing-transforming-global-sourcing-into-a-socio-economic-endeavor-2/

[22] Syaripah Ruzaini Syed Aris, Norjansalika janom, Noor Habibah Arshad, Siti Salwa Salleh, Norazam Mastuki., "A Preliminary Study for Crtical Success Factors of Micro Sourcing in Malaysia". 3rd International Conference on Research and innovations in Information Systems, 2013 (ICRIIS'13). "in-press".

[23] Arshad, N. H., Salleh, S.S., Aris, S.R.S, Janom, N., Mastuki, N., "Micro sourcing: Micro Sourcing Strategic Framework for Low Income Group International Journal of Advanced Computer Science and Applications, 2013, Vol. 4, No. 6, pp. 97-105.

[24] Salleh, S.S., Arshad, N. H., Aris, S.R.S, Janom, N., Mastuki, N., "Formulating Cohesive Digital Ecosystem of Micro Sourcing Business Process in Malaysia". Science and Information Conference (SAI), October, London, 2013, pp. 321-327.

Lightweight Symmetric Encryption Algorithm for Secure Database

Hanan A. Al-Souly, Abeer S. Al-Sheddi, Heba A. Kurdi

Computer Science Department, Computer and Information Sciences College
Imam Muhammad Ibn Saud Islamic University
Riyadh, Saudi Arabia

Abstract—Virtually all of today's organizations store their data in huge databases to retrieve, manipulate and share them in an efficient way. Due to the popularity of databases for storing important and critical data, they are becoming subject to an overwhelming range of threats, such as unauthorized access. Such a threat can result in severe financial or privacy problems, as well as other corruptions. To tackle possible threats, numerous security mechanisms have emerged to protect data housed in databases. Among the most successful database security mechanisms is database encryption. This has the potential to secure the data at rest by converting the data into a form that cannot be easily understood by unauthorized persons. Many encryption algorithms have been proposed, such as Transposition-Substitution-Folding-Shifting encryption algorithm (TSFS), Data Encryption Standard (DES), and Advanced Encryption Standard (AES) algorithms. Each algorithm has advantages and disadvantages, leaving room for optimization in different ways. This paper proposes enhancing the TSFS algorithm by extending its data set to special characters, as well as correcting its substitution and shifting steps to avoid the errors occurring during the decryption process. Experimental results demonstrate the superiority of the proposed algorithm, as it has outperformed the well-established benchmark algorithms, DES and AES, in terms of query execution time and database added size.

Keywords—Encryption; Security; Protection; Transposition; Substitution; Folding; Shifting

I. INTRODUCTION

The tremendous development of technology and data storage leads organizations to depend on database systems. Organizations store huge amounts of data in secured databases in order to retrieve them in a fast and secure way. Some of the stored data is considered sensitive and has to be protected.

In the presence of security threats, database security is becoming one of the most urgent challenges because much damage to data can happen if it suffers from attacks and unauthorized access. With databases in complex, multi-tiered applications, attackers may reach the information inside the database. Damage and misuse of sensitive data that is stored in a database does not only affect a single user; but possibly an entire organization [1]. We can categorize the attackers into three types: intruder, insider, and administrator. Intruders are external people who infiltrate a database server to steal or tamper with data. Insiders are authorized users in a database system, who conduct some malicious works. Administrators can be database administrators (DBA) or system administrators

(SA), and both have absolute rights to database systems. However, if they are malicious, the security of the database may be damaged [2]. Insider and administrator attackers have gathered more attention in recent years because they can access a database without any effort, and they use important data in a wrong way. Database encryption has the potential to secure data at rest by providing data encryption, especially for sensitive data, avoiding the risks such as misuse of the data [1]. In order to achieve a high level of security, the complexity of encryption algorithms should be increased with minimal damage to database efficiency, ensuring performance is not affected.

There are many research studies in the database security field. Some of them have efficient implementations. Also, many encryption algorithms have been proposed, some of which have appealing features but still need further development, one such algorithm is the Transposition, Substitution, Folding and Shifting TSFS algorithm, known as the TSFS algorithm [1]. The TSFS algorithm provides a high degree of security, using a number of features. However, it supports only numbers and alphabetic characters that are not enough to protect different types of sensitive data. Another deficit of the TSFS algorithm is during the substitution and shifting processes where some errors occur during the decryption process.

This paper provides a secure and efficient encryption method that encrypts only sensitive data without using special hardware. It enhances TSFS algorithm by extending its data set to special characters, and corrects substitution and shifting processes, by providing more than one modulo factor and four 16-arrays respectively in order to avoid the error that occurs during the decryption steps. Moreover, this paper draws a comparison between the enhanced TSFS algorithm (ETSFS) and two other famous encryption algorithms, namely Data Encryption Standard (DES) and Advanced Encryption Standard (AES) algorithms, and evaluates their performance in terms of query execution time and database added size.

The remaining parts of this paper are organized as follows: section 2 reviews existing work on database encryption techniques. Section 3 introduces the ETSFS algorithm and explains its procedure, while section 4 introduces the implementation of the ETSFS algorithm and suggested structure. Section 5 presents a comparative study between the algorithms, evaluates performance, reports results and

discusses them. Finally, section 6 concludes with a summary of contributions and makes suggestions for future research work.

II.   RELATED WORK

Due to the important role that encryption techniques play in securing database systems, numerous algorithms have emerged with different techniques and performance. Bouganim and Pucheral proposed a smart card solution to protect data privacy; the owners of databases can access the data using a client terminal that is supported by smart card devices [3]. This proposed solution is considered as a secure and an effective solution, but it is complicated and expensive [1]. Database encryption greatly affects database performance because each time a query runs, a large amount of data must be decrypted. Therefore, [4] suggests that encrypting sensitive data only can provide the needed security without affecting the performance.

In [1] [5] [6] encryption algorithms were proposed depending on encryption of sensitive data only. Kaur et al. proposed a technique to encrypt numeric data only using a fixed data field type and length [5]. However, this algorithm does not support encryption of character data. Agrawal et al [6] also, proposed an encryption scheme for numeric data with an important feature that allows queries or any comparison operations to be applied directly on encrypted data sets without decrypting them. The scheme uses indexes of database over encrypted tables, but it is only applied to numeric data, additionally, it has not investigated key management. In some application, where the data is backed up frequently, we need to control the access to data and support multilevel access. So, Hwang and Yangb proposed a multilevel database encryption system with subkeys, which can encrypt/decrypt the whole table, column or row. Also, this system can encrypt each row with different subkeys according to a security class of the data element. This system is based on the Chinese Remainder Theorem [7].

The DES algorithm is one of the famous encryption algorithms that uses a symmetric-key to change 64-bit of a plain text into 64-bit of a cipher text, using 56-bit of the key and 16 rounds. [8]. It is, now, considered as insecure for many applications; this is mainly due to the size of key, which is too small [9]. The work in [10] presents the AES algorithm as a replacement for the DES algorithm as a standard for data encryption. It is a symmetric-key algorithm that takes 128-bit for the plain text and 128, 192, or 256-bit for the key, the length of the key specifies the number of rounds in the algorithm.

Finally, Manivannan and Sujarani [1] proposed efficient database encryption techniques using the TSFS algorithm, which is a symmetric-key algorithm. Its main features include using transposition and substitution ciphers techniques that are important in modern symmetric algorithms as they have diffusion and confusion.

Also, it encrypts only the sensitive data, so, it limits the added time for encryption and decryption operations. The algorithm utilizes three keys and expands them into twelve sub-keys using the key expansion technique to provide effective security for the database. In order to improve the security, this algorithm uses twelve rounds and two different keys in each round.

However, TSFS algorithm applies only to alphanumeric characters; it does not accept special characters or symbols. More details about the TSFS algorithm is provided in [11], which builds a system that generates different numbers of secret keys based on the TSFS algorithm along with other algorithms to ensure a high security level of encrypted data.

III.   PROPSED ALGORITHM (ETSFS)

The main objective of this paper is to enhance the TSFS algorithm [1] and accordingly to provide a high security to the databases whilst limiting the added time cost for encryption and decryption by encrypting sensitive data only. The ETSFS algorithm can encrypt the data that consists of alphabetic characters from A to Z, all numbers and the following symbols: (*, -, ., /, :, @ and _). The ETSFS algorithm is a symmetric encryption algorithm, meaning each transformation or process must be invertible and have inverse operation that can cancel its effect. The key also must be used in inverse order.

ETSFS algorithm uses four techniques of transformations, which are transposition, substitution, folding and shifting. Fig. 1 presents the encryption algorithm, where the decryption algorithm reverses the encryption algorithm. The following sections describe the four techniques and contain the algorithms in pseudo-code format to be easy to understand:

A. Transposition

Transposition transformation changes the location of the data matrix elements by using diagonal transposition that reads the data matrix in the route of zigzag diagonal starting from the upper left corner after getting the data and pads it with *s if it is less than 16 digits [1]. Fig. 2 shows the transposition process when the entered data was: 6923@domain.Sa, where Fig. 3 shows the transposition algorithm in encryption side, then, Fig. 4 shows the transposition algorithm in decryption side.

```
Algorithm encryption (String data,
                      Array[12] keys )
Pre: data is plain text.
     keys is array that contains 12 4x4-key matrices.
Post: encryptedData is data after encrypting.

        Matrix[4,4] dataMatrix;
        String encryptedData;
        if (data length < 16)
                padd data by adding *'s;
        else if (data length > 16)
                cut the data after 16;
        end if
        dataMatrix = data;
        key  = expandKeys (keys);
        for (int i=0; i<12; i++)
                dataMatrix = transposition (dataMatrix);
                dataMatrix = substitution (dataMatrix, keys(i),
keys((i+1)mod 12));
                dataMatrix = folding (dataMatrix);
                dataMatrix = shifting (dataMatrix);
        end for
        encryptedData = dataMatrix;
        return encryptedData

End encryption
```
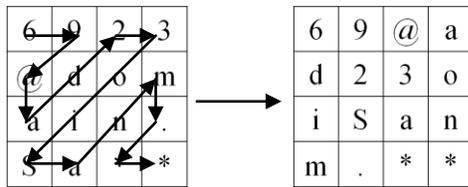
Fig. 1. Encryption algorithm.

Fig. 2.   Transposition example.

Algorithm transposition (Matrix data)
Pre: data is 4x4 matrix that contains the data should be encrypted.
Post: data is data after changing symbols location.

```
Matrix temp;
temp[0,0] = data[0,0];
temp[0,1] = data[0,1];
temp[0,2] = data[1,0];
temp[0,3] = data[2,0];
temp[1,0] = data[1,1];
temp[1,1] = data[0,2];
temp[1,2] = data[0,3];
temp[1,3] = data[1,2];
temp[2,0] = data[2,1];
temp[2,1] = data[3,0];
temp[2,2] = data[3,1];
temp[2,3] = data[2,2];
temp[3,0] = data[1,3];
temp[3,1] = data[2,3];
temp[3,2] = data[3,2];
temp[3,3] = data[3,3];
data = temp;
return data;
```

End transposition

Fig. 3.   Transposition algorithm.

Algorithm inverseTransposition (Matrix data)
Pre: data is 4x4 matrix, which contains the data should be decrypted.
Post: data is data after retrieving symbols location.

```
Matrix temp;
temp[0,0] = data[0,0];
temp[0,1] = data[0,1];
temp[0,2] = data[1,1];
temp[0,3] = data[1,2];
temp[1,0] = data[0,2];
temp[1,1] = data[1,0];
temp[1,2] = data[1,3];
temp[1,3] = data[3,0];
temp[2,0] = data[0,3];
temp[2,1] = data[2,0];
temp[2,2] = data[2,3];
temp[2,3] = data[3,1];
temp[3,0] = data[2,1];
temp[3,1] = data[2,2];
temp[3,2] = data[3,2];
temp[3,3] = data[3,3];
data = temp;
return data;
```

End inverseTransposition

Fig. 4.   Inverse transposition algorithm.

B. Substitution

The second algorithm is substitution transformation. It replaces one data matrix element with another by applying certain function [1]. If the element represents an alphabetic character, it then will be replaced with another character. If the element represents a number, it will be replaced with a number,

and if it represents a symbol, it will be replaced with a symbol. The encryption function [1] E for any given letter x is

$$E(x) = (((k1+p) \bmod M + k2) \bmod M \qquad (1)$$

Where p is the plain matrix element, k1 and k2 are the keys elements that have the same position of p, and M represents the size of modulo operation. The ETSFS algorithm takes three values for the modulus size instead of one value as in the TSFS algorithm. The described substitution process in [1] has confusion. Confusion happens if the data is composed of alphabetic and numeric digits, and the modulus size (M) will be 26 for any digit, as illustrated in the next example. If one element in the data was 4, k1=5, k2=5, M = 26, then the result of substitution process is 14 as the paper presents. This result causes two problems. The first problem, is that the length of the data will be changed and increased; for example, when the plan text size is 16 digits, the cipher text size will be 17 digits if one element only changes, and that contradicts the TSFS algorithm's feature. The second problem, since the inverse operation decrypts the data digit by digit also, is that then it will deal with each element in the cipher text individually (1 then 4). As a result, the decrypted data will be different from the data that have been encrypted. Therefore, the ETSFS algorithm gives M the following values: 26 if p is aliphatic, 10 if p is numerical and 7 if p is symbolic. The decryption function [1] D is:

$$D(E(x)) = (((E(x) - k2) \bmod M) - k1) \bmod M \qquad (2)$$

Since most of the programming languages such as Java and C++ deal with the modulus as the remainder of an integer division, some of the results may have minus sign, and this will create a problem because there is no data that have minus sign representation. So, one more step has been added to the ETSFS algorithm implementation to check if the result includes the minus sign, and then apply:

$$D(E(x)) = M - |D(E(x))| \qquad (3)$$

The following Fig. 5 shows the result of substitution. From the same example in fig. 5, if we implemented the decryption operation (2) on the first element, the result would be -4, so the ETSFS algorithm applies function (3) to get the correct result, which is 6. Fig. 6 and 7 show the substitution encryption algorithm and its inverse respectively.
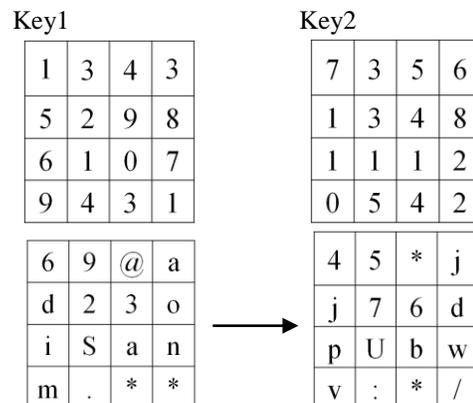

Fig. 5.   Substitution example.

```
Algorithm substitution (Matrix data,
                        Matrix key1,
                        Matrix key2)
Pre: data is 4x4 matrix.
     key1 and key2 are 4x4 matrix used to encrypt data.
Post: data is data after applying substitution encryption method.

        Matrix temp;
        int M;
        for (int i=0; i<4; i++)
            for (int j=0; j<4; j++)
                if (data[i,j] is alphabet)
                        M=26;
                else if (data[i,j] is number)
                        M=10;
                else if (data[i,j] is symbol)
                        M=7;
                end if
                temp[i,j]= (((k1[i,j]+ numric(data[i,j]) mod M)+k2[i,j]) mod M;
            end for
        end for
        data = temp;
        return data;

End substitution
```

Fig. 6.   Substitution algorithm.

```
Algorithm inverseSubstitution (Matrix data,
                               Matrix key1,
                               Matrix key2)
Pre: data is 4x4 matrix of data get from inverse Transposition technique.
     key1 and key2 4x4 matrix used to decrypt data.
Post: data is data after retrieving changes.

        Matrix temp;
        int M;
        for (int i=0; i<4; i++)
            for (int j=0; j<4; j++)
                if (data[i,j] is alphabet)
                        M=26;
                else if (data[i,j] is number)
                        M=10;
                else if (data[i,j] is symbol)
                        M=7;
                end if
                num=(numric(data[i,j])-k2[i,j]-k1[i,j]) mod M
                if (num<0)
                        num = M - |num|
                end if
            end for
        end for
        data = temp;
        return data;

End inverseSubstitution
```

Fig. 7.   Inverse substitution algorithm.

C. Folding

The third algorithm is folding transformation. It shuffles one of the data matrix elements with another in the same entered data, like a paper fold. The data matrix is folded horizontally, vertically and diagonally [1]. The horizontal folding is done by exchanging the first row with the last row. The vertical one is done by exchanging the first column with the last column. The diagonal fold is done by exchanging the inner cells, the upper-left cell with the down-right cell and the upper-right cell with the down-left cell. Fig. 8 shows the example after folding, while Fig. 9 shows the folding encryption algorithm. Next, Fig. 10 shows the folding decryption algorithm.
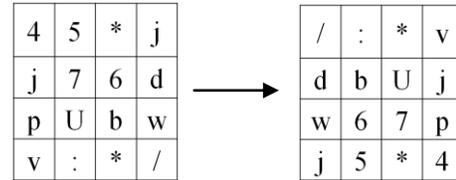

Fig. 8.   Folding example.

```
Algorithm folding (Matrix data)
Pre:  data is 4x4 matrix of data get from substitution technique.
Post: data is data matrix after applying folding technique.

        Matrix temp;
        temp[0,0] = data[3,3];
        temp[0,1] = data[3,1];
        temp[0,2] = data[3,2];
        temp[0,3] = data[3,0];
        temp[1,0] = data[1,3];
        temp[1,1] = data[2,2];
        temp[1,2] = data[2,1];
        temp[1,3] = data[1,0];
        temp[2,0] = data[2,3];
        temp[2,1] = data[1,2];
        temp[2,2] = data[1,1];
        temp[2,3] = data[2,0];
        temp[3,0] = data[0,3];
        temp[3,1] = data[0,1];
        temp[3,2] = data[0,2];
        temp[3,3] = data[0,0];
        data = temp;
        return data;

End folding
```

Fig. 9.   Folding algorithm.

```
Algorithm inverseFolding (Matrix data)
Pre:  data is 4x4 matrix of data get from inverse substitution technique.
Post: data is data matrix after applying inverse folding technique.

        Matrix temp;
        temp [0,0] = data[3,3];
        temp [0,1] = data[3,1];
        temp [0,2] = data[3,2];
        temp [0,3] = data[3,0];
        temp [1,0] = data[1,3];
        temp [1,1] = data[2,2];
        temp [1,2] = data[2,1];
        temp [1,3] = data[1,0];
        temp [2,0] = data[2,3];
        temp [2,1] = data[1,2];
        temp [2,2] = data[1,1];
        temp [2,3] = data[2,0];
        temp [3,0] = data[0,3];
        temp [3,1] = data[0,1];
        temp [3,2] = data[0,2];
        temp [3,3] = data[0,0];
        data = temp;
        return data;

End inverseFolding
```

Fig. 10. Inverse folding algorithm.

D. Shifting

The last part of the algorithm is the shifting transformation, which provides a simple way to encrypt using a 16-array element of numeric digits to exchange a letter with another. Each element of the array must contain the numeric representation of the data. Each digit must appear only once in

each element of the array. The digits can appear in any order [1]. In shifting process, the algorithm replaces each element in the data matrix by its position within its array element. The ETSFS algorithm uses four 16-arrays instead of one array as the TSFS algorithm uses, because the described shifting process in [1] has confusion. For example, if an element in the plain text is 4 and its position within the array is 15, then the shifting process in [1] returns 15, which is causing the same two problems that were described in substitution transformation. So, the ETSFS algorithm separates each type from other. The ETSFS algorithm uses four 16-arrays, one for numeric, one for symbols, but because it is difficult to enumerate all symbols in this project; the suggested ETSFS algorithm considers only two types of symbols. Symbols that are used in emails (-, ., @, _) and symbols that are used in IP addresses (/, :). The last two 16-arrays are used for alphabetic, where one for capital letters and the other for small letters. We used that to enhance TSFS algorithm and make it is sensitive for the type of letter. The process illustrated in Fig. 11. Fig. 12 and 13 show the shifting encryption algorithm and its inverse respectively.

The previous encryption process is considered as the result of the first round of the ETSFS algorithm. The output of the first round goes as an input to the second round and the output the second round goes as an input to the third round. This process continues up to the 12th round and the output of this round is the cipher text of the given plain text and that cipher text is stored in the database. For keys, in each round, it selects two keys for encryption. In encryption, each round (i) selects the key (i) and the key (i+1), at round 12 it selects key (12) and key (1). In decryption, the keys are selected in reverse order. The Fig. 14 shows the steps of expand keys as [1] suggested.

IV. IMPLEMENTATION

A Java-based project has been built to test the ETSFS algorithm correctness and performance. The implementation uses three-tier architecture, as represented in Fig. 15. The three-tier separate the functions into interface, processing and data management functions. The multi-tier architecture allows developers to create flexible and reusable applications. In addition, this architecture provides "encryption as a service" to facilitate the interaction between the interface and the encryption/decryption model, and makes the process of encryption or decryption transparent to application [2]. In this paper, the interface-tire is used to enter and retrieve data from the database. The processing-tier is used to garner the data or query from the interface-tier and then to complete the encryption or decryption processes to apply the query over the secure database. It stores the keys in a separate file instead of storing them in the database to increase the security. Finally, a data management-tire stores the data.

Depending on the suggested architecture, the implementation structure was developed as shown in Fig. 16. This Figure illustrates all classes and their connections with each other. It shows the attributes of each class and functions headers. The classes include:

A. Maic Class

In general, at the beginning, the user can enter the information that will be encrypted. In this implementation, the main class reads data from a file to obtain equivalent results when measuring the performance. The interface part is responsible for taking the data from file and sending it to the translator part to save it in the database. Another function for the interface is to retrieve the data form the database by using the translator part.

I/P	Array Element	O/P
/	0 1 2 3 4 5 6	/
:	1 2 3 4 5 6 0	/
*	2 3 4 5 6 0 1	@
v	3 4 5 6 7 8 9 10 11 12 13 14 15 … 23 24 25 0 1 2	s
d	4 5 6 7 8 9 10 11 12 13 14 15 16 … 24 25 0 1 2 3	z
b	5 6 7 8 9 10 11 12 13 14 15 16 … 24 25 0 1 2 3 4	w
U	6 7 8 9 10 11 12 13 14 15 16 … 24 25 0 1 2 3 4 5	O
.	.	.
.	.	.
4	1 5 4 6 0 7 2 8 3 9	2

Fig. 11. Shifting example.

Algorithm shifting (Matrix data,
                    MAtrix arrayNumber,
                    Matrix arrayAlpha,
                    Matrix arraySymbol)
Pre: data is 4x4 matrix of data gets from folding technique.
       arrayNumber is 16x10 dimension array used for numeric data.
       arrayAlpha is 16x26 dimension array used for alphabetic data.
       arraySymbol is 16x7 dimension array used for symbol data.
Post: data is data matrix after applying shifting technique.

```
  Matrix temp;
  char charOfData;
  loop from i=0 to i=3 do
        loop from j=0 to j=3 do
            charOfData = data[i,j];
            if (charOfData is number )
                  loop from k=0 to k=9 do
                      if (arrayNumber[(3xi)+i+j][k]== charOfData )
                                  temp[i,j] = k;
                                  break;
                            end if
                      end loop

            elseIf (charOfData is alphabet)
                  loop from k=0 to k=25 do
                      if (arrayAlpha[(3xi)+i+j][k]== charOfData )
                                  temp[i,j] = Alpha that have order (k);
                            end if
                      end loop
            else
                  loop from k=0 to k=6 do
                      if (arraySymbol[(3xi)+i+j][k]== charOfData )
                                  temp[i,j] = Symbol that have order (k);
                            end if
                      end loop
            end if
        end loop
  end loop
  data = temp;
  return data;
```

End shifting

Fig. 12. Shifting algorithm.

Algorithm inverseShifting (Matrix data,
                    Matrix arrayNumber,
                    Matrix arrayAlpha,
                    Matrix arraySymbol)
Pre: data is 4x4 matrix of data get from inverse folding technique.
       arrayNumber is 16x10 dimension array used for numeric data.
       arrayAlpha is 16x26 dimension array used for alphabetic data.
       arraySymbol is 16x7 dimension array used for symbol data.
Post: data is data matrix after applying inverse shifting technique.

```
  Matrix temp;
  char charOfData;
  int index;
  loop from i=0 to i=4 do
        loop from j=0 to j=4 do
            charOfData = data[i,j];
            if (charOfData is number)
                  index = the order of charOfData number in number order;
                  temp [i,j] = arrayNumber[(3xi)+i+j][index];
            elseIf (charOfData is alphabet)
                  index = the order of charOfData alphabet in alphabetical
order;
                  temp [i,j] = arrayAlpha[(3xi)+i+j][index];
            else
                  index = the order of charOfData symbol in symbol order;
                  temp [i,j] = arraySymbol[(3xi)+i+j][index];
            end if
        end loop
  end loop
  data = temp;
  return data;
```

End inverseShifting

Fig. 13. Inverse shifting algorithm.

Algorithm expandKeys (String [3] initialKeys)
Pre: initialKeys contains of Key1, Key2 and Key3 as initial keys.
Post: expandedKeys is array of size 12, contains 12 expanded keys Matrcis.

```
  Matrix[12] expandedKeys;
  Matrix[4,4] tempKey;
  for (int i=0; i<3; i++)
        if (initialKeys[i] length < 16)
            padd data by adding 0's;
        else if (initialKeys[i] length > 16)
            cut the data after 16;
        end if
        change initialKeys to number based on the position in alphabets a-z;
        tempKey = initialKeys[i];
        for (int j=0; j<4; i++)
            for (int h=0; h<4; i++)
                expandedKeys [j+ix4].row(h) = tempKey.row(h) after
shifting by (j+h)%4 times;
            end for
        end for
  end for
  return expandedKeys;
```

End expandKeys
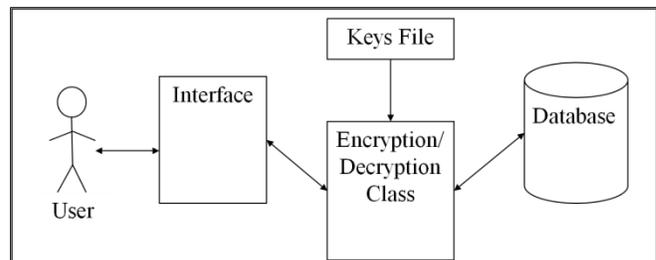
Fig. 14. Expand keys algorithm.
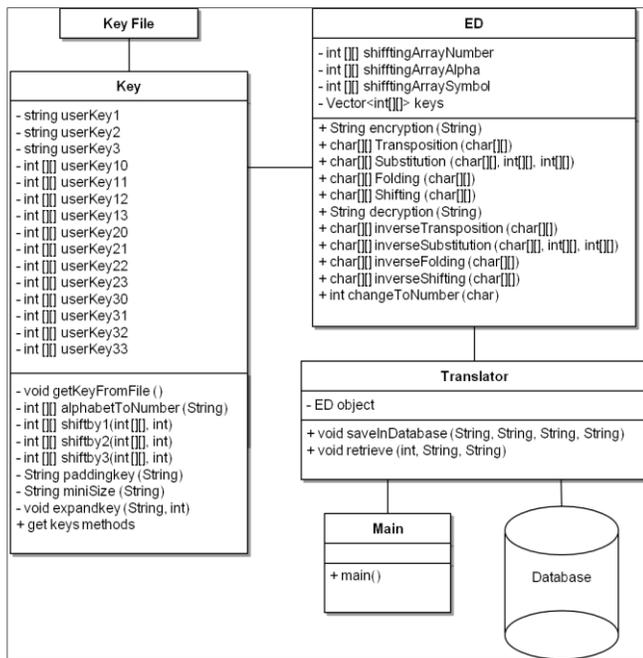

Fig. 15. Implementation architecture.

Fig. 16. Implementation UML diagram.

B. Translator Class

The middle part contains three parts. One part is used for keys; the other one, which is the translator, is used to deal with the database, and the last part is used for encryption/decryption operations. During the experiment, the translator class receives the data from the interface part by the SaveInDatabase() method, and draws a connection with the database so as to apply insert query in the database after encrypting the received data. Here the function depended on the application and the table used in the database. We used one table named "person". It contains four columns: name, phone, mail and job. The first three columns contained sensitive data. So, SaveInDatabase() method receives four data and encrypts the first three data by using the encryption/decryption (ED) class, then applies the insert query to add this data into the database.

Also, the translator has another method used to apply the select query from the database. In this project, we focused only on the encryption/decryption algorithms rather than how the query should be translated or mapped to a query that can be applied on the encrypted database. So, we just used two direct specific select queries for selecting the data, one for retrieving the complete table and the other for selecting a query depending on a condition which is the information of a person how has specific name. After retrieving the data from the database, the translator class decrypts it by using the ED class, and then printed it in the screen. In the second query, before the method does select operation from the database, this method must encrypt the data that is used in the condition by using the ED class to do the comparison operations over encrypted data.

C. Encryption/Decryption Class

This class is the most important class for this work. After the ED class receives the data from the translator class, it checks on some of points. It checks the length. The length must be equal to 16; if it is less than 16, the ED class pads the data

with (*) symbol up to 16 digits. In this project, we choose the (*) symbol for padding because it is not used in the data. If it is more than 16, the ED class deducts of the excess. After that, the ED class changes the data form to 4x4 array form to apply the algorithm on an array form. Then the ED class applies the same following scenario twelve times. First, it applies transposition transformation and changes the location of the elements in 4x4 data array by using diagonal transposition, and the result array of this step considers as input to the second step. In the second step, the ED class applies substitution transformation and replaces one data with another by using two keys in each round. Each element in the data array is one of three types: number, alphabet or symbol, and each of them has different modulo. If the element is alphabet or symbol, first it changes to a number based on the position in the alphabets a-z (a-0, b-1, ... z-25), then the ED class applies the equation of encryption on this number. We selected the symbols' positions as following (*, -, ., /, :, @, _) depended on precedence in ASCII codes. The result array of this step considers as input for the third step. In the third step, the ED class applies folding transformation and shuffles each element in 4x4 data array with another one in the same array, and the result array of this step considers as input for the fourth step. Fourth step, the ED class applies shifting transformation and replaces each digit of data array by its position within its array element. Each element in the data array is one of three types: number, alphabet or symbol, and each of them have different arrays set. There are four defined 16-arrays, one for numeric, two for alphabetic where one for capital letter and other one for small letter and last array for symbols. If the element in the data array is alphabet or symbol, first the ED class changes the element into a number based on the position in the alphabets a-z (a-0, b-1, ... z-25), then the ED class replaces the number with its position within its array set. The result array of this step considers as input for the first step again in the next round. After the twelve rounds finish and the data encrypted, the ED class sends the result to the translator part to store it in the database.

Also, the ED class has other function is responsible of decryption operations. This function receives the encrypted data from the translator part. The ED class first prepares the data into 4x4 array form to apply the algorithm on array form. After that, it sends the data array to the decryption methods in reverse sequence, started with inverse shifting, inverse folding, inverse substitution then inverse transposition. All of these methods follow the same scenario that applies in encrypted part but with inverse operations. This scenario also applies twelve times, the output of the each round considered as input to the second round. After the twelve rounds finish and the data decrypted, the ED class sends the result to the translator class.

D. Key Class

Key class reads three initial keys from the file. Then it checks on some points. Checking on the length, the length must be equal to 16. If it is less than 16, key class pads the key with 0s. If it is more than 16, key class deducts off excess. Then key class converts all digit of the keys to numbers based on position in the alphabets a-z (a-0, b-1, ... z-25). After that, it changes the keys form to 4x4 array form to apply the operations on array form.

After that, the keys are expanded based on shifting the rows into twelve keys, where each key is expanded into four keys. Each output key is stored in 4x4 array form, and the expanding technique implements as the following: the first output key has four rows where row 0 is not shifted, row 1 is shifted by one time, row 2 is shifted two times and row 3 is shifted three times. In the second output key, row 0 is shifted one time, row 1 is shifted 2 times, row 2 is shifted three times, and row 3 is not shifted. In third output key, are row 0 is shifted two times, row 1 is shifted three times, row 2 is not shifted and row 3 is shifted one time. In the last output key, row 0 is shifted three times, row 1 is not shifted, row 2 is shifted one time and row 3 is shifted two times.

V. COMPARATIVE STUDY

This section presents a comparative study between the DES algorithm, the AES algorithm and the ETSFS algorithms. It explains the experiment in order to evaluate their performance to establish the best algorithm amongst all possible algorithms. It then reports the results and discusses them.

The DES algorithm uses too short key length. Within the rapid advances, DES is breakable, for example in 1998 the Electronic Frontier Foundation built a DES Cracker for less than $250,000 that can decode DES messages in less than a week [12]. The AES algorithm is used as a standard for data encryption. It is stronger and faster than the DES algorithm [13]. DES and AES algorithms have open source code, and they are supported by Java libraries.

A. Experiment Setup

The experiment compared the ETSFS algorithm with the DES and AES algorithms, which have open source code. To test the algorithms, the following materials were used:

- Programming language: Java.
- Application platform: NetBeans IDE 6.9.1.
- Development: Java Development Kit (JDK) 1.6.
- Database management system: MySQL Server 5.6.
- Java external Library: Connector-java 5.1.23-bin.jar to connect the java with MySQL server.
- Visual database design tool: MySQL Workbench 5.2 CE used for database design, modeling and SQL development.
- Operating system: Windows Vista Home Premium, 32-bit.
- Hardware computer: Dell XPS M1330 laptop, Intel(R) Core(TM)2 Duo CPU T7300 2.00GHz and 3.00 GB for RAM.

In the experiment, to obtain fixed and fair comparison between the algorithms, same entered data and same functions that are responsible for accessing the database were used in all algorithms. Furthermore, each algorithm was tested with the following data size: 100, 500, 1000, 1500 and 2000 rows. For each size, the experiment was repeated three times and then the average value for each timer was calculated to eliminate the effect of the computer processing issues and insure near fair real value.

B. Evaluation Metrics

Execution time (Second): The evaluation performance of encryption/decryption processes conducted in terms of the execution time of insert and select SQL queries. In this project, we followed the same approach presented in [14] to calculate the execution time. We used a timer to calculate the execution time of the query from the beginning of its work until it finishes successfully. Three timers were used in each algorithm to calculate the execution times for three types of quires:

- **Insert:** We calculated the insertion execution time to know how much time does the insertion operation consumes with encryption processes. Insert query example: *INSERT INTO person VALUES (encrypted name, encrypted phone, encrypted mail, job).*

- **Select all:** To know how much time does the query take to select all the rows and decrypt the encrypted fields with decryption processes. Select all query example: *SELECT name, phone, mail, job FROM person.*

- **Select with condition:** To know how much time the query takes to encrypt the data in the condition, then compare this data with the encrypted data inside the database to retrieve the required data, and decrypt the encrypted selected fields with decryption processes. Select with condition query example: *SELECT name, phone, mail, job FROM person WHERE name = encrypted name.*

1) Database size (Kilobyte): One of the most important factors to determine about the database is its overall size [15]. So, we used the database size criterion to compare the database sizes between the three algorithms after storing the encrypted data. We used it to know the impact of the encryption process on the database overall size.

C. Results and Discussion

In this section, we stated the results of our experiments in terms of execution time and database size aspects. For each criterion, we presented the average values for each one of the three different algorithms with different sizes of data. Also, we interpreted the findings and results, and showed the relationships among the algorithms.

1) Execution time:
- **Insert queries execution time:** First, Fig. 17 shows the relationship between encryption time during insertion the data in the database and the number of tuples for each algorithm. In general, when the size of data increases, the time also increases. From the results, it is obvious that the AES algorithm consumes the longest time for encrypting and the ETSFS algorithm consumes the least time for that. The speed of the AES encryption depends on the number of rounds and the key generation in each time it does the encrypt function. Also, each round in the AES algorithm needs round key from the key expansion algorithm. So the AES algorithm depends on the speed of the round key

generation [16]. In contrast, the ETSFS algorithm needs the key generation only one time before doing the encrypt function. That is mean the ETSFS algorithm has the best execution performance when applying insert queries that include encryption processes.

- **Select all query execution time and select with condition quires execution time:** Fig. 18 shows the relationship between decryption time during searching in database with using a query to retrieve the complete table and the number of tuples, and Fig. 19 shows the relationship between decryption time but with using the queries that depend on a condition and the number of tuples for each algorithm. The DES algorithm consumes least time when applying the select queries. The instability of the relationship may occur owing to the time taken in connecting and disconnecting the database or file operations. Moreover, hardware issues may cause inconsistent relations. The AES algorithm consumes more time than the ETSFS algorithm for selecting queries because in the AES algorithm, the decryption module depends mainly on the key expansion module [16]. Also, the execution of the query that retrieves the complete table consumes little more time than those that depend on the condition because the former retrieves more tuples than the latter; this tuples need more decryption operations.

2) Database Size: In term of storage, Fig. 20 shows the relationships between the number of tuples and the size of the database in kilobyte after storing the encrypted data with the use of the three algorithms with different data sizes. Where the data size increases, the difference is evident between the algorithms. The results show that the ETSFS algorithm has consumed the smallest space among other algorithms. This happened because the size of the encrypted data in ETSFS algorithm does not increase more so than the original size; rather, it keeps the size as it is. However, AES and DES algorithms extend the data size to multiple of 128-bit and 64-bit respectively [12]. The AES algorithm needs largest storage space to store the data, followed by DES algorithm, which needs larger storage space than the ETSFS algorithm.

The experiment proves that the ETSFS algorithm can secure the data successfully when the data sets of the ETSFS algorithm are increased, and also that the substitution and shifting techniques are corrected. The insert queries execution time is not affected because the algorithm encrypts the sensitive data only, and the database size is not affected because the algorithm does not increase the size of encrypted data. In general, the improved performance comes without compromising query processing time or database size.
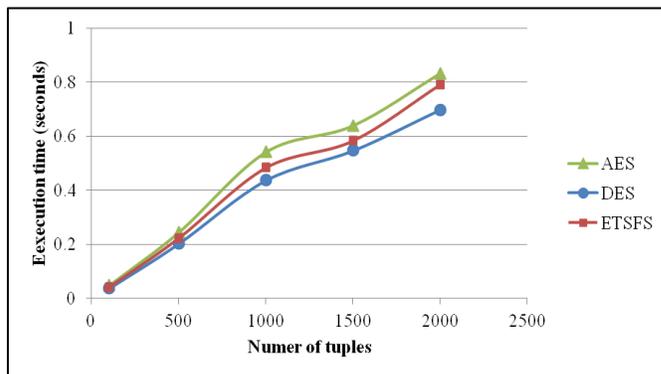

Fig. 18. The relationship between the execution time of select all rows operation and the number of rows for the three algorithms.
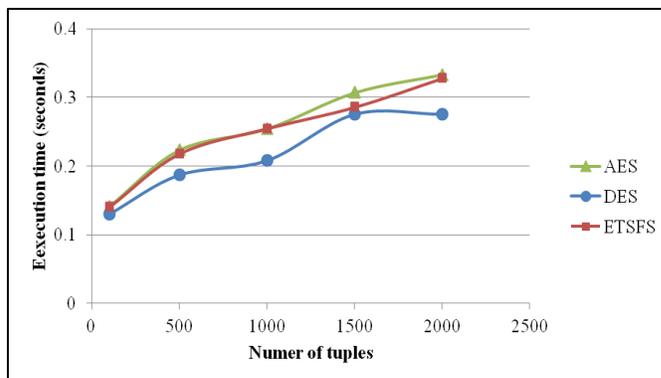

Fig. 19. The relationship between the execution time of select operations that depend on condition and the number of rows in the database for the three algorithms.
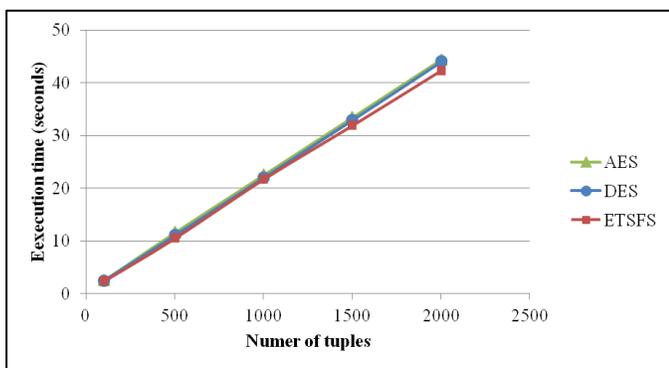

Fig. 17. The relationship between the insert operations execution time and the number of inserted rows for the three algorithms.
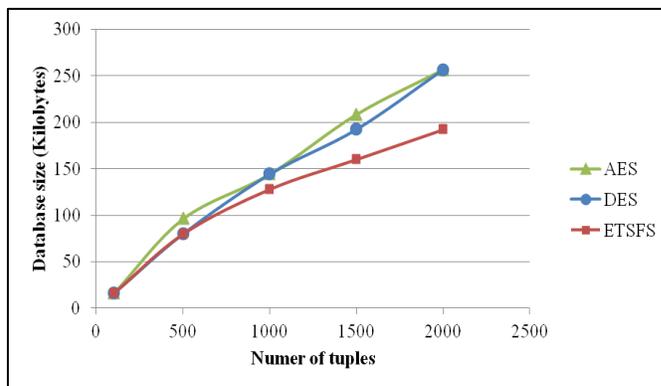

Fig. 20. The relationship between the database size after applying encryption and the number of rows for the three algorithms.

VI. CONCLUSION AND FUTURE WORK

Data-storing and exchanging between computers is growing fast across the world. The security of this data has become an important issue for the world. The best solution centred on securing the data is using cryptography, along with other methods. This paper proposes the enhancement of the TSFS algorithm to support the encryption of special characters, correct substitution process by providing more than one modulo factor to differentiate between data types and prevent increasing the data size, as well as correcting the shifting process for the same reasons by providing four 16-arrays. The experimental results have shown that the ETSFS algorithm successfully encrypted important symbols, as well as alphanumeric data. The improved performance comes without compromising query processing time or database size. Using well-established encryption algorithms as benchmarks, such as DES and AES, the proposed ETSFS algorithm was shown to have consumed the smallest space and encryption time compared to the other algorithms.

Due to time constraints, it was difficult to cover all special symbols in this paper; however, the ETSFS algorithm can be extended to include other symbols with slight modification to the encryption/decryption processes. For future work, it is intended that this algorithm be improved so as to accommodate any size of data, rather than only 16 digits. Furthermore, it is intended to further evaluate the security of ETSFS algorithm by establishing the number of operations and the time attackers need to recover the keys and accordingly hack the encrypted data.

REFERENCES

[1] D. Manivannan, R .Sujarani, Light weight and secure database encryption using TSFS algorithm, Proceedings of the International Conference on Computing Communication and Networking Technologies, 2010, pp. 1-7.

[2] L. Liu, J. Gai, A new lightweight database encryption scheme transparent to applications, Proceedings of the 6th IEEE International Conference on Industrial Informatics, 2008, pp.135-140.

[3] L. Bouganim, P. Puncheral, Chip-Secured data access: Confidential data on untrusted servers, Proceedings of the 28th International Conference on Very Large Databases, 2002, pp. 131-142.

[4] Z. Yang, S. Sesay, J. Chen, D. Xu, A Secure Database Encryption Scheme, Proceedings of Consumer Communications and Networking Conference, 2005, pp. 49-53.

[5] K. Kaur, K. Dhindsa, G. Singh, Numeric to numeric encryption: using 3KDEC algorithm, Proceedings of IEEE International Conference on Advance Computing, 2009, pp. 1501-1505.

[6] A. Rakesh, K. Jerry, S. Ramakrishnan, Order preserving encryption for numeric data, Proceedings of ACM SIGMOD International Conference on Management of Data, 2004, pp.563-574.

[7] M. Hwang, W. Yangb, Multilevel secure database encryption with subkeys, Data and Knowledge Engineering 22 (1997) 117-131.

[8] Z.Yong-Xia, The Technology of Database Encryption, Proceedings of the 2nd International Conference on MultiMedia and Information Technology, 2010, pp. 268-270.

[9] S. Bhatnagar, Securing Data-At-Rest, Literature by Tata Consultancy Services.

[10] I. Widiasari, Combining advanced encryption standard (AES) and one time pad (OPT) encryption for data security, The International Journal of Computer Applications 57 (2012) 1-8.

[11] M. Raj, M. Kumar, D. Manivannan, A. Umamakeswari, Design and development of secret session key generation using embedded crypto device-ARM-LPC 2148, Journal of Artificial Intelligence 6 (2013) 134-144.

[12] H. Alanazi, B. Zaidan, A. Zaidan, H. Jalab, M. Shabbir, Y. Al-Nabhani. New comparative study between DES, 3DES and AES within nine factors, Journal of Computing 2 (2010) 152-157.

[13] A. Mousa, E. Nigm, S. El-Rabaie, O. Faragallah, Query processing performance on encrypted databases by using the REA algorithm, The International Journal of Network Security 14 (2012) 280-288.

[14] A. Sterbenz, P. Lipp, Performance of the AES candidate algorithms in java, Proceedings of the AES Candidate Conference, 2000, pp.161-165.

[15] Oracle, Siebel Performance Tuning Guide, Available from: http://docs.oracle.com/cd/E14004_01/books/PDF/PerformTun.pdf 2012.

[16] T. Subashri, R. Arunachalam, B. Kumar, V. Vaidehi, Pipelining architecture of AES encryption and key generation with search based memory, The International journal of VLSI design & communication systems (VLSICS) 1 (2010) 1-13.

Advancing Research Infrastructure Using OpenStack

Ibad Kureshi, Carl Pulley, John Brennan, Violeta Holmes, Stephen Bonner, Yvonne James
School of Computing and Engineering
University of Huddersfield
Huddersfield, UK HD1 3DH
Email: hpc-rg@hud.ac.uk

Abstract—Cloud computing, which evolved from grid computing, virtualisation and automation, has a potential to deliver a variety of services to the end user via the Internet. Using the Web to deliver Infrastructure, Software and Platform as a Service (SaaS/PaaS) has benefits of reducing the cost of investment in internal resources of an organisation. It also provides greater flexibility and scalability in the utilisation of the resources. There are different cloud deployment models - public, private, community and hybrid clouds. This paper presents the results of research and development work in deploying a private cloud using OpenStack at the University of Huddersfield, UK, integrated into the University campus Grid QGG.

The aim of our research is to use a private cloud to improve the High Performance Computing (HPC) research infrastructure. This will lead to a flexible and scalable resource for research, teaching and assessment. As a result of our work we have deployed private QGG-cloud and devised a decision matrix and mechanisms required to expand HPC clusters into the cloud maximising the resource utilisation efficiency of the cloud.

As part of teaching and assessment of computing courses an Automated Formative Assessment (AFA) system was implemented in the QGG-Cloud. The system utilises the cloud's flexibility and scalability to assign and reconfigure required resources for different tasks in the AFA. Furthermore, the throughput characteristics of assessment workflows were investigated and analysed so that the requirements for cloud-based provisioning can be adequately made.

I. INTRODUCTION

Since 2010, the University of Huddersfield has established a private cloud. The primary aims for this system is to further the research goals in advanced computer systems. The cloud was quickly adopted by other academics and this system went from being a research machine to an integral part of the campus grid infrastructure. As a whole, this infrastructure is known as the Queensgate Grid (QGG), and includes: several high performance computing (HPC) clusters; a render farm; and a high throughput (HTC) cycle stealing system. Now this private cloud provides not only Infrastructure as a Service (IaaS) but platform and software services as well.

Heavily used in delivering courses like 'Network Security' and 'Operating Systems', the QGG-Cloud has served as a 'Virtual Laboratory'. With minimal hardware and very little technician time, the students were provided with a learning environment to meet their academic needs. Different flavours of the Microsoft Windows®and Linux operating systems were delivered over IP. This has meant that: the technicians do not have to exhaust time dual/triple booting lab machines; students can be given access to many different operating systems; and with more enhanced access privileges than if they were

using a physical machine plugged into the University network. Previously the specialist labs that could dual/triple boot were not available out of hours, due to their physical locations, but using the QGG-Cloud students can get access to their learning environment from the 24 hours labs within the library [1].

To help researchers and students undertaking research projects, different applications are placed in the cloud as snapshots of typical lab PCs. This system of creating snapshots also ensures backward compatibility of software. Even if new versions of the software or operating systems are released, researchers can still recreate previous conditions when proofing their prior research. Delivering Software as a Service in this manner allows the users to get different (in most cases higher) hardware specifications to do their science. Provisioning faster CPU's or larger memory configurations ensures the University and research capital is not wasted on an endless cycle of purchases. For most of their lifetime the cutting edge desktop machines are either idle, not being used for their intended purpose or not being used to their maximum capability [2]. The approach of centrally providing high memory type configurations leads to better access for all users within the institution, improved utilisation and better management of power hungry devices.

The QGG-Cloud in its current incarnation is an Open Stack Grizzly deployment using RDO over CentOS 6.4 [3]. The central head or access node provisions images on quad socket AMD based servers. These servers have 4GB of RAM per core. The system can provide up to 96 standard m2.medium configurations. Internally the cloud has two gigabit networks and externally is linked to the University's 10 gigabit research computing network.

In this paper we expand on work described in the paper 'Using OpenStack to Improve Student Experience in an H.E. Environment'[1]. In the previous work we briefly described how the private IaaS cloud supported our existing research computing infrastructure by providing machine configurations not available in the traditional HPC clusters. In Section II, of this paper, we describe how the challenges of dynamically surging our traditional HPC workloads to the cloud have been met. The ability to dynamically scale beyond the rigid hardware of the clusters was achieved by making enhancements to the open-source HPC batch system TORQUE and the job scheduler MAUI.

The growing numbers of students enrolling for University and online programming courses requires the use of automated evaluation of student assignments in order to meet that demand. Previous work[1] demonstrated a system (based on a custom Domain Specific Languages (DSL)) for specifying
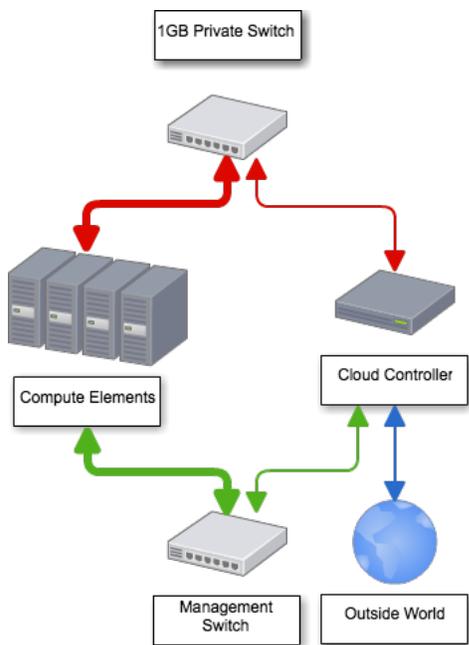
Fig. 1: Layout of the OpenStack based QGG-Cloud

Automated Feedback Assessments (AFA) and was delivered to academics using a Platform as a Service. We expand on that work here by examining an Akka actor[4] implementation of an assessment workflow. In Section III we use actual student submissions[1] along with the original submission frequency model, to simulate different loading scenarios for our assessment workflows. In using an actor based implementation, we additionally investigate the extent to which claims of scaling and failure resilience hold.

II. SURGE COMPUTING

A. Background

Cloud computing has been described as the next paradigm in computing. It promises to deliver computing power as the fourth utility direct to the end user. While large enterprises have yet to move their IT infrastructure into the cloud, startups and small businesses have benefited greatly from the cloud. The shorter barrier to entry means that companies can deliver their product or enhance their infrastructure as needed [5].

Where small to medium enterprises (SME) have successfully employed cloud computing is in scaling their existing infrastructure, as and when required. Whilst they are still reliant on some internal resources, in moments of heavy load the SME can scale their resources within a public cloud. The companies can therefore guarantee customers a high quality of service with minimal investment. This form of provisioning is known as surge computing [6].

Most current work is focused on surging Web 2.0 type workloads within the cloud [7][8][9]. Using load balancing and heartbeat monitoring tools, Apache web servers and mySQL databases are provisioned to meet such high demands.

---

[1]Our corpus of student data consists of 2164 submissions made by 163 students over a 210 day period.

There are many proponents of cloud computing whom additionally advocate migrating high performance computing workloads to the cloud [10][11][12]. However, there has been little existing work in creating a HPC Job Manager and Scheduler that is truly dynamic and elastic. The HTCondor project is a dynamic HPC scheduler, but it lacks elasticity, working on the principle that execution nodes can start up and then connect to the control node. The control node itself does not 'hunt' for execution endpoints. Even with such dynamism HTCondor can be utilised for HPC type workloads. It has been utilised by pharmaceutical companies to create large clusters on Amazon EC2 [13]. The major limit to the dynamism currently provided is that based on the workload on the head node there is no mechanism to automatically generate execution endpoints.

Traditional HPC Job managers are very rigid. Job managers like TORQUE and Grid Engine need to be given endpoint information, e.g. hostnames or IP addresses, at startup. If a change is required then the whole job management suite needs to be restarted. This obviously makes such schedulers inelastic. Recently however, IBM (in collaboration with Platform Computing) have released a version of the Platform LSF that dynamically creates nodes within elastic environments to meet such HPC needs. The Enterprise Edition of the MOAB HPC Suite now also includes dynamic cloud based provisioning functionality. Penguin Computing too have a cloud based initiative for enterprise clients [14].

Whereas the previous two solutions are proprietary, our efforts to surge enable the local HPC have solely utilised open source GNU/GPL job management systems.

B. Motivation

The University of Huddersfield's research computing grid comprises of several high performance compute clusters. They differ from each other by core speed, interconnect speed, co-processor availability and memory configurations. On average most systems have a 2GB RAM per core core memory configuration. The maximum available memory configuration is 4GB RAM per core, but these nodes cap off at 16GB RAM per node [15]. Researchers at Huddersfield do have access to Shared Memory Processor (SMP) and POWER systems based high memory configurations through academic partnerships with other institutions. Unfortunately, there are some applications within the QGG that are not able to run on non x86 systems. Others have license limitations that prevent the software from being used off-campus. Additionally, there are instances where researcher generated models can not be partitioned to run on a distributed memory clusters [16].

An example scenario has occurred with researchers from the school of Human and Health Sciences (HHS). Using a commercial *NIX based rendering package, the researchers generate high resolution visualisations of blood flow. Under most conditions the rendering package divides the frames between nodes and speeds up the render time. However, for the HHS renders, each single frame can consume more than the maximum available 16GB of RAM. Due to the commercial nature of the software package, the application can not be ported to other architectures, nor can it be deployed on a partner institutions HPC system. Before HPC and then cloud deployment of this package, HHS bought 8GB RAM

workstations, followed by upgrades to 16GB and then finally a single 32GB RAM workstation. The purchasing power of the research group has quite clearly been affected by a single computational problem.

The purchased workstations are power hungry in an idle state and, for the majority of the time, are used for standard office work! Provisioning for such applications within the QGG-Cloud allows for delivering different hardware configurations that the researcher could not possibly get under their desk. The key however is to remove the complexity of the cloud and to seamlessly integrate this system within the existing HPC infrastructure so that the researchers workflow is not significantly disrupted.

C. Implementation

In the initial development, we designed a wrapper script for PBS/TORQUE (TORQUE being the commonly deployed batch system within the QGG). Using Python extensions for Keystone, and a MySQL database to hold configurations, the wrapper script decides if a job should be directly submitted to the underlying torque system or surged to the QGG-Cloud.

In section II-C1 we outline the decision making process undertaken before jobs are surged to the cloud. In section II-C2 we provide detailed explanations for the mechanisms required to expand the TORQUE cluster into the cloud.

1) Decision Metrics: When a job is submitted to the cluster, the wrapper intercepts the job file. If the requested resources (in the job file) match those resources that can be found within the hardware scope (of the traditional cluster), then the job carries on through the system as normal. If the requested resources exceed what is available as bare metal, then the wrapper begins to provision a node in the cloud. Using information defined in the configuration database, along with the status of the load on the cloud, decisions are taken regarding the provisioning. A flow chart showing the steps is shown in Figure 2.

If the requested resources can fit on pre-defined cloud limits, then the wrapper initiates a compute node instance within the cloud and submits the job to TORQUE, along with a flag that ensures that this job will be routed to the new cloud instance. If the requested resources exceed the limits of the local cloud service, and the user/administrator has configured credentials to surge to a public cloud, then the VM is instantiated within the public cloud. In the event that there are no public cloud credentials, the job is returned to the user with an error message.

To maximise the utilisation efficiency of the private cloud, the wrapper is able to create a hardware flavour within cloud. This flexibility is essential in a private cloud setting. Generally provision options on clouds tend to be configured in binary increments. So after a 16GB RAM flavour the next flavour will have 32GB of RAM. Therefore if a user requests 20GB of RAM for their simulation there will be 12GB of RAM booked and not used. It will also limit the total number of surge instances created.

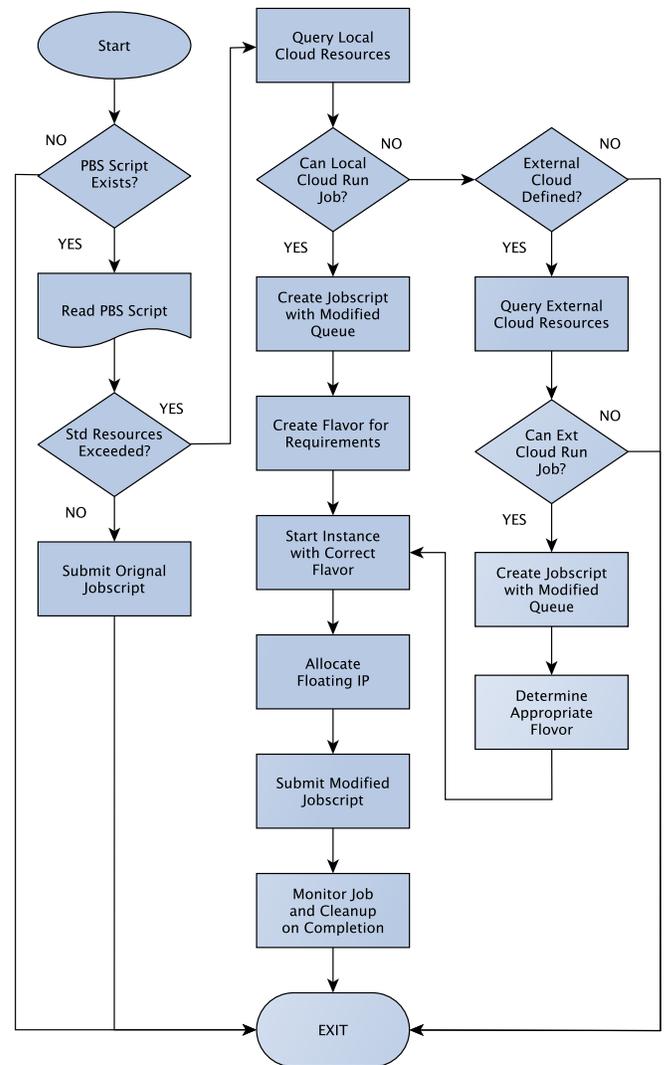The following section covers the actual methodology of implementation for the wrapper.

Fig. 2: Flowchart depicting the decision making process in HPC Cloud Surging

2) Surge Wrapper for TORQUE: Users on the surge enabled HPC system have a custom script named 'qsub' in their default paths. This qsub is not the standard TORQUE submission executable. The script passes the users submission to the surge wrapper. The surge wrapper parses the arguments supplied by the user to PBS/TORQUE. If a job is submitted with more than the job file argument, the wrapper automatically passes the job down to the scheduler. To invoke the surge a user has to define their job within a PBS job file.

Using PBS based functions, the wrapper queries the underlying HPC system to assess the available capabilities. Once it has ascertained that the requested resources are not natively available via TORQUE, it calls the function to surge the job into the cloud. Using the limits specified by the configuration database, the wrapper then generates a hardware flavour template. This is injected into OpenStack. The configuration database holds information such as: how elastic the system is; the largest flavour instance possible; etc.
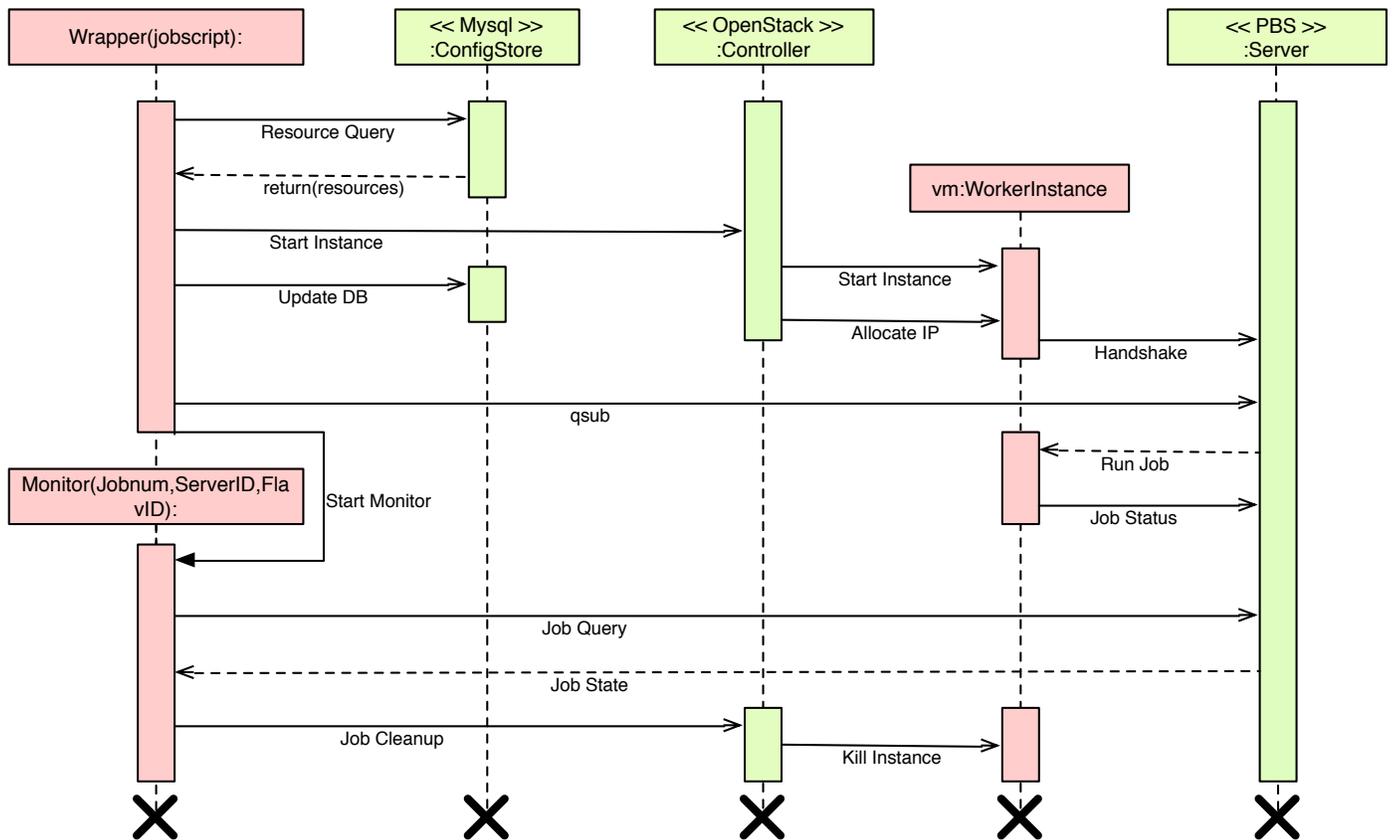
Fig. 3: Control bus depicting the sequence of events in an HPC Surge

With the hardware flavour injected into the system, the wrapper starts a virtual machine instance utilising a pre-configured image. This image mirrors a typical node within the HPC system. Next, the job script is regenerated with a special queue name and handed over to TORQUE. The wrapper also starts a small service to monitor the job's progress.

On the system TORQUE is pre-configured to recognise 'n' number of nodes (where n is equal to the maximum nodes that can be generated in the cloud). These nodes are locked to the special queue. So when the wrapper submits the job with the modified queue, TORQUE waits for the instance in the cloud to start up and then matches the job to the cloud instance. From the TORQUE point of view, a job has arrived which can only be assigned to a particular set of nodes and, one of those nodes has just become available.

Within the cloud, OpenStack generates a new hardware flavour and instantiates an image against this flavour. It then subtracts the requested resources from the pool of resources the surge can request. This is fed back to the surge wrapper to ensure the cloud does not get overly subscribed.

Upon completion of the job, the TORQUE MOM demon becomes idle and, once the job monitoring service detects this, injects a poison pill to terminate the VM instance. The custom hardware flavour is additionally removed from the OpenStack environment. Within TORQUE, the node becomes offline. Figure 3 depicts the sequence of events that take place when a job is surged to the private cloud.

Whether being used in a public or private cloud environment, the systems scalability will always be controlled. In a public cloud the availability of resources purchased (or some form of financial cap) will prevent an indiscriminate number of nodes being spawned. In a private cloud the elasticity itself is limited. OpenStack controls the number of cores and instances a user can run through group policies. In OpenStack parlance, this is know as 'Project Quotas'. In the situation where the surge 'project' on OpenStack is out of resources to spawn a new node, the new job is simply queued. If the new job, e.g. job-B, can run on already spawned virtual machines, then job-B is sent to TORQUE in the modified queue. The monitoring service that is awaiting the completion of already started/running jobs, is passed the new jobs information. Upon completion of the running jobs, the monitor refrains from injecting the poison pill, allowing TORQUE to push job-B to the now idle node. If the spawned instances do not meet the requirements of the new job, then job-B is held in a queue external to TORQUE. Running jobs will terminate as normal, and instances are brought down returning their resources to the quota. The job is then reconsidered for execution when the monitor reports a change to the load on the cloud.

If a requested hardware configuration can never be met, an error is returned to the user. Information pertaining to the maximum number of nodes, smallest configuration of nodes, largest configuration possible, static flavours (if using a public cloud) and credentials, are available to the wrapper from the MySQL configurations database.
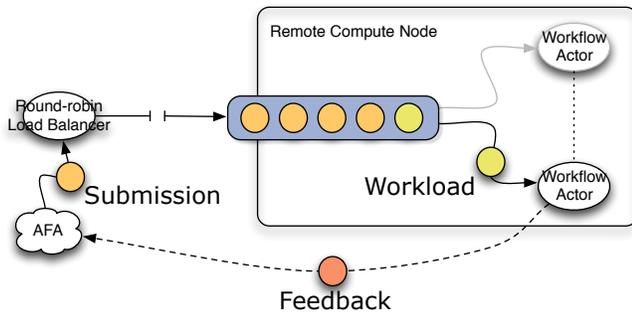
Fig. 4: Assessment Workflow

Fig. 5: Student Coursework Submission Frequency

The HPC Surge completed our institutional demands for Infrastructure as a Service. In the next section we evaluate the Automated Feedback Assessment framework, which is delivered to our users as Platform as a Service.

III. AFA WORKFLOWS

In [1] we investigated how the HPC cloud could be used to automatically generate formative feedback for submitted student assessments [2]. Using a combination of various technology stacks (i.e. JClouds, Chef, Apache Camel and Akka actors), [1] presented a domain specific language (DSL) that:

- could define complex assessment workflows

- was able to support a (extensible) variety of data collection, analysis and feedback delivery methods

- and simplified the deployment and configuration of the underlying computing infrastructure.

In this paper, we focus on the assessment workflows that AFAs encapsulate and investigate their design for throughput and scalability. Our assessment workflows are profiled using a corpus of student work[3] taken from a real-world (topological sorting) Java programming exercise.

A. Workflow Design

Assessment workflows are encapsulated by an AFA that is responsible for managing the consuming of coursework submissions, their injection into the assessment workflow and the subsequent processing and delivery of the (workflow returned) submission feedback[1]. As assessment workflows are the resource intensive component of an AFA, here we propose resilient and scalable designs for them.

[17] investigates how a combination of testing, control-flow graph similarity matching and software verification can be used to effectively assess and grade student code. Our work completes this work by providing a framework within which their assessment strategy, or workflow, may be run. Unlike [17], we base our work on a more complex programming exercise, and so face additional challenges in using software
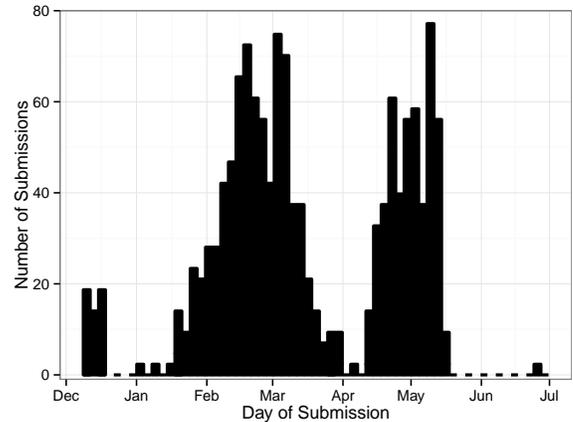
verification. Future work plans to explore the extent to which we may address these issues using Separation Logic based solutions such as: jStar[18] and Krakatoa[19].

Figure 4 shows our assessment workflow deployment strategy. Using a round-robin load balancing router, we send messages (i.e. submissions) to a series of remote cloud compute nodes. On each cloud compute node, messages are received and processed by a collection of worker actors that all share and process the same mailbox (configured via a balancing dispatcher). After a worker actor has processed a message, it replies with its result (i.e. feedback) to the message's original sender.

On going work is investigating how best to implement message persistence when cloud compute nodes fail or restart. Currently, we have used durable mailboxes (implemented using filesystems) to provide resilience against restarts. Future work will investigate using event sourcing[20] to rebuild system state by replaying submission events from message transaction logs.

B. Methodology

Our aim is to investigate the throughput characteristics of realistic assessment workflows so that requirements for cloud based provisioning can be adequately made. We first describe a workflow model and then how we plan to load it with submission data.

As reported in [17], program based assessments consist of three workload tasks: testing; control-flow graph similarity matching; and software verification. The resource requirements for each of these three tasks are dependent upon the specific submission and their processing may occur concurrently. Hence, we approximate our workflow model by simply performing testing only in our simulations. For our simulations, we configure our experiments to use a single (local) compute node, with 8 worker balanced workflows.

In order to measure quality of service, we constrain queued submissions so that they must be processed within a given time period (we choose a value of 5 minutes). Any feedback replies generated after the timeout period will be rejected. We implement this timeout using an Akka actor that proxies
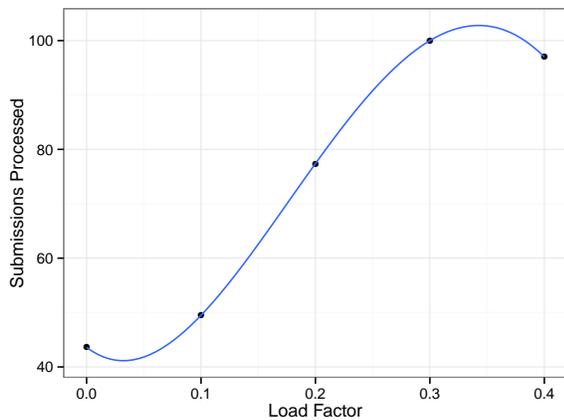
---

[2]All code is open sourced and publicly available from: http://github.com/carlpulley/cloud-paper.

[3]The corpus consists of 2164 submissions made by 163 students over the period of 201 days.

Fig. 6: Load Factor vs Submissions Processed

Load Factor	Submissions Processed (%)	Avg. Processing Time (s)
0.4	97.0	0.42
0.3	100.0	0.42
0.2	77.4	0.42
0.1	49.5	0.40
0.0	43.6	0.42

TABLE I: Resource Usage Data

the workflow actor using the implicit timeout of an *ask* or request/reply message pattern.

Different delivery loads are simulated using the actual submission frequency (see figure 5) of our real-world student data scaled with a constant *load* factor.

For each *load* factor, profiling experiments were ran for a fixed time of 15 minutes, with data collected at the termination of that experiment. All simulation code was ran on a 2.4GHz Quad-core Intel Xeon Mac Pro with 26GB of RAM, using a JVM configured with 256MB of stack and 1GB of heap space. Resource monitoring was performed using the Typesafe console. All other Akka actor configuration was set to the default values.

C. Data Analysis

By examining the AFA's message store, we extract the time of first submission, the last accepted feedback and the number of submissions and accepted feedbacks. From this data we can calculate (see table I) the percentage of submissions processed (our throughput) and a submissions average processing time (a quality of service measure).

With a load factor of 0.4, only 97.0% of submissions were processed. This is an artefact of the time period over which the experiments were run. If they had been run for a longer period (e.g. 20 minutes), then existing feedback processing would have completed and the expected 100% throughput for this load factor would have been achieved.

Throughout, our average processing time remains reasonably constant. Moreover, as we flood the actor queues with submission messages (i.e. the load factor gets smaller), the number of successful feedbacks generated decreases. This is expected behaviour. Using a graph plot of load factor vs

submissions processed (see figure 6), along with an estimate of acceptable feedback loss and compute node startup time, we can now predict the point at which more compute nodes should be launched. By monitoring our feedback message failure rates, our Akka actor workflow can be configured to elastically adapt to the changing workflow demands.

IV. FURTHER WORK

One of the biggest shortfalls with present job schedulers is their rigidity. The TORQUE batch queuing system needs to know about all execution endpoints and their capabilities before accepting a job. If a user requests more hardware than can be provided, the job is not accepted. This is why our implementation required a wrapper to sit above TORQUE assessing the requirements of incoming jobs and then provisioning for the demand. Our aim is to integrate the wrapper into a hybridised version of TORQUE. In this new elastic TORQUE, a job will not be rejected because of a lack of resources. Using the same principals of the surge wrapper, TORQUE will assess if it can initiate a node in the cloud to provision the required hardware. A job will only be rejected if the job's requirements can truly not be met.

The current implementation of the surge tool lacks interoperability with other cloud providers and middlewares. The final version of the wrapper tool will be more modular and have provisions to plug in to different cloud middlewares and providers. While current public clouds maybe too costly for large scale HPC provisioning, the public cloud can be used to surge time critical workloads to, or to get access to hardware configurations (such as GPU etc.) that are not available in house.

Within the University of Huddersfield's research computing grid, there is currently research taking place to make traditional HPC schedulers more intelligent [21]. Using an open-framework bench-marking suite known as the 'Application and System Performance Profiler', job schedulers are being given more information and are thus more aware of the performance characteristics of the software being executed on the HPC clusters. Heuristic information, gathered from previous jobs, is also being utilised in the scheduling decisions. Using the dynamic scaling features that the elastic-TORQUE project provides, the Intelligent scheduler can lead to cloud environments being utilised more efficiently for HPC workloads. Performance profiles will provide the HPC job manager with estimated resource requirements and the corresponding run times, so that the scheduler can factor in a new dimension to its scheduling algorithm: cost. This will truly open up the cloud for HPC and other research computing enterprises.

V. SUMMARY AND CONCLUSION

Cloud computing has great potential to provide scalable and flexible computational resources in commercial and academic environments. However, concerns regarding intellectual property rights and security risks, combined with the high cost of acquiring resources through a commercial provision have stimulated research into private clouds for education and research institutes.

In this paper we have presented the results of research in deploying and utilising cloud technology to create the QGG-Cloud at the University of Huddersfield, UK. This private

cloud is deployed as an OpenStack Grizzly using RDO over CentOS 6.4. It supports users needs for High Performance Computing resources by providing machine configurations not available in traditional HPC clusters, delivered as IaaS. To improve resource utilisation within the cloud we have implemented an HPC job manager and scheduler that are dynamic and elastic. This was achieved by making an enhancement to the open-source HPC system TORQUE and the job scheduler MAUI. Future work will focus on developing an intelligent scheduler to further improve utilisation of cloud environments for HPC workloads.

Responding to the demand for quicker and more insightful coursework feedback from computing students, we have implemented Automated Feedback Assessment (AFA) system within the university private cloud, delivered as PaaS. Based on the analysis of student coursework submission frequency, we have devised assessment workflow deployment strategy and investigated the throughput characteristics of the workflows. This has allowed workflows to be configured to elastically adapt to changing demands, and predict when more computing nodes should be launched in the cloud.

Building on the University's existing investment in internal HPC clusters and campus grid technology, and using open source cloud software - OpenStack, we have deployed the QGG-Cloud. This private cloud provides an effective solution for advancing the HPC research infrastructure and a flexible and scalable system used for research, teaching and assessment at the university.

References

[1] Bonner, Stephen, Pulley, Carl, Kureshi, Ibad, Holmes, Violeta, Brennan, John and James, Yvonne (2013). *Using OpenStack To Improve Student Experience in an H.E. Environment.*, Science and Information Conference 2013, London, pp.888-894. TheSAI

[2] Sundararajan, E.; Tan Bing Yean; Rahmani, M., *"Using computer labs for distributed computing,"*, 2011 International Conference on Electrical Engineering and Informatics (ICEEI), pp.1,6, 17-19 July 2011

[3] "Redhat Open Stack" RDO. Redhat, n.d. Web. http://openstack.redhat.org/.

[4] Wyatt, Derek (2013). *Akka Concurrency.* Artima Developer.

[5] Foster, I., Zhao, Y., Raicu, I., & Lu, S. (2008, November). *Cloud computing and grid computing 360-degree compared.* In Grid Computing Environments Workshop, 2008. GCE'08 (pp. 1-10). IEEE.

[6] Van den Bossche, R., Vanmechelen, K., & Broeckhove, J. (2010, July). *Cost-optimal scheduling in hybrid iaas clouds for deadline constrained workloads.* In Cloud Computing (CLOUD), 2010 IEEE 3rd International Conference on pp. 228-235. IEEE.

[7] Chieu, T. C., Mohindra, A., Karve, A. A., & Segal, A. (2009, October). *Dynamic scaling of web applications in a virtualized cloud computing environment.* In IEEE International Conference on e-Business Engineering, 2009. ICEBE'09. pp. 281-286. IEEE.

[8] Vaquero, L. M., Rodero-Merino, L., & Buyya, R. (2011). *Dynamically scaling applications in the cloud.* ACM SIGCOMM Computer Communication Review, 41(1), 45-52.

[9] Mao, M., Li, J., & Humphrey, M. (2010, October). *Cloud auto-scaling with deadline and budget constraints.* In International Conference on Grid Computing (GRID), 2010 11th IEEE/ACM. pp. 41-48. IEEE.

[10] Nurmi, D., Wolski, R., Grzegorczyk, C., Obertelli, G., Soman, S., Youseff, L., & Zagorodnov, D. (2009, May). *The eucalyptus open-source cloud-computing system.* In 9th IEEE/ACM International Symposium on Cluster Computing and the Grid, 2009. CCGRID'09. pp. 124-131. IEEE.

[11] He, Q., Zhou, S., Kobler, B., Duffy, D., & McGlynn, T. (2010, June). *Case study for running HPC applications in public clouds.* In Proceedings of the 19th ACM International Symposium on High Performance Distributed Computing. pp. 395-401. ACM.

[12] Bientinesi, P., Iakymchuk, R., & Napper, J. (2010). *HPC on competitive cloud resources.* In Handbook of Cloud Computing. pp. 493-516. Springer US.

[13] Brodkin, Jon. "$1,279-per-hour, 30,000-core Cluster Built on Amazon EC2 Cloud.", Ars Technica, 20 Sept. 2011. Website

[14] Bernstein, J. McMahon, K. (2012). *Computing on DemandHPC as a Service.* Penguin Computing

[15] Kureshi, Ibad. *The Queensgate Grid - Systems*, HPC Resource Centre RSS. HPC Research Group, University of Huddersfield, 04 Apr. 2010. http://hpc.hud.ac.uk/?page_id=446.

[16] Kureshi, Ibad. *The Queensgate Grid - Applications*, HPC Resource Centre RSS. HPC Research Group, University of Huddersfield, 04 Apr. 2010. http://hpc.hud.ac.uk/?page_id=175.

[17] Vujošević-Janičić, Milena, Nikolić, Mladen, Tošić, Dušan and Kuncak, Viktor (2013). *Software Verification and Graph Similarity for Automated Evaluation of Students' Assignments.* Information and Software Technology, Volume 55, Number 6, pp.1004-1016.

[18] Distefano, Dino and Parkinson, Matthew J. (2008). *jStar: Towards Practical Verification for Java.* Proceedings of the 23rd ACM SIGPLAN conference on Object-oriented programming systems languages and applications, OOPSLA '08, pp.213-226.

[19] Filliâtre, Jean-Christophe and Marché, Claude (2007). *The Why/Krakatoa/Caduceus Platform for Deductive Program Verification.* 19th International Conference on Computer Aided Verification. Lecture Notes in Computer Science, Volume 4590, pp.173-177. Springer.

[20] Evans, Eric (2003). *Domain-Driven Design: Tackling Complexity in the Heart of Software.* Addison Wesley.

[21] Kureshi, I., Holmes, V., & Cooke, D. J. (2012, September). *Robust Mouldable Scheduling Using Application Benchmarking For Elastic Environments.* In Local Proceedings of the Fifth Balkan Conference in Informatics (No. 920, pp. 51-57). University of Novi Sad, Serbia.

Modifying the IEEE 802.11 MAC to improve performance of multiple broadcasting of multimedia data in wireless ad-hoc networks

Christos Chousidis

School of Engineering and Design
Brunel University
London, UK

Rajagopal Nilavalan

School of Engineering and Design
Brunel University
London, UK

Abstract— **Multimedia applications over wireless networks have dramatically increased over the past years. Numerous new devices and applications that distribute audio and video over wireless networks are introduced every day and all of them demands a reliable and efficient wireless standard. Either operating as independent ad-hoc networks or as terminal parts of wired networks or the internet, wireless networks are frequently facing the need to broadcast multimedia data from multiple sources to multiple users. IEEE 802.11 standard (Wi-Fi) is the primary technology in wireless networking today. However, it has some inherited problems when it comes to broadcasting caused mainly by the lack of an acknowledgment mechanism. These problems do not allow the standard to take full advantage of the bandwidth offered by its latest amendments. In this paper two independent modifications of the medium access control (MAC) mechanism of the standard, are proposed along with the expanded use of the CTS-to-Self protection mechanism. The main objective of this study is to explore the ability of the modified MAC mechanisms to improve broadcasting performance while are operating in conjunction with a regular wireless network, and also to define the cases were the use of CTS-to-Self-protection mechanism can improve the overall performance of the network. The results show that the overall performance can be improved using these alternative MAC methods. Also, the cases where the CTS-to-Self technique can additionally contribute to the network performance are defined and analyzed.**

Keywords— *wireless networks; broadcasting; multimedia; linear increase of CW; EBNA; CTS-to-Self*

I. INTRODUCTION

The IEEE 802.11 standard (Wi-Fi) is the main technology in wireless networking and its use has an exponential raise [1]. Every day more and more devices and application are adopting this standard and many of them are using this networking platform to transmit multimedia type data [2][3]. Internet TV and radio, VoIP, video conferencing and telepresence, network gaming, and live audio networking are some of the recreational and professional applications that demand reliable and efficient wireless networks [4][5]. Many of those applications are using ad-hoc networking practices and broadcasting or multicasting transmission in order to stream their data. Broadcasting is a good practice in media networking because it can distribute simultaneously data to multiple users. IEEE 802.11 standard supports broadcasting without any type of feedback (e.g.

acknowledgment, ACK) from the recipients. Therefore, broadcasting does not provide any kind of delivery guarantee which in some cases, like live media transmission, is not critical as the retransmission of real-time data creates additional problems when it comes to delay and jitter. IEEE 802.11 implements a carrier sense multiple access mechanism, with collision avoidance (CSMA/CA). However, collisions are still happening and due to the lack of ACK, are leading to lost data. CSMA/CA mechanism uses a random backoff technique to reduce the probability of collisions and fairly arbitrate the wireless medium by allocating random waiting time values from a predefined contention window (CW) to every station intended to transmit. The lack of positive ACK of a transmitted packet causes an exponential increase of the CW and therefore a decrease in the probability of collision. In broadcasting however, as long as ACK is not implemented, the CW remains constant and always holds the minimum value. In the case of saturated networks and multiple broadcasting environments, the probability of collision is increased; affecting negatively the network's throughput [2]. An additional technique used by the CSMA/CA mechanism to arbitrate the medium access is the distribution of the network allocation vector (NAV) which contains information about the time that the network will be occupied by the station (STA), who recently gained access to the wireless medium. This technique demands an exchange of Request-to-Send, Clear-to-Send control messages (RTS/CTS) which again requires a unique recipient and therefore cannot be implemented in broadcasting. CTS-to-Self control message is an alternative to RTS/CTS process used in the cases were a CTS is not possible. It is broadcasted from a station with destination address its own address and lower transmission rate. This technique is strictly used as protection mechanism only for mixed-mode environments where extended rate physical (ERP-802.11g) and/or high throughput (HT-802.11n) devices coexist with legacy 802.11 technologies [6]. As long as ACK and NAV distribution is not implemented in broadcasting, saturated media broadcasting STAs are suffering from two main problems: large number of collisions and data loss due to buffer overflow caused by the excessively busy wireless medium.

To address the above problems two modifications of the IEEE 802.11 MAC mechanism are proposed in this paper. The amendments focus on two main areas, the random backoff algorithm and the NAV distribution. The IEEE 802.11 MAC is

reprogrammed to send a CTS-to-Self message prior to every broadcasting packet using the operational-high data rate. The random backoff process is also modified using two different approaches. In the first one, a linear increase of the CW is applied taking into account the changes in the number of broadcasting STAs. In the second approach, again a linear increase of the CW is performed according to the variation of the broadcasting STAs, but in this case an exclusive backoff number allocation (EBNA) algorithm is implemented. This algorithm allocate exclusive backoff numbers to each STA while maintain fairness in waiting periods between STAs over the time.

The main objective of this study is to explore the ability of the modified MAC mechanisms to improve broadcasting performance while are operating in conjunction with a regular wireless network. For this reason, the above modifications are tested in a "mixed" network where broadcasting and unicasting STAs coexist. The performance of the network is studied as a whole in order to investigate the possibility of different MAC mechanisms to function together.

The remaining of this paper is organized as follows: In section II, the 802.11 MAC process is summarized and the drawbacks of random backoff algorithm in the case of multiple broadcasting are analysed. In section III, the proposed modifications of the MAC process and the use of CTS-to-Self protection mechanism, are thoroughly described. In Section IV, the simulation's characteristics are analysed and comparative results are presented and commented while in Section V, the conclusions of this work are presented.

II. ANALYSIS AND DRAWBACKS OF IEEE 802.11 MEDIUM ACCESS MECHANISM

A. General description

The IEEE 802.11 MAC is mainly designed for wireless unicast communication and for unlimited number of users in the network. In Distributed Coordination Function (DCF) which is its primary medium arbitration method, Random Backoff in conjunction with virtual and physical carrier sense provides a level of protection from collisions. The 802.11 2007 standard provide an additional protection mechanism using RTS/CTS or CTS-to-Self control frames. This is mainly used for Network Allocation Vector (NAV) distribution in mixed-mode environments where different 802.11 technologies coexist. Although RTS/CTS it is used to address the hidden node problem, CTS-to-Self is used strictly as a protection mechanism for mixt-mode networks using data rates and modulation method that legacy 802.11 technologies can understand. NAV is distributed by setting the duration field of the control frame with the time in microseconds required in order for the two parties to complete transmission including ACK. It is clear however that there is no MAC-Level recovery mechanism in broadcasting [7]. In multimedia broadcasting the focus must be on preventing the loss of packets and the collisions instead of recovery and retransmission. NAV distribution is possible in broadcasting, only in mixed mode networks, by using the CTS-to-Self control frame [4]. CTS-to-Self is a standard CTS frame transmitted with a destination address of the transmitting station. The transmitting STA cannot hear its own transmission in a half-duplex medium but

all nearby STAs are alerted that a frame broadcast is pending and they can also update their NAVs with the value included in the duration field of the CTS-to-Self frame. As mentioned above, the use of CTS-to-Self is strictly limited in mixed-mode environments and it is using lower data rates that reduce throughput and increase delay. The possibility of modifying the 802.11 MAC to use CTS-to-Self as a main NAV distribution method, using also high data rates will significantly contribute to the performance of the protocol especially in broadcasting. However, the use of CTS-to-Self alone cannot eliminate the collisions occurrence which is caused by the drawbacks of 802.11 MAC Random Backoff mechanisms. This mechanism significantly contributes in collision avoidance but cannot totally eliminate them, especially when the number of STAs increases. In heavy data loads, there is a high likelihood that two or more STAs will choose the same backoff value. In this case the collision cannot be avoided regardless of the use of CTS-to-Self. For this reason an alternative EBNA algorithm can be used to overcome the Random Backoff algorithm drawbacks in the case where multiple broadcasting is taking place

B. Analysis of IEEE 802.11 MAC algorithm

IEEE 802.11 MAC Layer is the lowest part of the Link Layer and it is placed between the Physical (PHY) and the Logical Link Control (LLC) sub-layer. MAC architecture is based on two basic coordination functions, Point Coordination Function (PCF) and Distributed Coordination Function (PCF). PCF is a contention free access method which provides polling intervals to allow uncontended transmission opportunities (TXOP) for participating STAs. This function is outside the scope of this paper, first because it demands the use of an AP and second, because the manufacturers never implemented it into their devices. In this study the fundamental DCF contention-based access mechanism is used.

DCF's timing diagram is illustrated in figure 1 and its function is described as follow. A STA with a packet to transmit waits for the channel to become idle. When an idle period equal to DCF Inter-Frame Space (DIFS) is detected, generates an initial Backoff time value. This value indicates the period that the STA has to additionally defer before transmitting. The random Backoff process is the most important mechanism used in IEEE 802.11 CSMA/CA to prevent collisions. CW increases exponentially for every retransmission, (unique per station). Under low utilization, stations are not forced to wait very long before transmitting their frame. If the utilization of the network is high, the protocol holds stations back for longer period of times to avoid the probability of multiple stations transmitting at the same time. When we are referring to Contention-Based access, random Backoff is actually the primary mechanism for contention. This value is extracted from the following formula:

Backoff_Time = INT (CW x Random (0, 1)) x aSlotTime   *(1)*

Random (0, 1) is a pseudo-random number between 0 and 1 drown from a uniform distribution. CW is an integer within the range of values CWmin and CWmax. CWvalues=2x-1 (x starts from an integer defined by the station and goes up to 10). For example, for x=4, CW4=24-1=15, CW5=31, CW6=63 ….. CW10=1023. The aSlotTime duration is the value of the

correspondingly named PHY characteristics. The Backoff timer is decremented with one slot as long as the channel is idle. When a transmission is detected, the Backoff timer freezes and start to decrease again when the channel is sensed idle for a DIFS. When the timer reaches zero the data packet is finally transmitted.
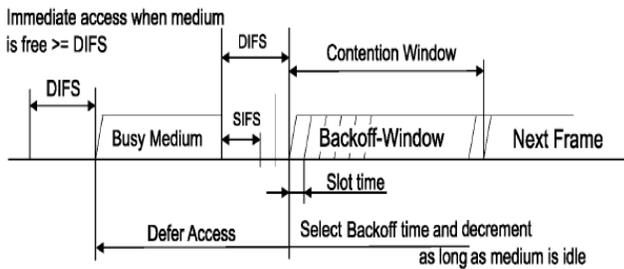

Fig. 1.   IEEE 802.11 basic access method

C. Drawbacks of random backoff in 802.11 broadcasting

There is plenty of research on the Reliable Broadcasting over wireless ad-hoc networks and many protocols have been proposed [8] [9] [10]. These protocols can be divided into four main categories according to the methods they use.

1) Simple Flooding Methods: Requires each node to retransmit all packets

2) Probability Based Methods: Use some basic understanding of the network topology in order to assign a probability to a node to rebroadcast.

3) Area based Methods: Rebroadcasting is based on the possible additional area that will be covered.

4) Neighbour Knowledge Methods: Maintain a state of neighbours, obtained by "Hallo" messages. This stage is used in the decision to retransmit

All the above methods require a sort of retransmission which is unsuitable for media networking. Reliability in media broadcasting is reduced by the drawbacks of random backoff process, which cause channel access delay and collisions no matter the available bandwidth of the wireless technology that is used.

The IEEE 802.11 standard defines that the CW size exponentially increases for each retransmission attempt of the same packet. However, as there is no retransmission in broadcasting, the CW size always holds the CWmin value. Under high utilization due to the increasing number of STA and/or high data production, CWmin appears to be extremely small. In this case we are facing two major problems. The first one is that it is possible for a STA that just completed a transmission and has a new packet to send, to choose zero as its initial backoff time and start transmitting immediately after a DIFS. As we can see from (1), backoff time is a random outcome based on a uniform distribution but its range increases proportionally with the size of CW. This consecutive transmission will give other STAs no chance to backoff. This problem is refereed as the backoff counter consecutive freeze process (CFP), and was extensively analysed by Xianmin Ma and Xianbo Chen in [11]. They show that the solution would be the ability to increase CW in broadcasting. The second and

most significant problem in the case of wireless media broadcasting is that there is a high likelihood for two or more STAs to choose concurrently equal backoff value. It is easy to understand that when we have fifty or more STAs producing continuous data and they are performing the backoff process using a CW=15 (like in 802.11g & 802.11n) this is highly possible. In this case a collision is occurring and a data packet is lost as there is no recovery mechanism and no time for retransmission.

For a given size of CW and "N" number of STAs broadcasting in a wireless network, the probability "*p*" of collision is given by the equation:

$$p = 1 - (1 - \frac{1}{CW})^{N-1} \qquad (2)$$

Figure 2 shows the calculated and simulated probability of collision in a multiple broadcasting environment using various CW sizes. A series of simulations performed for this study using OPNET modeler. The IEEE 802.11 MAC was modified to allow broadcasting using different than the minimum, CW sizes. CW size takes the values: 16, 32 and 64. The simulation parameters are: Packet Size 1024 bytes, Physical Characteristics 802.11g-24Mbps, Bit Rate 400 Mbps. As we can see from figure 2, for a CW=16 which is the broadcasting CW size according to the standard, the probability of collision for saturated network of 15 broadcasting STAs reach the value of 50% and approaches 100% when 50 STAs are broadcasting simultaneously.
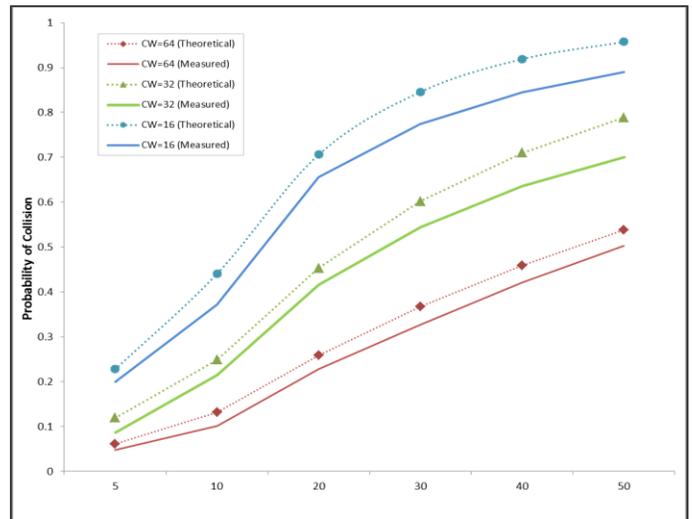

Fig. 2.   Probability of Collision

III.   MODIFIED IEEE 802.11 MAC MECHANISM

As it mentioned earlier in this paper in order to override the inability of the 802.11 protocol in handling multiple broadcasting media data, a modified MAC mechanism is proposed. The amendments focus on two main areas, the NAV distribution and the random Backoff algorithm.

A. Linear increase of CW

According to the IEEE 802.11 standard, in broadcasting the CW remain unchanged and always hold its minimum

value. This results in busy networks for broadcasting STAs to significantly increases the probability of collision (Fig 2). In this modification, the CW dynamically change according to variation of the broadcasting STAs in the network. For this purpose a variable (*No_of_STAs*) indicating the number of broadcasting STAs is created. The number of Backoff slots is a random value selected using a uniform distribution from a CW which is given by the formula using:

$$CW = CWmin + No\_of\_STAs \quad (3)$$

B. Exclusive Backoff Number Allocation algorithm (EBNA)

In order to prevent STAs from choosing similar Backoff numbers which leads to a collision an EBNA algorithm is implemented [12] [13]. This algorithm increases the size of CW according to the number of broadcasting STAs in the network. It is also designed to maintain fairness while allocating exclusive Backoff values for each transmission attempt. In order to do this, the algorithm needs two external variables, the total number of STAs in the wireless network (*No_of_STAs*) and the Station ID (*STID*) that every STA obtains upon joining the network (fig:3, line 1 & 2). The CW is always given by:

$$CW = No\_of\_STAs * 2 \quad (4)$$

The algorithm divides the CW in two equal groups. Values in the groups are allocated as follow:

$$group1 \leqslant No\_of\_STAs/2 \quad (5)$$
$$group2 > No\_of\_STAs/2$$

For each transmission attempt a random value between 1 and 2 is generated in order to select one of the two groups (fig:3, line 3). If group1 is selected the algorithm allocates to the STA a Backoff value equal to its STID, in other case the value given by the algorithm it is a projection of the STID value to group 2 and it is given by the formula:

$$Backoff\_slots = [(No\_of\_STAs * 2) - STID] + 1 \quad (4)$$

For a network with 10 STAs the station with STID=2 the *Backoff_slot* variable will take randomly one of the backoff values 2 or 19 while a station with STID=6 will take the Backoff value 6 or 15 (fig. 4). The pseudo-code describing the above process is illustrated in figure 3.

```
1.....No_of_STAs=10     % total number of stations
2.....STID=2            % current station
3.....Group=rand (1,2)
4.....if (group=1)
5...........Backoff_slots=STID
6.....else
7............ Backoff_slots=No_of_STAs*2-STID+1
8.....end if
```

Fig. 3. Exclusive Backoff Number Allocation algorithm pseudo-code
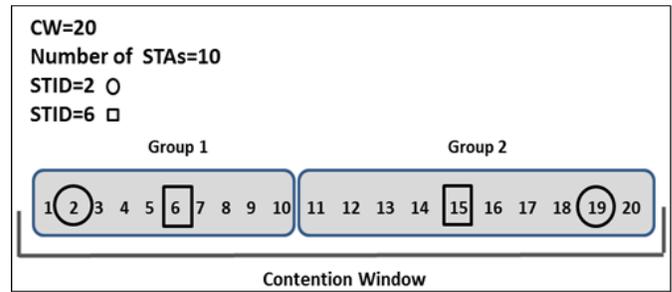

Fig. 4. EBNA example

Figure 5 shows a 3D snapshot of the Backoff number allocation process for both modifications, while figure 6 shows the fairness of the EBNA algorithm over the time as it is illustrated for to different broadcasting STAs during the same simulation by presenting the mean average of Backoff values for these STAs. As we can see, when the network reaches the steady state (approximately 5 sec) the average waiting time is equal for both STAs.
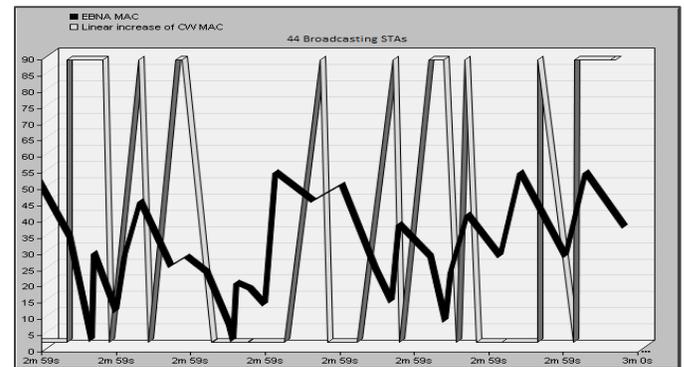

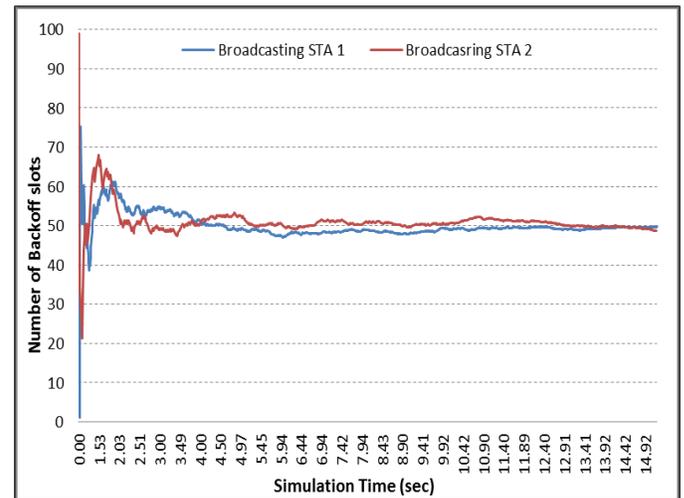Fig. 5. backoff number allocation process

Fig. 6. Fairness in the long run, using the EBNA algorithm

C. CTS-to-Self modofication

CTS-to-Self control message is an alternative to RTS/CTS process currently used in broadcasting only in cases where legacy technologies coexist with an ERP (802.11g) or HT (802.11n) physical, (mixed-mode networks). It is achieved by sending a CTS-to-Self control frame in appropriate (usually lower) data rate and modulation that all STAs can understand. CTS-to-Self frame contains in its "duration" field the time that all non-transmitting STAs must defer before trying to access the medium.

In order for the CTS-to-Self to be used as the main protection mechanism in broadcasting, two major modifications have been done in the 802.11 MAC algorithm. First, while the structure of the CTS-to-Self packet format remained as it was described in the standard; the MAC has been reprogrammed in order to transmit a CTS-to-Self control message prior to every data transmission. Thus, when a STA complete its random Backoff countdown and finds the medium idle, instead of broadcasting a data packet, sends a CTS-to-Self control packet which contains information concerning the time period that the medium will be reserved due to the forthcoming data packet transmission. In the unfortunate event that another STA completes the random Backoff simultaneously, the two CTS-to-Self packets are colliding causing a jam in the network lasting significantly less than a data packet collision (Fig 7).
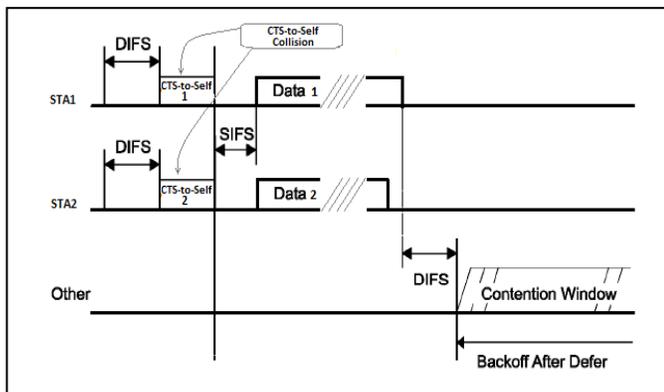

Fig. 7. CTS-to-Self and data Collisions

In any other case the STA waits for a sort inter frame period (SIFS) and transmits the data packet. Another modification in the MAC algorithm, which has been done in this project, was to reprogram the transmission rate of the CTS-to-Self message. CTS-to-Self transmission parameter has been modified to always adjust with the selected data rate used for data transmission.

IV. SIMULATION CHARACTERISTICS AND RESULTS

A. Simulation characteristics

The network simulation platform used in this study is OPNET Modeler 17.1. The simulation is based on IEEE 802.11g PHY, with a bit rate of 54Mbps. The topology is based on an ad-hoc network in a single BSS, with 56 unicast STAs located in the middle and broadcasting STAs randomly surrounding the unicast group in a 50x50m surface. The number of broadcasting STAs is gradually increased from 4 to 44, taking sequentially the values 4, 8, 16, 24, 34, and 44. The

simulation duration is 3 min. This is enough time for the system to reach its steady state. Three separate simulations have been conducted where all stations were relocated and also a different seed number has been set during the simulation execution. The presented results are the average values, in those cases where significant differences occurred.

All data traffic generation parameters for unicast and broadcast traffic are listed in table 1.

TABLE 1: TRAFFIC GENERATION PARAMETERS

Unicast traffic	
Start Time	Normal Distribution (0.5, 0.1)
On-State	180 sec
Off-State	0 sec
Interarrival Time	Normal Distribution (0.1, 0.005)
Packet Size	2200 bytes
Broadcast traffic	
Start Time	Normal Distribution (1, 0.01)
On-State	180 sec
Off-State	0 sec
Interarrival Time	Constant (0.0243)
Packet Size	1100 bytes

Figure 8 shows the network configuration for a population of 56 unicast and 44 broadcast STAs. The resulting load transmitted by each broadcasting STA is approximately 370Kbps while unicast STAs are transmitting with a bit rate of 77Kbps. The 56 unicast STAs are sending data to randomly chosen destinations including the existing broadcasting STAs. That cause to the broadcasting stations an additional load due to the ACK messages they have to transmit when a unicast packet is addressed to them. All broadcasting STAs are set to unlimited buffer size.
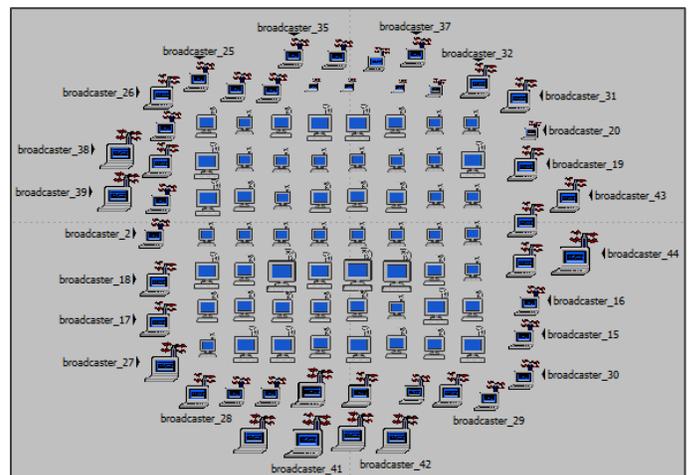

Fig. 8. Network configuration

B. OPNET code modifications

All modifications to the standard IEEE 802.11 MAC were made using OPNET Modeller network simulation environment. OPNET modeller is a powerful simulation tool which allows users to have full access to the executed code and

gives the ability to create and modify complex communication protocols. It has its own C++ library and it is using state machines to design and implement processes. For our implementation the OPNET wireless station node model is used.

a) Random Backoff modifications

Both cases described in III A and B, are implemented in OPNET by modifying the *'wlan_dispatch'* process and the *'wlan_mac'* child process. In the *wlan_mac* process (BKOFF_NEED state), the Backoff slots allocation algorithm has been changed according to the above described logic. The *No_of_STAs* variable is entered manually in the source code. For the *STID* parameter, a custom attribute is created and it is set during the "run" process. It is important to mention here that OPNET use the same *"wlan_dispatch"* for all models running the IEEE 802.11 standard. In our case, two MAC algorithms has to be used simultaneously; the classic 802.11 MAC used by the unicast STAs and the modified 802.11 MAC used by the broadcasting STAs. For this reason, a modified *'wlan_dispatch'* was also created.

b) CTS-to-Self modifications

A number of modification have been made mainly in *Function Block* of the *wlan_mac* process in order to force this modified child process to create a CTS-to-Self message prior to each data packet transmission using also the operational bit rate which in this case is 54 Mbps.

C. Organising the simulation

In order to study the performance of the modified MAC algorithms proposed in this paper and also to investigate the effect of the modified CTS-to-Self protection mechanism, five independent simulations have been carried out for each increase of the broadcasting STAs population. The number of unicast STAs as well as their traffic generation parameters remains unchanged during all simulations. Broadcasting STAs are altering their MAC algorithm according to the following combinations:

- Broadcasting using the classic IEEE 802.11 MAC
- Broadcasting using the Linear Increase of CW without CTS-to-Self
- Broadcasting using the Linear Increase of CW with the use of CTS-to-Self
- Broadcasting using the EBNA algorithm without the use of CTS-to-Self
- Broadcasting using the EBNA algorithm with the use of CTS-to-Self

D. Results

a) Throughput

Figure 9 shows the throughput performance for the five different simulations described above. Throughput is measured as the overall load of successfully transmitted data. It is important to note that in the case of broadcasting, a successfully transmitted packet it appears to be measured multiple times, depending on the number of recipients in the wireless network.

Thus, if n is the total number of STAs in the network and Ai is the successfully transmitted load from each STA, the overall measured throughput is given by the formula:

$$Throughput = (n - 1) \sum_{i=0}^{n} Ai \qquad (6)$$

As it is shown from figure 9, when the number of broadcasting STAs is small all methods are performing equally. When the number of Broadcasting STAs increases all modified medium access processes are performed slightly better than the classic 802.11 MAC. As it is expected, the EBNA algorithm gives the best results because it guarantees that there are no collisions between broadcasting STAs. All collisions happening at this stage are between broadcasting and unicasting traffic.
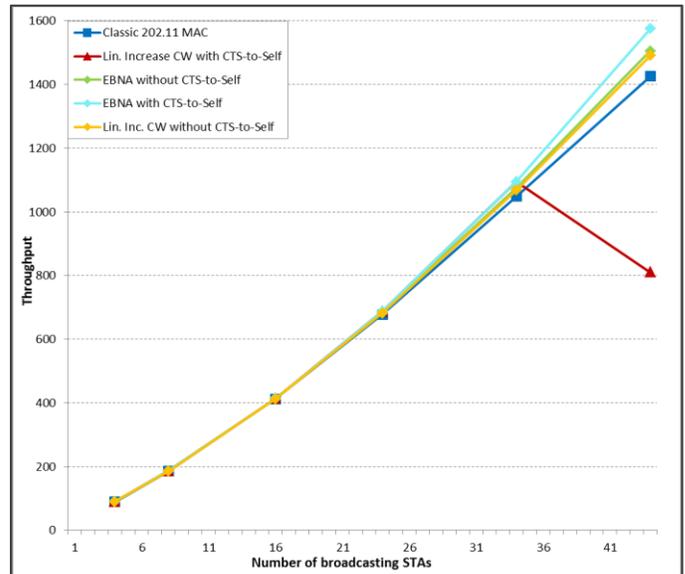

Fig. 9.   Overall Throughput

The EBNA with the use of the CTS-to-Self protection mechanism it performs better than the linear increase without CTS-to-Self. This is happening because although the EBNA guarantees no collision between broadcasting traffic, the additional traffic of the CTS-to-Self collides with the unicast packets. Finally, we can see that when linear increase of CW is used in conjunction with CTS-to-Self, the additional traffic created by this message leads to a network breakdown due to buffer overflow

b) End-to-End delay

This statistic shows the overall end-to-end delay for the entire network. As it is expected, the delay increases in all modified MAC processes because in all of them a wider CW it is used. It remains however in acceptable levels. Classic 802.11 has the lowest delay which however represents a smaller number of successfully delivered packets. Figure 10 shows the overall end-to-end delay for all the five cases. Here again, in the linear increase of CW with the use of CTS-to-Self when the number of broadcasting STAs increases the network breaks down and the delay become prohibitively high.
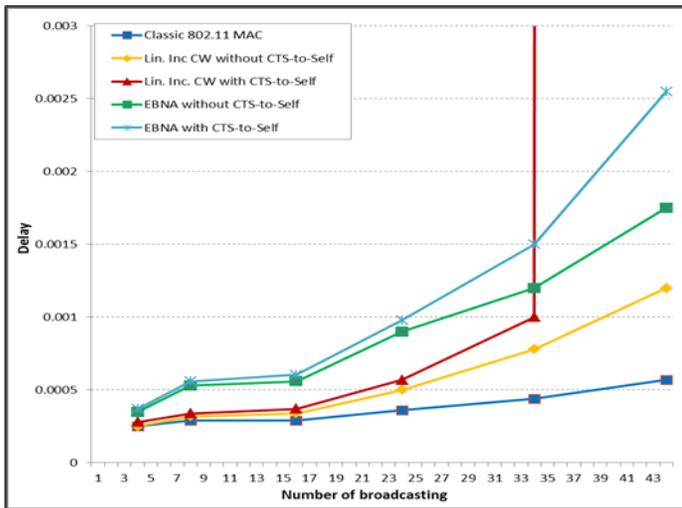
Fig. 10. End-to-End delay

c) Retransmission attempts

This statistic measures the average retransmission attempts for each packet for the entire network. It is directly affected by the throughput, and as we can understand concerns only the unicast transmission. Broadcasting STAs have no chance for retransmission. Figure 11 shows the retransmission attempts for the overall network

d) Backoff Slots

The average number of backoff slots is measured in this statistic. As it is mentioned above, there are two different types of MAC algorithms running simultaneously in all simulations. The unicasting STAs use the classic 802.11 MAC and the broadcasting STAs which comply with the cases described in section IV, C. Therefore, two separate statistics for the number of backoff slots are collected in this study. Figure 12 shows the average backoff slots measured in each Broadcasting STA. Figure 13 shows the average backoff slots measured in each Unicasting STA.
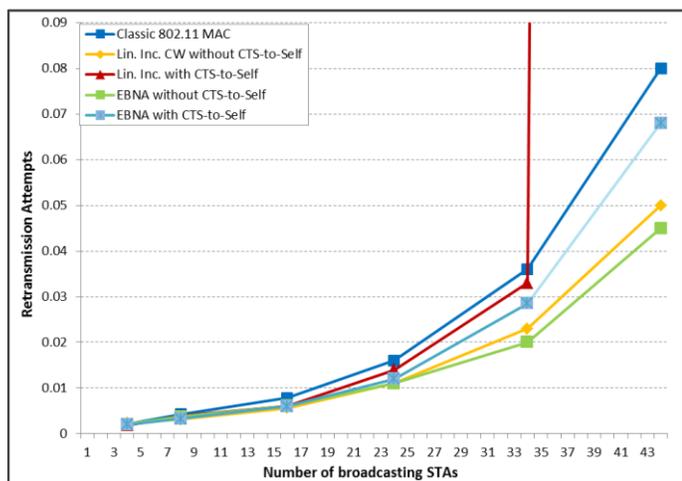

Fig. 11. Retransmission Attempts

In broadcasting, the use of the EBNA algorithm cause longer waiting times mainly due to the bigger size of the CW. The linear increase of CW gives acceptable low number of

backoff slots up to 34 broadcasting STAs. For bigger number of STAs, the network collapses and in the case that CTS-to-Self is used, the number of backoff slots increases dramatically. An interesting observation in the graph of figure 12 is that the increase of the number of backoff slots is not vertical. That gives the evidence that network's breakdown take place in the unicasting and not in broadcasting STAs.

In the unicasting STAs (Fig. 13) where the classic 802.11 MAC is used and also the generated traffic is lower, the number of backoff slots is significantly low and uniform. Exception is the case of linear increase of CW window with the use of CTS-to-Self where the network collapses.
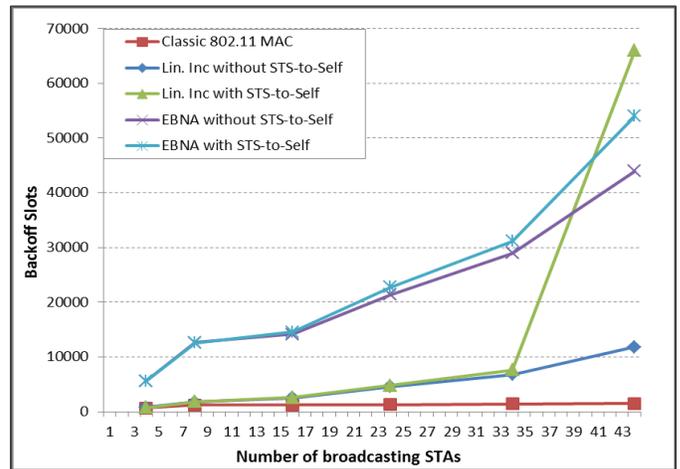

Fig. 12. Number of backoff slots for Broadcasting STAs
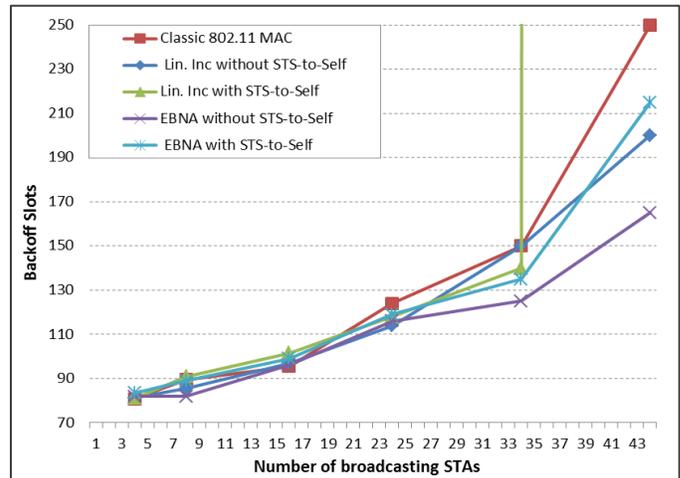
Fig. 13. Number of backoff slots for Unicasting STAs

e) Collisions

This statistic describes the total number of collisions encountered in the entire network during each simulation. This is not a standard OPNET statistic. In order to obtain this measurement the OPNET wlan_mac process is equipped with a counter which increases every time the collision flag in OPNET is set. The accuracy of this custom statistic was validated using the OPNET collision status statistic which indicate the present of a collision but cannot provide the total number of collisions. It is shown from figure 14 that the lowest

number of collisions is achieved when the EBNA without CTS-to-Self is used. This is because the EBNA algorithm is designed to minimize collisions in broadcasting STAs.
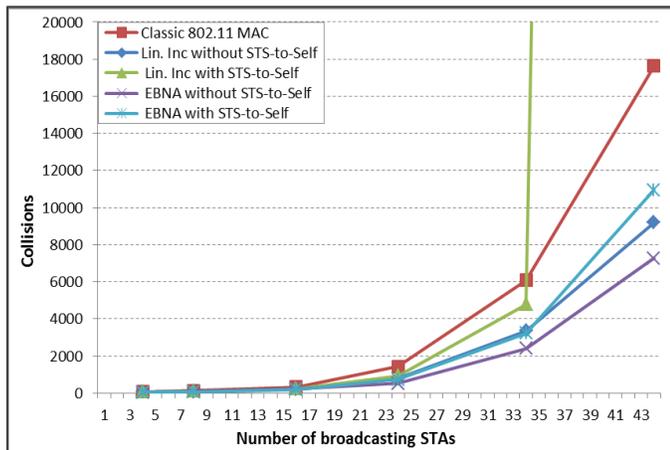

Fig. 14. Number of backoff slots for Unicasting STAs

We can clearly see here that classic 802.11 MAC cause a significant increase of collision when is used as medium access in multiple broadcasting wireless network.

V.   CONCLUSIONS

In this paper we explore the ability of the modified IEEE 802.11 MAC mechanisms to improve broadcasting performance while are operating in conjunction with a regular wireless network, and also to define the cases were the use of CTS-to-Self-protection mechanism can improve the overall performance of the network. For that reason two alternative MAC mechanisms are proposed. The first one is a linear increase of the CW, directly proportional to the number of broadcasting STAs in the wireless network. The second is simple algorithm which allocates exclusive backoff numbers to each STA that needs to backoff, keeping fairness between STAs during the random backoff process. In addition an extended application of the CTS-to-Self control message is proposed in order to be used for the distribution of network allocation information. All the above techniques have been independently tested, and have demonstrated that they can improve the performance of multiple broadcasting of multimedia type data in wireless ad-hoc networks [14].

In this study all the above modifications are tested in a "real life" scenario in which the modified MAC mechanism coexists in the same BSS with classic IEEE 802.11 devices. The performance of the network is studied as a whole in order to investigate the possibility of modified and classic wireless MAC mechanisms to function together.

The analysis of the results showed that the modified MAC mechanism can coexist with the classic Wi-Fi and also to improve the performance in the case of multiple broadcasting of media data. More specific, a better overall throughput can be achieved while the expected increase of the delay remains in acceptable level. When it comes to the use of CTS-to-Self mechanism, although it slightly improves throughput, it turns that is not an appropriate technique in this "mixed" networks. The significant increase of the traffic that is created by CTS-to-Self messages in combination with the inability of classic 802.11 to manage broadcasting, decrease the performance of the network and in some cases leads to collapse.

Future research can focus in the design of an adoptive EBNA algorithm where the size of the CW will be dynamically controlled taking into account the current traffic of the wireless network. This will significantly decrease delay and will increase overall performance of broadcasting in the wireless network.

REFERENCES

[1]   Yamaha System Solution "white paper", an introduction to networked audio, Yamaha Comersial Audio 2006, C.Bakker, H.Hamamatsu, T.Harison

[2]   AES White Paper: Best Practices in Network Audio, JAES Volume 57 Issue 9 pp. 729-741; September 2009

[3]   A. Floros T. Karoudakis, Delivering high-quality audio over WLANs, AES 116th Convention, 2004.

[4]   Kevin Gross, Audio Networking, Application and Requirements, JAES Volume 54, No.1/2, January/February, 2006

[5]   Nakjung Choi; Yongho Seok; Taekyoung Kwon; Yanghee Choi, "Leader-Based Multicast Service in IEEE 802.11v Networks," Consumer Communications and Networking Conference (CCNC), 2010 7th IEEE , vol., no., pp.1,5, 9-12 Jan. 2010

[6]   Fluke networks White Paper, Ensuring 802.11n snd 802.11 a/b/g Compatibility, 2011

[7]   IEEE Standard, Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, 2007M. Young, the Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

[8]   Brad Williams Tracy Camp, Comparison of Broadcasting Techniques for Mobile Ad Hoc Networks, 3rd ACM international symposium on Mobile ad hoc networking & computing, 2002

[9]   Shiann-Tsong Sheuy, Yihjia Tsai and Jenhui Chen, A Highly Reliable Broadcast Scheme for IEEE 802.11 Multi-hop Ad Hoc Networks, Communications,. ICC 2002. IEEE International Conference, 2002.

[10]   Jenhui Chen and Muwen Huang, Broadcast Engagement ACK Mechanism to Support Reliable Broadcast Transmission in IEEE 802.11 Wireless Ad Hoc Networks, Vehicular Technology Conference, 2004.

[11]   Xiaomin Ma Xianbo Chen, Performance Analysis of IEEE 802.11 Broadcast Scheme in Ad Hoc Wireless LANs, IEEE Transaction on Vehicular Technology, VOL.57, NO.6, November 2008.

[12]   C Chousidis, R Nilavalan, A Floros, Enhancement of IEEE 802.11 in Handling Multiple Broadcasting Audio Data in Wireless ad-hoc Networks, Journal of the Audio Engineering Society, Vol 61, Issue 4, April, 2013

[13]   C Chousidis, R Nilavalan, Improving multiple broadcasting of multimedia data traffic in wireless ad-hoc networks, Science and Information Conference 2013.

[14]   C Chousidis, R Nilavalan, L Lipa, Expanding the use of CTS-to-Self mechanism to improving broadcasting on IEEE 802.11 networks, IEEE Wireless Communications and Networking Conference (WCNC) 2014, (under review).

Constructing and Monitoring Processes in BPM using Hybrid Architectures

José Martinez Garro

Universidad Nacional de La Plata
UNLP
La Plata, Argentina

Patricia Bazán

LINTI – Universidad Nacional de La Plata
UNLP
La Plata, Argentina

Abstract— With the entrance of BPM in the Cloud, a change in the conception and design of Business Processes has been produced. Distributed environments, in this context offer computing possibilities which are advantageous for processes, especially in a decomposition context. This last concept has been introduced in BPM allowing processes to be executed in a cloud environment as well as in an embedded one. This situation takes advantage of both approaches under criteria like sensitive data, high computing performance and system portability. An unexplored aspect in current bibliography is process monitoring over a decomposed environment. In the present article we introduce the analysis of some concepts presented in current bibliography, and we propose also the architecture for a distributed process monitoring system. In this architecture we consider different design factors like location transparency, and the data needed for instance tracking over a cloud system.

Keywords — BPM; Cloud Computing; Execution; Monitoring.

I. INTRODUCTION

In this article we face the problem of including a Business Process Management System (BPMS) in a cloud oriented collaborative environment, with the particularity that it is an external environment to the organization. It is one of the purposes of this work to make a current bibliography analysis in sections II to VII, where we describe the different variants of a cloud model, its benefits and cons, hybrid architectures with embedded systems and the problem of monitoring a distributed process. Then, from section VIII we introduce the architecture of a process monitoring application. Finalizing the document we present some conclusions about the current state of the art and future work proposals in this research line.

II. RELATED WORKS

There are different trends in what comes to BPM in the cloud, but they are different if we are talking about research fields or trends in the market. Currently we can find research works tending to analyze the different paradigms of BPM (whether in the cloud or embedded), and how they escalate according to user's needs, connectivity that grows and mobile device incorporation. In [1], [2] and [3] especially we found trends like adaptive workflows and complex events.

These references support the idea of the hybrid architecture and the necessity of monitoring a distributed process using a centralized application. Regarding the other references, we will cite each one of them in every related topic.

In relation to the commercial market, we find fewer advances than in the research area. Most of the available BPMS in the cloud are very similar to the embedded ones, and the concepts introduced in the present work, in [2] and [29] like decomposed processes (or dynamic services) are not present. At the same time, most BPMS support local process monitoring, which is not equivalent to monitor a process instance distributed in different servers. In this paper we introduce further our approach for a monitoring application that gathers information from different servers in a complex architecture and displays it seamlessly.

III. BPM AND CLOUD

With the fast technological development in the context of application launching and execution using cloud based architectures, companies that began to choose this model are facing new problems. In particular, collaborative business processes with several interaction areas offer an optimization potential through the combination of cloud computing and BPM. A common factor between both paradigms is the flexible and agile approach. The cloud based computing model may be considered as an enabler for an improved combination of service oriented architectures, and also an agile procedure for Business Process Management. But this potential depends on the conditions imposed by the different frameworks, which can be viewed from technical and financial aspects.

A. Technical view:

From a technical point of view there are three dimensions in order to design, implement and successfully operate the different BPM tools in a cloud environment. These dimensions are: programming, integration and security.

- Programming: complex and distributed systems are easily reachable in current IT. In connection with obtaining more usability and flexibility, this complexity represents new requirements for Software Engineering. To solve this problem it is necessary to adopt new languages. So, based on new concepts and innovative techniques, the efforts invested in the development phase have been reduced to convert the complexity of these new aspects into a manageable element.
- Integration: this category can be divided in data integration, function integration and process integration. Under the light of the new challenges involved, the current topic plays an important role

in different scenarios. For example, a cloud based workflow can control distributed activities beyond the companies' border, mainly due to its easy accessibility. For a simplified execution of several process instances it is necessary to have integration interfaces and structured methods that allow joining the new components under the considered process.

- Security: this concept can be divided into three categories: functional security, information security and data security. All these categories have a significant relevance for BPM, especially in regard to business process grids and distributed process servers. Functional security specifies how the current status corresponds with the desired functionality status. The information security is focused in unauthorized changes or information extractions, as well as data security is in charge of the process related data.

Even more, from a technical point of view the question on "what processes are more appropriated to be executed in a cloud-based architecture" should be responded. Possible risks, such as insufficient integration options, location and integrity problems as well as programming interfaces should be taken in consideration.

B. Financial view

There are two dimensions from the financial point of view:

- Availability: the services provided by a cloud infrastructure can be accessed at any time because of the high availability model. Based in a high abstraction level, the customization and installation are significantly easier. In addition with this simplification, the final user is capable of working with the service immediately.

- Investment risk: in the context of the different variable billing models (for example "pay per transaction") the use of a cloud based service results in certain charges. These charges contain relevant costs given by transferences and transactions [1] [2] [3] [4].

IV.   BENEFITS AND DRAWBACKS

Cloud based BPM provides users the possibility of using software in a "pay per use" way, instead of forcing them to make big investments in BPM software, hardware and maintenance, versus the traditional licensing applications. Systems can escalate up and down according to the user's needs. This means they do not have to worry about the over/ under resource provisioning because of the high adaptability provided currently by cloud service providers, as we can see in Figure 1.

The current model, on the other hand, has several low points. By putting a BPMS in the cloud, users may lose control over sensitive data. This aspect results major considering that business processes inside an organization may manage important information for it and its members. On the other hand, the non-high computational activities' efficiency and effectiveness cannot be increased by putting them in the cloud,

but rather these activities may get more expensive. For example, an activity which is not intensively computational could need to process a certain amount of data. The transference of these data to the cloud could take more time than the transmission to an embedded version installed locally. That transference could result bigger than the real necessity of processing. Even more, the cost of the activity may increase due the data transference. This element is one of the billing concepts in a cloud computing system because of the high connection availability [1] [2] [5] [6].
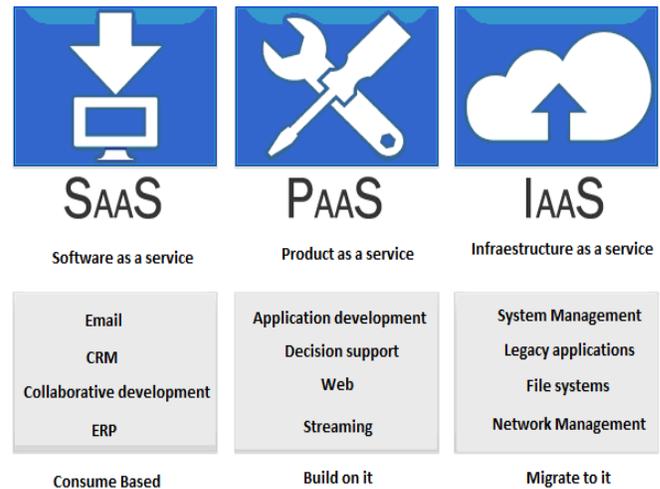

Fig.1.   Service model

V.   SERVICE MODEL

A. Infrastructure as a service (IaaS)

When an application is moved to an IaaS model, the cloud user is responsible for the operating system, the middleware and the applications running on the virtual machine. The action of installing BPM software in an IaaS cloud solution is comparable to installing an embedded BPMS, since everything except the hardware is managed by the cloud user. Furthermore, the user has to make some security decisions in order to avoid intrusions. According to this, possible security measures are: port blocking, access control policies and updating the applications and the operating system frequently.

B. Product as a Service (PaaS)

By positioning a workflow based application in a model like "Product as a Service", the responsibilities for the user and the cloud provider are different. The execution engine is assumed as a part of the platform, so it is offered by the service provider. Users must upload their processes to run them in the cloud. The engine can be used by several users since the platform is shared. The responsibility for data storage and management is no longer in charge of the user, who has to deal with several security issues:

- The process models should not be readable by intruders in posession of a description file.

- Process models should not be altered by intruders.

- Process models should not be deployed in other servers.

In order to achieve these requirements, the process model descriptions should be encrypted and signed. The encryption ensures that process models are not readable by intruders. By the action of signing them, it can be assured that a file is only valid for a particular execution engine, and using it to point to another execution engine will provoke an error. This turns into utility considering that the same server can be accessed by different users in a shared environment.

Storing the application database can be an issue also. Data should be encrypted in order to not be readable by intruders. Data encryption in a relational database generates expressivity issues with queries using relational operators. For example, joins can have problems in an encrypted data context.

C. Software as a Service (SaaS)

By moving an application to a SaaS model, the cloud provider is now responsible for the application itself. The application is no longer an asset of the enterprise cloud user but it is offered by the cloud provider. The application may be given to multiple cloud users in a single or multiple tenant architecture. In a single tenant paradigm, an execution engine is installed for each process model. In a multi-tenant environment, multiple users and process models are served by a unique engine. The data stored by the cloud provider should be assured in order to prevent unintended accesses, both by the service provider or other users in the cloud. The same measures we have mentioned in the previous subsection related with signing and encryption can be applied to solve this problem.

In a multi-tenant architecture, different users access the same execution engine. The data used by one user should not be accessible to other cloud users. There are two possible solutions for this problem: in the first place, a database for each cloud user can be created. As an alternative, a column to each table where the user identifier is saved can be added. It is necessary to observe the scalability of both solutions: the amount of users could increase, and because of that, the need of resources too [1] [6] [7] [8].

VI. COMBINATION OF EMBEDDED AND CLOUD SCHEMES

Privacy protection is one of the barriers to execute BPM in a cloud environment. Not all users desire to put their sensitive data outside the organization. Besides, it is necessary to observe product's portability and versions, and their availability in a cloud system. Another not minor problem is the efficiency.

The intensive computing activities may obtain benefits in the cloud due to the scalability and the computing force high availability. The non intensive computational tasks, on the other hand, not always take advantage of this context. The performance of one activity running in an embedded environment should be better than in the cloud because of the data that are transferred in order to execute the activity. These activities could also result expensive due to the fact that data transference is a billing criterion in the cloud [11] [12].

- Architecture: in most BPM solutions, the process engine, the activities and the process data are located in the same side, even in an embedded or cloud solution. There are some papers introducing the PAD model

(Process - Activity - Data) of Figure 2 as a distribution possibility for BPM in the cloud. In this approach, the process model, the involved activities and the data are separately distributed. The PAD model defines four possibilities of distribution:
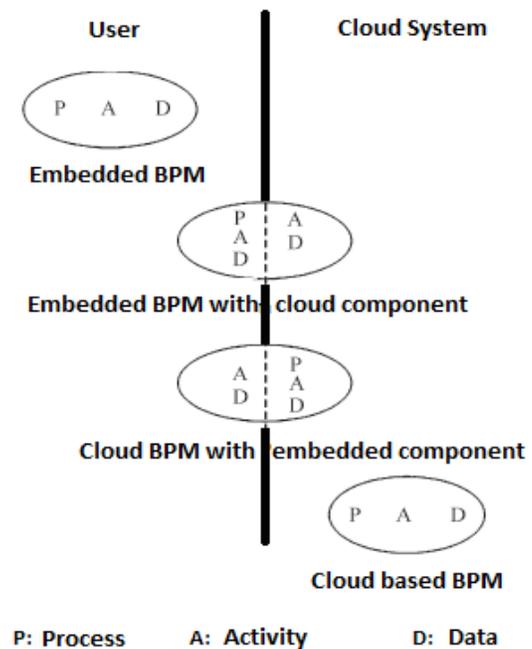

Fig.2. PAD Distribution Schema [6]

1) The first pattern is the traditional alternative where all elements are distributed over the final user side.

2) The second pattern is useful when the user already has a BPMS, but the high computing activities are located in the cloud to increment their performance.

3) The third pattern is useful for the users who still do not have a BPMS, so they can use the cloud system in a "pay per use" way. In this approach the activities with low computing intensity or the ones with sensitive data management can be located on the final user side.

4) The fourth pattern is the cloud based model where all the elements are located in the cloud.

- Business processes consist of two kinds of flows: control and data. Control flows regulate the execution of activities and their sequence, while data flows determine how the information is transferred from one activity to the other inside the process. BPM engines must deal with the control of both kinds of flows. A data flow could contain sensitive data, so when a BPMS is deployed in the cloud, the content of those flows should be protected. An example of the proposed architecture could be a scenario where the engine in the cloud only deals with data flows using reference identifiers instead of real data. When an activity needs sensitive data, the data being transferred to the activity are managed under user supervision in an encryption

tunnel. Sensitive data are saved in the final user side, and non sensitive data are saved in the cloud. This schema allows that sensitive data do not travel indiscriminately through the web.

- Optimal distribution: the cloud system costs have been an object of study in different articles. There are several formulas to calculate the optimal distribution of activities, since they can be located in the cloud or in an embedded system. The calculation takes in consideration time costs, monetary costs and privacy risk costs. By using the formulas, users can make cost estimations about deploying part of their applications in an embedded or cloud system alternatively [2] [5] [6] [9].

VII. PROCESS DECOMPOSITION

It is possible to generalize the distribution and identify a fifth pattern where the process engine, the activities and data are deployed in the cloud and in the final user. This solution presents two potential benefits:

1) The process engine regulates control and data flows. One activity receives data from the process engine and after its execution the produced data are passed again to the process engine. Consider now a sequence of activities located in the cloud, while the process engine is deployed in the final user. Each activity uses data produced by the previous activity as an income. Data are not passed directly from one activity to the other but they are sent to the process engine first. Since data transference is one of the billing factors in this model this kind of situations could become more expensive when large amounts of data are transmitted between activities. To avoid this problem a process engine can be added to the cloud, in order to regulate the control and data flows between activities located inside it. When a sequence of activities is located in the cloud, data are regulated by the process engine in the cloud. This reduces the amount of data to be transmitted between the cloud and the embedded system.

2) When the cloud is not accessible, users can execute business processes in a complete way in the embedded system until the former one is available again.

In order to run a single business process between two separated engines, it should be split into two individual processes. It could be convenient for the users to take a distribution list of the process and its activities. The process can be automatically transformed into two business processes, one in the cloud and the other in the embedded system. The communication between both systems can be described using a choreography language, like BPEL. Besides, the distribution list can be created automatically according to the optimal distribution formulas mentioned in subsection VI [13] [14].

Business process monitoring is more complicated now, since the process has been divided into two or more parts. As a solution, a monitoring tool can be developed for the original process, through the combination of the individual process monitoring details. This point will be analyzed further.

A possible approach to manage the process decomposition is to identify its structure and semantics. When the control and data dependencies are identified, the consequences of moving some activities from the embedded system to the cloud and vice versa can be researched. When the activity distribution consequences are known, a transformation model can be created.

Then, a business process and a list with marks are used to create two separated processes, one for the cloud and another for the final user. Also, a choreography description can be generated in order to describe the communication between both processes using some standard language, like BPEL [6] [10].

VIII. HYBRID SCHEMES IMPLEMENTATION

The possibility of locating a BPMS in an external space to the organization (for example in a cloud computing architecture with a SaaS model) makes feasible to access it from inside the organization through an Internet connection, as well as from any other external point. Considering this fact, besides the possibility of having clients accessing from mobile devices, the access points to the cloud are incremented.

This generates the following issues about process execution, and their corresponding proposed solutions:

- **Process Decomposition:** as exposed in Section V, the fact of putting a BPM server in the cloud generates the problem of what to do with sensitive data management. Facing this problem, this solution can be enounced: in case of publishing the corporative database (or at least part of it) in a cloud environment is not a viable choice according to the organizational security policies, the decomposition of the process is going to be necessary in order to implement a hybrid scheme. In this scenario, the high computing activities can be located inside the cloud in order to take advantage of the computing performance, and the activities that make use of corporative sensitive data are located inside the organization in an embedded installation.

- **Decomposed process synchronization:** the disaggregated process is formally divided into sections according to the amount of involved servers. According to this, it is going to be necessary to solve how to synchronize the servers in order to ensure the execution sequence. There are, in theory, different ways to implement the synchronization, such as by using messages or event monitoring. Using messages, the end event of each process part invokes the start event of the next one. This can be made through start and end message type events, included both in the last version of standard BPMN (Business Process Management Notation), where the execution of the end event of a process throws a message to the BPMS in order to notify the finalization, and require the execution of a process previously parameterized. The notifications can be implemented by using a message queue and a daemon for pooling. This daemon receives messages and initiates instances of the required process. In this way, each server in this hybrid model (the embedded

and also the cloud based ones) must have a copy of the pooling service in order to receive the finalization notifications and later notify the process engine. The result of this is to initiate instances corresponding to the requested definition [15] [16].

- **Decomposed process monitoring:** the biggest problem of having a process partitioned orientation is to monitor the different distributed instances, and at the same time to accomplish an integrated model of them under the optic of the "real process" which they belong to. In order to solve this inconvenient the following solution can be analyzed: in first place it is necessary to associate the different instances with the original process, in order to recover them from the existing servers. Once they are recovered, some kind of application in charge of gathering data and showing them seamlessly should be provided. The most important thing in this aspect is to accomplish monitoring transparency for the user, without forcing him to distinguish the server where each activity has been executed. This fact provides thus an integrated visualization of the different instances by seeing them as a unique entity. The implementation of the current feature should be made by a cloud resident web application, located there in order to access every involved server, whether cloud or embedded, and to ensure user access from any point. For this purpose it is important for the application to have a catalog with every existing server in the architecture, with their location information updated. Each involved server will have a copy of a web service which receives a process definition identifier and returns information about every existing instance associated with the sent definition. The returned information includes instance identification, current status (running, completed, suspended), current activity in case of non-completion status, start and end date. According to this, the cloud resident web application sends an invocation of the web service with the selected process definition as a parameter to each server, and receives the information of the associated instances. Then this information will be visualized in a web interface where the user can select a particular instance and observe its details. For this purpose the application contains a web service to require to each server the details of the associated activities. The information returned includes identification of the activity, associated participant, start date, current status and end date. After receiving this information the web application will allow the user to observe some activity details transparently, without indicating the server information where they were executed. This helps to accomplish location transparency [6] [17] [18].

IX. MONITORING PROCESSES IN THE CLOUD

As we have seen previously, the biggest problem about using a partitioned process model is to gather and monitor the different distributed instances (either in an embedded system or in the cloud), and at the same time to accomplish an integrated view under the optic of the "original process" which they belong to. To face this inconvenient we have designed a solution considering distributed and intercommunicated components forming an architecture, which is described as follows.

On the one hand, it is going to be necessary to associate the different process instances initiated in a chain, with the purpose of gathering information about them accessing the different involved servers. The execution model of decomposed processes consists of linking each instance flow to the corresponding partitioned processes. Thus, when an instance finishes in a server, it initiates automatically a new instance corresponding to the next process partition, depending on the distribution architecture. For this purpose, each node in the architecture should be capable of establish communication with the next node in order to initiate new instances, and gather in this way information about them. Namely, given a new instance which was initiated in a node of the architecture, we should be able to obtain, not only its data but every instance generated by it in another server [29].

A. Bonita Open Solution: API and connectors.

There are several ways of implementing instance flow linking. In our case we have selected Bonita Open Solution [30] as the BPMS. In this way, once the original process was partitioned over the servers, following criteria like sensitive data storing, data transferring and application portability, we have used the API and connectors provided by the BPMS in order to create instances and recover their information using Java classes. These classes use the API as libraries, including functions like server authentication, instance launching, instance information gathering and process variable setting. These classes are invoked from the process definition using connectors.

It was also included in each process definition the information needed for the communication with another Bonita server inside the architecture, and in this case, by using connectors, launch new instances in that server. Thus, every instance when is finished will execute the connector which allows initiating a new instance by using the API, linking in this way automatically the process execution flow [19] [20].

B. Centralized front-end

As it was described initially in section VIII, a monitoring application must be developed in order to show integrated data related with distributed instances. Facing the execution link, it is very important for each instance to be able of storing, not only their own information but the one associated with the instances created by them over other servers. In this way, by accessing the initial instance of the process, it is possible to recover the information associated to the next instance, and so on in order to obtain the complete flow of the process. Once recovered the execution chain in the different servers, it must be provided an application for visualization in charge of gathering data and show them seamlessly.

The most important thing in this aspect is to accomplish monitoring transparency for the user: he should not be forced to distinguish the server where the activity was executed but he should visualize seamlessly the different instances and observe them as a unique entity. The implementation of this feature was made through a web application located in the cloud, following the criteria established in Section VII. This application was placed there in order to access each involved server, being them cloud or embedded, guaranteeing in this way user access from every point. For this purpose it is important for the application to have a catalog with the existing servers in the architecture considering their location information updated. Each of these servers has a copy of a web service (*getInstanceService*), which receives a process definition id and returns information of each instance existing in the server associated with the definition sent as a parameter. The information returned includes instance id, current status (executing, completed, suspended), current activity if the instance is not finalized, start and end date. In this way, the application located in the cloud sends to each server a web service invocation with the selected process definition as a parameter, and receives the information of the associated instances. Then, this information is visualized in a web interface, where the user can select a particular instance and observe its details. In order to make this, the application has another web service (*getInstanceActivityService*) used to get from each server the details of each activity associated to the instance. The returned information includes activity id, participant, start date, current status and end date. Once ended this collection phase, we need to remember that each instance contains also the information of the different instances initiated over the different servers in the architecture. In this way, the web application will have to concatenate the information received about the different instances and allow to the user to observe the monitoring details in a transparent and integrated way, without indicating him (unless he asks for it, for administration purposes) the information of the server where each activity was executed, accomplishing in this way location transparency [21] [22] [23].

C. Application's architecture

We can observe in Figure 3 the different distributed components identified in the architecture design, as well as the internal relationship between them and the user.

The solution is composed by three main nodes: the cloud, the embedded or traditional system and the monitoring application. The cloud works as the container of several elements: the BPMS, the monitoring application, the REST API used by the developers in order to integrate the applications with the process engine, and eventually a geolocation service which allows assigning to mobile clients the most convenient version of the service according to where they are.

On the other side we find the embedded type components, namely traditional BPM applications which belong to the organization, and because of different reasons like data sensibility or application portability, it could be decided not to locate them in the cloud. These nodes, functionally talking, take a role which is equivalent to the cloud node's behavior,

even when they have access restrictions and lower computing force compared with the first ones.
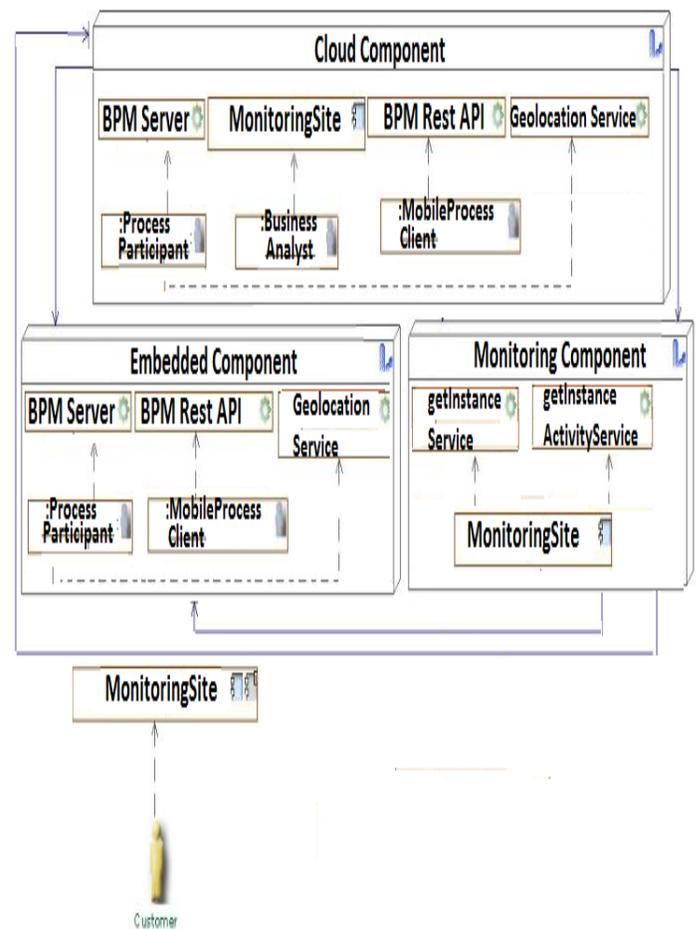

Fig.3.   Application architecture and user location

The third component is related with monitoring. It is used by the monitoring application, and is in charge of returning information about instances and activities which were executed in every node of the distributed architecture. The web services getInstance and getInstanceActivity were constructed jointly with the monitoring application, and are executed on demand by this one. They are communicated with the process servers through an API (in our case, the Bonita one), and are in charge of returning, in first place, information about the instances initiated on each servers, and once these were accessed, return data about the activities that compose them [24] [25] [26].

D. Component communication

If we consider every component present in the architecture, we have analyzed the communication between each one of them through an application communication diagram. There we can observe the most important involved applications, their main actors and the interaction of the different distributed software components.

We can see at the same time the different user profiles involved in the execution of the components represented in the architecture.

While the preponderant role in the process execution is the activity's participant, the monitoring site results are important for the business analyst, as well as for the architecture administrators which can optimize the services or process components (Figure 4).

A feature in common between the process execution application and the monitoring one is the location transparency. Users should not be necessarily notified about the execution environment change, in case we are considering a decomposed process where the activities are located in different servers. This is very useful in order to allow users to have a unified vision of the process, more than a partitioned one, which main existence reason is related with taking advantage of technical resources.

We can also visualize in Fig. 4 how both the execution and the monitoring components access indistinctly to the cloud or embedded nodes, in order to gather information about each instance initiated in the distributed servers [27] [28].
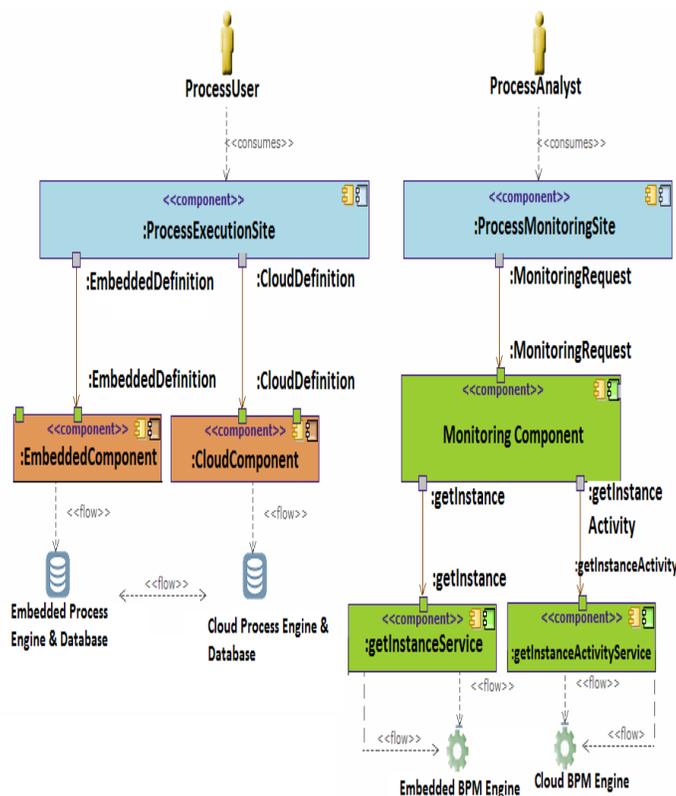

Fig.4.   Application Communication Diagram

X.   CONCLUSIONS

As we could observe, BPM as well as many other specialties in IT, have suffered changes due to the different service models in the cloud. This has forced specialists to consider new process design and implementation variants which allow using different advantages offered by the quoted paradigm. Facing the possibility of using unlimited computing force and high availability, some new decomposition process schemes appear in order to divide a process along some distributed server architecture.

Even when this approach allows using efficiently technological resources and protecting the organization's sensitive data, it is not necessarily easy to implement, and many times depends on the subjacent cloud infrastructure and the selected process server. In the present article we have used Bonita Open Solution because it is open source, and has an API which allows, through using connectors, accessing the different servers of the architecture. Without this last component it is very difficult to initiate new instances in different servers, and accomplish in this way the execution link of a decomposed and distributed process.

On the other side, as we said previously, even when process decomposition is a highly explored subject in current literature, the scenario is not the same with distributed process monitoring. This topic, at a glance, is not easily soluble. In a traditional business process model, the information source to monitor is in the same node that executes and monitors processes, while in a distributed environment instances are located in different servers. For this reason, different mechanisms are needed in order to gather data about executed instances, as well as to link them and provide an integration perspective under the light of the original process.

Currently, our research interest is focused on improving the monitoring application, allowing different filters for the users. The objective of this is to monitor efficiently each node of the architecture and optimize eventually the performance in some of them.

On the other side, it results important also to analyze different modifications to the BPMN notation present in current bibliography, which would allow including in process models semantic associated with decomposition, as well as interconnections between distributed servers.

REFERENCES

[1]   T. Kirkham, S. Winfield, T. Haberecht, J. Müller, G. De Angelis, "The Challenge of Dynamic Services in Business Process Management", University of Nottingham, United Kingdom, Springer, 2011

[2]   M. Minor, R. Bergmann, S. Görg, "Adaptive Workflow Management in the Cloud – Towards a Novel Platform as a Service", Business Information Systems II, University of Trier, Germany, 2012

[3]   M Mevius, R. Stephan, P. Wiedmann, "Innovative Approach for Agile BPM", eKNOW 2013: The Fifth International Conference on Information, Process, and Knowledge Management, 2013.

[4]   Dr. Manuel Goetz, "Integration of Business Process Management and Complex Event Processing", Germany, 2012.

[5]   M. Gerhards, V. Sander, A. Belloum, "About the flexible Migration of Workflow Tasks to Clouds -Combining on and off premise Executions of Applications", CLOUD COMPUTING 2012: The Third International Conference on Cloud Computing, GRIDs, and Virtualization, 2012.

[6]   Evert Duipmans, Dr. Luis Ferreira Pires, "Business Process Management in the cloud: Business Process as a Service (BPaaS)", University of Twente, April, 2012.

[7]   S. Aleem, S. Molnar, and N. Mohamed, "Collaborative Business Process Modeling Approaches: A Review", In Proc. of the 2012 IEEE 21st International workshop on Enabling Technologies: Infrastructure for Collaborative Enterprises, pp. 274-279, June 2012.

[8]   Dirk Fahland, Wil M.P. van der Aalst Eindhoven, "Simplifying Discovered Process Models in a Controlled Manner", University of Technology, The Netherlands, 2012.

[9]   Hubert Scheuerlein, Falk Rauchfuss, Yves Dittmar, Rüdiger Molle, Torsten Lehmann, Nicole Pienkos, Utz Settmacher, "New methods for

clinical pathways –Business Process Modeling Notation (BPMN) and Tangible Business Process Modeling (t.BPM)". Springer-Verlag 2012.

[10] Marielba Zacarias, Paula Ventura Martins, "Collaborative methods for Business Process Discovery", Portugal, Springer-Verlag 2012.

[11] Jiri Kolar, Tomas Pitner, "Agile BPM in the age of Cloud technologies", Scalable Computing: Practice and Experience, 2012.

[12] Andreas Lehmann and Dirk Fahland , "Information Flow Security for Business Process Models - just one click away", University of Rostock, Germany, 2012.

[13] Rafael Accorsi, Thomas Stocker, Günter Müller, "On the Exploitation of Process Mining for Security Audits: The Process Discovery Case", Department of Telematics, University of Freiburg, Germany, 2012.

[14] Aleš Frece, Gregor Srdić, Matjaž B. Jurič, "BPM and iBPMS in the Cloud", Proceedings of the 1st International Conference on Cloud Assisted ServiceS, Bled, 25 Octubre 2012

[15] Dr. Luis Ferreira Pires, "Business Process Management in the cloud: Business Process as a Service (BPaaS)", University of Twente, April, 2012.

[16] S Balzert, P Fettke, P Loos, "A Framework for Reflective Business Process Management", 45th Hawaii International Conference on System Sciences, USA, 2012.

[17] Marco Brambilla, Piero Fraternali, and Carmen Vaca, "BPMN and Design Patterns for Engineering Social BPM Solutions", Politecnico di Milano, Piazza L. da Vinci 32, Milano, Italy, 2012

[18] Marco Brambilla, Piero Fraternali, Carmen Vaca, Stefano Butti, "Combining Social Web and BPM for Improving Enterprise Performances: the BPM4People Approach to Social BPM", WWW 2012, European Projects Track, Abril 16–20, Lyon, France, 2012.

[19] S. Balzert, P. Fettke, P. Loos, "Enhancement of traditional Business Process Management with reflection – a new perspective for Organizational Learning", Institute for Information Systems (IWi) at German Research Center for Artificial Intelligence (DFKI), Germany, 2012.

[20] Huang Hua, Zhang Yi-Lai, Zhang Min, "A Survey of Cloud Workflow", Jingdezhen Ceramic Institute, Jingdezhen, Jiangxi, 333001, China,

Proceedings of the 2nd International Conference On Systems Engineering and Modeling (ICSEM-13), 2013.

[21] Toàn Nguyên, Jean-Antoine-Désidéri, "Resilience Issues for Application Workflows on Clouds", Project OPALE, INRIA Grenoble Rhône-Alpes, ICNS 2012: The Eighth International Conference on Networking and Services, Grenoble, France, 2012.

[22] Markus D• ohring and Birgit Zimmermann, "vBPMN: Event-Aware Workflow Variants by Weaving BPMN2 and Business Rules", SAP Research, Darmstadt, Germany, 2011.

[23] Zhenyu Fang, Changqing Yin, "BPM Architecture Design Based on Cloud Computing", School of Software Engineering, Tongji University, Intelligent Information Management, Shanghai, China, 2010.

[24] Duipmans E Pires L. da Silva Santos L. "Towards a BPM Cloud Architecture with Data and Activity Distribution" 2012 IEEE 16th International Enterprise Distributed Object Computing Conference Workshops. ISBN 978-0-7695-4786-2/12

[25] T. Anstett, F. Leymann, R. Mietzner, and S. Strauch, "Towards bpel in the cloud: Exploiting different delivery models for the execution of business processes," in Proceedings of the 2009 Congress on Services - I. Washington, DC, USA: IEEE Computer Society, 2009, pp. 670–677.

[26] Roder, A.; Lehmann, M.; Kabitzsch, K., "Monitoring service choreographies," Industrial Informatics (INDIN), 2011 9th IEEE International Conference on , vol., no., pp.142,147, 26-29 July 2011. doi: 10.1109/INDIN.2011.6034852

[27] T. Dornemann, E. Juhnke, and B. Freisleben, "On demand resource provisioning for bpel workflows using amazon's elastic compute cloud," in Proceedings of the 2009 9th IEEE/ACM International Symposium on Cluster Computing and the Grid, ser. CCGRID '09. Washington, DC, USA: IEEE Computer Society, 2009, pp. 140–147.

[28] W. Fdhila, U. Yildiz, and C. Godart, "A flexible approach for automatic process decentralization using dependency tables," in *ICWS*, 2009, pp. 847–855.

[29] J.Martinez Garro, P.Bazan "Constructing hybrid architectures and dynamic services in Cloud BPM¨ Science and Information Conference 2013 October 7-9, 2013 | London, UK.

[30] Bonita Open Solution http://es.bonitasoft.com/. October, 2013.